




RESEARCH ARTICLE

Using spatiotemporal information in weather radar data to detect and track communal roosts

Gustavo Perez^{1,†} , Wenlong Zhao^{1,†} , Zezhou Cheng¹, Maria Carolina T. D. Belotti² , Yuting Deng², Victoria F. Simons², Elske Tielens³, Jeffrey F. Kelly³, Kyle G. Horton², Subhansu Maji¹ & Daniel Sheldon¹

¹College of Information and Computer Sciences, University of Massachusetts Amherst, Amherst Massachusetts, 01003, USA

²Department of Fish, Wildlife, and Conservation Biology, Colorado State University, Fort Collins, Colorado, USA

³Department of Biology, University of Oklahoma, Norman, Oklahoma, USA

Keywords

Aeroecology, convolutional neural networks, deep learning, machine learning, NEXRAD, ornithology, swallow roost, weather radar

Correspondence

Gustavo Perez, College of Information and Computer Sciences, University of Massachusetts Amherst, 140 Governors, Drive, Amherst, MA 01003. Tel: +(413) 545 4843; Fax: +(413) 545 1249; E-mail: gperezsarabi@umass.edu

Funding Information

This material is based upon work supported by the US National Science Foundation under Grant Nos. 2017554 to Colorado State University, 1749833 and 2017756 to the University of Massachusetts Amherst, and 2017582 to the University of Oklahoma.

[†]Equal contribution.

Editor: Vincent Lecours

Associate Editor: Matthew Van Den Broeke

Received: 1 October 2023; Revised: 18 March 2024; Accepted: 21 March 2024

doi: 10.1002/rse2.388

Remote Sensing in Ecology and Conservation 2024; **10** (5):567–583

Introduction

Weather radar is one of the most promising technologies for studying flying animals, with networks of weather surveillance radars around the globe continuously monitoring the airspace and detecting birds, bats, and insects in addition to precipitation (Kunz et al., 2008). The US Next Generation Weather Radar (NEXRAD) weather radar

Abstract

The exodus of flying animals from their roosting locations is often visible as expanding ring-shaped patterns in weather radar data. The NEXRAD network, for example, archives more than 25 years of data across 143 contiguous US radar stations, providing opportunities to study roosting locations and times and the ecosystems of birds and bats. However, access to this information is limited by the cost of manually annotating millions of radar scans. We develop and deploy an AI-assisted system to annotate roosts in radar data. We build datasets with roost annotations to support the training and evaluation of automated detection models. Roosts are detected, tracked, and incorporated into our developed web-based interface for human screening to produce research-grade annotations. We deploy the system to collect swallow and martin roost information from 12 radar stations around the Great Lakes spanning 21 years. After verifying the practical value of the system, we propose to improve the detector by incorporating both spatial and temporal channels from volumetric radar scans. The deployment on Great Lakes radar scans allows accelerated annotation of 15 628 roost signatures in 612 786 radar scans with 183.6 human screening hours, or 1.08 s per radar scan. We estimate that the deployed system reduces human annotation time by $\sim 7\times$. The temporal detector model improves the average precision at intersection-over-union threshold 0.5 ($AP^{IoU=0.5}$) by 8% over the previous model (48% \rightarrow 56%), further reducing human screening time by 2.3 \times in its pilot deployment. These data contain critical information about phenology and population trends of swallows and martins, aerial insectivore species experiencing acute declines, and have enabled novel research. We present error analyses, lay the groundwork for continent-scale historical investigation about these species, and provide a starting point for automating the detection of other family-specific phenomena in radar data, such as bat roosts and mayfly hatches.

network (Crum & Albery, 1993), in particular, has archived 25 years of data from 143 radar stations covering nearly the entire contiguous US (Ansari et al., 2018) and offers the possibility of monitoring flying animals at an unprecedented scale and resolution (Bruderer, 1997; Dokter et al., 2011; Gauthreaux et al., 2003; Gauthreaux & Belser, 1998; Gauthreaux, 1970). These data have fueled a growing number of studies at increasing scales about bird

populations, including studies about nocturnal migration and stopover behaviors (Buler & Dawson, 2014; Buler & Diehl, 2009; Cohen et al., 2021; Farnsworth et al., 2016; Gauthreaux et al., 2003; Horton et al., 2018), demography (Dokter, Farnsworth, et al., 2018), the effects of artificial light on migration (McLaren et al., 2018; Van Doren et al., 2017, 2021), systems to forecast migration (Van Doren & Horton, 2018), and landmark findings about the declines (Rosenberg et al., 2019) and shifting phenologies of North American birds (Horton et al., 2020). NEXRAD data have also produced insights into changing bat (Stepanian & Wainwright, 2018) and insect (Stepanian et al., 2020) populations in North America, while similar research programs are being carried out on other continents (Nilsson et al., 2019; Nussbaumer et al., 2019; Shamoun-Baranes et al., 2014).

Communal roosts occur throughout the US, especially in late summer and fall (Russell & Gauthreaux, 1998), and are often dominated by a single species, especially Purple Martins (*Progne subis*) and Tree Swallows (*Tachycineta bicolor*) (Bridge et al., 2015; Laughlin et al., 2016), depending on habitat and time of year. The roosts may also gather Barn Swallows (*Hirundo rustica*), Bank Swallows (*Riparia riparia*), and Cliff Swallows (*Petrochelidon pyrrhonota*), with occasional sightings of individuals of Northern Rough-Winged Swallows (*Stelgidopteryx ruficollis*) and Violet Green Swallows (*Tachycineta thalassina*) joining roosts of the aforementioned species (Winkler, 2006). These species of martins and swallows are aerial insectivores, which are rapidly declining in North America (Fraser et al., 2012; Nebel et al., 2010; Rosenberg et al., 2019), so taxon-specific results about their ecology are of great interest. In general, swallow roosts are one of a relatively small number of radar phenomena that can be traced to family level due to its distinctive expanding ring pattern (Horn & Kunz, 2008; Russell & Gauthreaux, 1998) (See Fig. 1A).

The NEXRAD archive includes more than 0.5 petabytes of data and 240 million scans, each of which may contain a variety of patterns corresponding to different types of precipitation, clutter, or biological scatterers. Therefore, methods are needed to automatically recognize, discriminate, and track different types of biological scatterers to collect measurements at large scales. AI algorithms based on convolutional neural networks have shown tremendous success at related visual recognition tasks, and are excellent candidates for recognizing biological patterns in weather radar. Past work has focused on discriminating precipitation from broad-scale bird migration in radar data through the use of AI classification (Horton et al., 2019; RoyChowdhury et al., 2016; Van Doren & Horton, 2018) or segmentation (Lin et al., 2019) methods. An AI-assisted system to monitor martins and

swallow roosting activity in past, present, and future radar scans could provide information urgently needed for basic science and conservation of these species. Past studies have used radar to study swallow roosts using human annotation to identify roosts in radar (Bridge et al., 2015; Kelly & Pletschet, 2017; Laughlin et al., 2013, 2016; Winkler, 2006), but cannot be easily repeated or expanded due to the cost of human effort.

In this paper, we investigate the design, deployment, and analysis of an AI-assisted system capable of extracting research-grade roost annotations from NEXRAD data as shown in Figure 1. Building on a prior system (Cheng et al., 2020), we make a number of novel contributions. We construct a standardized dataset of roost annotations to remove labeling style differences and to support standard training and evaluation methods. We enhance the detector neural network architecture to be able to process extra channels of radar moments at different elevations and from multiple scans, which allow the model to recognize the distinctive expanding movement dynamics of roost rings improving from 48% to 56% average precision at intersection-over-union threshold 0.5 ($AP^{IoU=0.5}$).¹ We develop open-sourced software for the automated roost detection and a user interface where humans can visualize and screen the machine predictions. We deploy the AI-assisted system on 21 years of radar scans from 12 radar stations near the Great Lakes region of the US, successfully extracting research-grade historical roost data for 8% of the radar stations in the contiguous US that cover a region with large swallow populations. The types of biological research possible with these data are demonstrated in our case study as well as recent research (Belotti et al., 2023; Deng et al., 2023) that quantify long-term phenological patterns of aerial insectivores and perform long-term analyses of the persistence of roosts. We report on estimated human annotation cost saving of: $\sim 7\times$, model error analyses and the applicability of the system to bats that are another species of flying animals whose roosts are observable from radar data.

Materials and Methods

In this section, we give preliminaries of the radar data and the rendering process (§ 2.1). Then, we describe our standardized dataset used for system training and evaluation, the development and deployment of our proposed AI-assisted system for roost annotation (§ 2.2), the incorporation of temporal information to improve the detection model (§ 2.3), and a biology case study to

¹ Average precision is an evaluation metric commonly used in the visual object detection literature in computer vision and described in the Evaluation paragraph of Appendix B.

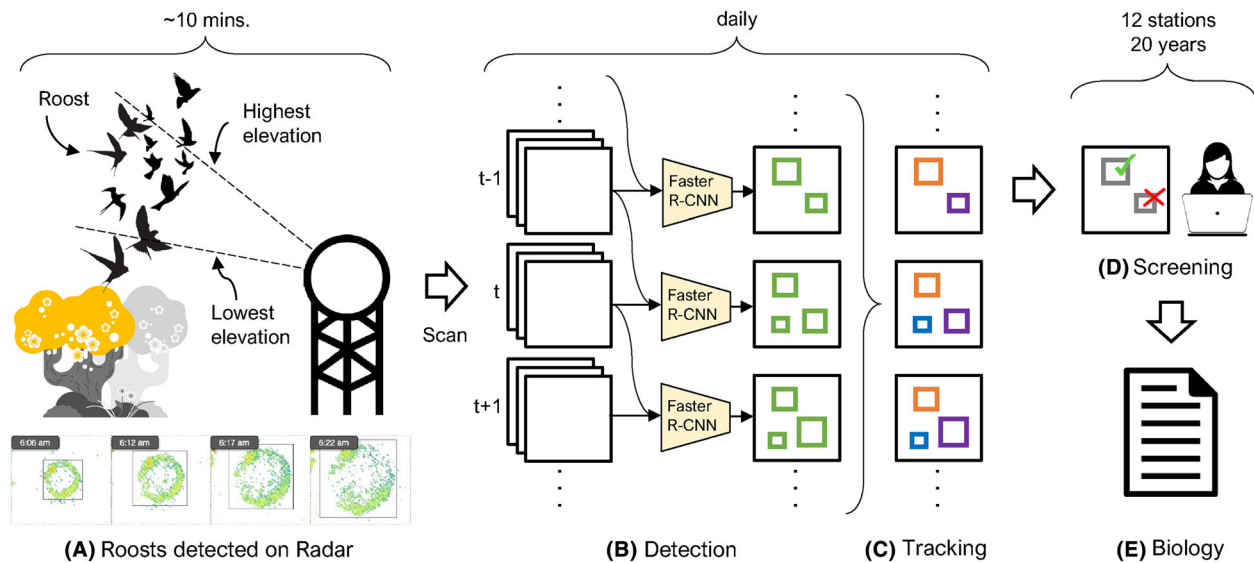


Figure 1. The proposed AI-assisted system. (A) The exodus of birds (e.g., swallows and martins) from nightly roosting locations are visible on weather radars as expanding rings in consecutive radar scans, particularly at the lowest elevations. (B and C) Our system detects individual roosts in each scan (shown as green boxes) using a spatiotemporal object detection model and organizes these into tracks corresponding to individual roosts every day (boxes of the same color correspond to the same track). (D) We have developed a web-based interface to enable quick screening of tracks, allowing us to analyze 20 years of data across 12 radar stations. (E) The screened data can provide insights about population trends and phenology of the tracked birds.

demonstrate the usefulness of our collected annotations (§ 2.4).

Radar data

Preliminaries

The Weather Surveillance Radar-1988 Doppler (WSR-88D, also called NEXRAD) network operated by the U.S. National Weather Service contains 143 stations in the contiguous U.S. and 16 stations in Alaska, Hawaii, and other U.S. territories. It has archived radar data products since the 1990s. The products are collected through *radar volume scans* (“scans”), each consisting of *sweeps* that collect various radar moments at several elevation angles (Fig. 4A top). In each sweep, the radar antenna rotates 360° around the vertical axis to sample cone-shaped “slices” of the surrounding airspace. The scans happen every 4–10 min.

Conventionally, WSR-88D radars collect 3 radar moments at various elevation angles. *Reflectivity factor* measures the density of objects in the atmosphere. *Radial velocity* measures the speed at which objects are moving relative to the radar station using the Doppler shift of the reflected radio waves. Reflectivity-weighted mean and standard deviation of radial velocity are collected as the *radial velocity* and *spectrum width* radar moments. In this

paper, our detectors are trained using subsets of these products. Between 2011 and 2013, the radar stations were upgraded to also collect 3 dual polarization (“dual-pol”) radar moments (Stepanian, 2015; Stepanian et al., 2016): *differential reflectivity*, *differential phase*, and *correlation coefficient*. These products result from both horizontal and vertical radar waves and can effectively be used to identify rain (Cheng et al., 2020; Dokter, Desmet, et al., 2018; Stepanian et al., 2016; Zrnić & Ryzhkov, 1998). We use dual-pol products, whenever available, to reduce false roost detections caused by rain during deployment.

Rendering

WSR-88D radar sweeps produce two-dimensional arrays in polar coordinates indexed by range and azimuth (antenna pointing direction in the horizontal plane), with fixed antenna elevation angle. We use nearest neighbor interpolation (Parker et al., 1983) to resample the 300 × 300 km region centered at the radar station in each sweep onto a fixed 600 × 600 Cartesian grid. Each pixel corresponds to 500 m. This rendering is known as a *plan position indicator* and corresponds to a top-down view of the cone shown at the top of Figure 4A. Each radar volume scan is rendered as an array of the shape “number of radar moments × number of elevations × 600 × 600.”

AI-assisted system: development and deployment

We release our standardized dataset v0.1.0 with roost annotations² and scripts and the library³ to develop and deploy machine learning models for automatically recognizing roosts. These reusable resources can continue to be enriched and collect more annotated data, now that hundreds of radar stations are performing new scans every few minutes.

Roost dataset preparation and release

We obtain the same training, validation, and testing splits of 88 972 radar scans used by (Cheng et al., 2020) and convert them into the COCO (Lin et al., 2014) format that is commonly adopted for computer vision datasets. After removing the scans with rendering errors, the three splits have 53 266, 11 599, and 23 587 scans, respectively, and 88 452 scans in total. These radar scans were manually annotated by different annotators for prior ecological research (Laughlin et al., 2016). The training, validation, and testing splits have 37 619, 5139, and 10 942 roost labels. Each label records the position and radius of a circle that approximates the roost. We convert the circle labels into their bounding boxes. Since the annotators have different annotation styles, Cheng et al. (2020) propose a latent-variable model and an expectation-maximization algorithm (Dempster et al., 1977) to jointly learn a detection model and scaling factors specific to annotators. We adopt their learned factors to scale and standardize the annotations in our dataset.

Roost detector training and configuration

We train a Faster R-CNN detector with a ResNet101-FPN backbone and 45 anchors ranging from 16×16 to 512×512. The backbone is pretrained on ImageNet classification and MS-COCO detection. The detector takes three channels as input, for which we selected reflectivity at 0.5° and 1.5° and radial velocity at 0.5°. We scale the three-channel input to 1200×1200 since Faster R-CNN that we use performs better on images of this size. The deployed detector achieves a 48.74 average precision at intersection-over-union threshold 0.5 ($AP^{IoU=0.5}$) on the test set described in [AI-Assisted System: Development and Deployment](#). During deployment, we keep the 100 top-scoring roost detections with predicted probability

scores of at least 0.05 in each scan for further processing. We select this low score threshold to ensure high recall, so that the predictions capture almost all roosts and could be later screened to remove false positives, see Appendix B for preliminaries about object detection with neural network models and Appendix E for ablation studies that support our design choices.

Tracking

We follow (Cheng et al., 2020) and employ a simple greedy heuristic to associate single-frame detections into tracks (Ren, 2008). We start with high-scoring detections and add unmatched detections in neighboring frames with high overlap. We apply the Kalman filter (Kalman, 1960) to smooth the roost tracks using a linear dynamical system for the bounding box center and radius. The linear system captures the dynamics of roost formation and expansion with parameters estimated from the ground truth annotations (e.g., the rate of expansion of roost bounding boxes).

Post-processing

As an optional step, when additional information about rain and wind farms are available, we apply a post-processing step to identify tracks with false-positive detections due to these sources. We consider a detection as rain and remove it if the majority of pixels in its bounding box has co-polar cross-correlation coefficient $\rho_{HV} > 0.95$, following the common rule of identifying rain (Dokter, Desmet, et al., 2018). We check whether the detections correspond to wind farms using known turbine locations from the U.S. Wind Turbine Database (Hoen et al., 2019), and if so, mark them as false positives.

Deployment

We deployed the above described models to collect information about swallow and martin roosts using data from 12 radar stations in the Great Lakes region (See Fig. 2). The radar stations include KAPX, KBUF, KCLE, KDLH, KDTX, KGRB, KGRR, KIWX, KLOT, KMKX, KMQT, and KTYX. We use this deployment experiment to verify the effectiveness of the AI system for studying roosts and as a baseline for further improving the AI system.

Swallow and martin roosts are usually detected by radars during their early morning dispersal, from mid-June to mid-September, after they finished breeding and before starting their southwards migration. Previous studies suggest that radar signatures found during this time period in North America are most likely the result

²<https://github.com/darkecology/roost-dataset>.

³<https://github.com/darkecology/roost-system/tree/fd2530fa8dba59a976da815c43dff9da8ad8e09d>.

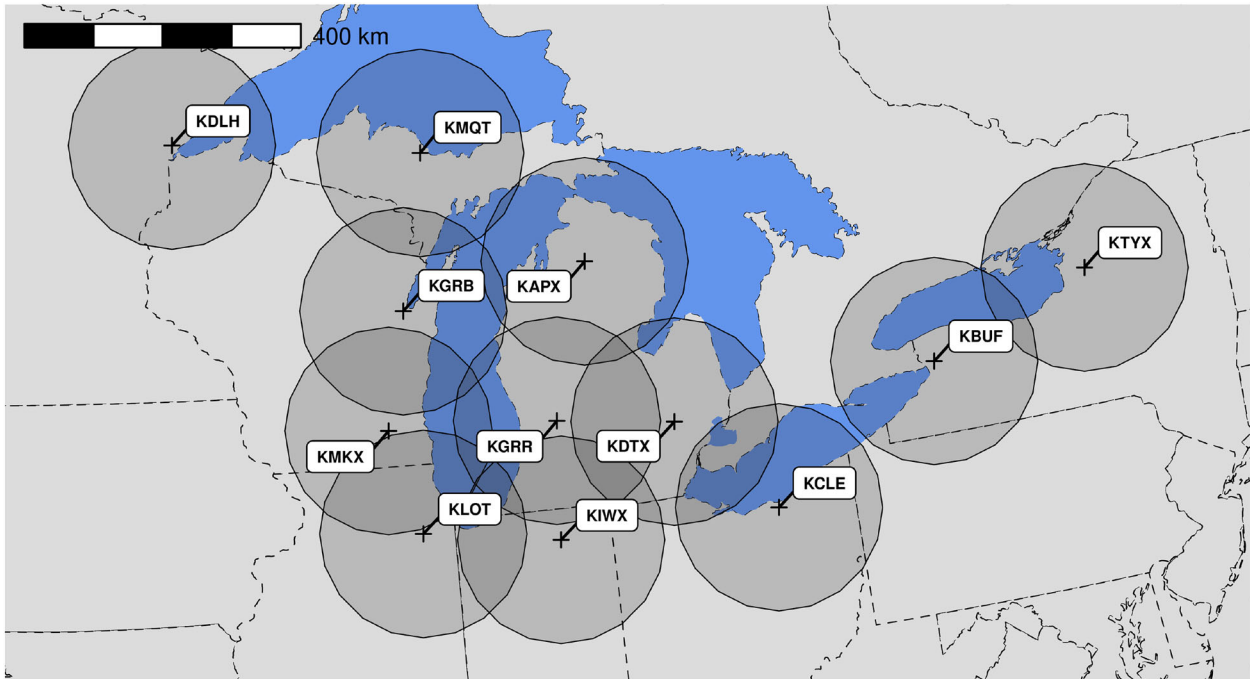


Figure 2. A map for 12 WSR-88D radar stations around the Great Lakes, which are a large group of five interconnected fresh-water lakes in the mid-east of North America.

of a Hirundinidae roost dispersal Bridge et al. (2015); Kelly and Pletschet (2017). We thus downloaded scans from 30 min before to 90 min after local sunrise between June 1 and October 31 of the 21 years ranging from 2000 to 2020. Each station's local sunrise times is obtained using the PyEphem (Rhodes, 2011) and pytz Python packages. In each 120-min window, we set 41 reference times spaced 3 min apart; we selected the scan closest to each reference time to render (Radar Data) and process, or none if there were no scans within 3 min of the time.

User interface and screening

We built a web-based interface to display the remaining tracks together with the underlying radar imagery for further screening by humans (Fig. 3). Any track with at least 2 detections, an average detection score of at least 0.15, and at least 1 detection with a score at least 0.5 was given the initial label of *roost*, considered “high-confidence tracks,” and displayed with full opacity; other tracks were given the initial label of *non-roost*, considered “low-confidence tracks”, and displayed faintly with low opacity. Our biologist teammates screened the system predictions, classifying predicted tracks into 7 categories. We removed from the analysis scans that had more than half of its pixels filled with weather or with the effects of anomalous propagation (which occurs when the radar

beam bends toward the ground and captures ground clutter). In these scans, we would not be able to identify roosts if they were present. In other scans, clear roosts were labeled as *roost*; roosts contaminated by weather, anomalous propagation, and unknown noise were labeled as *weather-roost*, *ap-roost*, and *unknown-noise-roost*; duplicated or incomplete tracks were labeled as *duplicate* and *bad-track*.

Roost detection with temporal information

Many studies (Ribani & Marengoni, 2019; Yosinski et al., 2014; Zhuang et al., 2019) have shown that pre-trained networks lead to faster convergence and greater robustness to hyperparameter settings when training object detectors. However, to use networks pretrained on three-channel (RGB) images, we must account for the domain shift and the different number of channels available due to various radar products, elevation angles, and sweeps from temporally adjacent scans. We propose a learnable adaptor, which maps an arbitrary k -channel array $x \in \mathbb{R}^{k \times n \times m}$ to a 3-channel image (see Fig. 4). This is implemented as a single convolutional layer with three filters of size $k \times 1 \times 1$, each of which implements a learned linear mapping from \mathbb{R}^k to \mathbb{R}^1 . While other choices are possible, including a nonlinear adaptor and replicating filters in the first layer of the network to match the shape

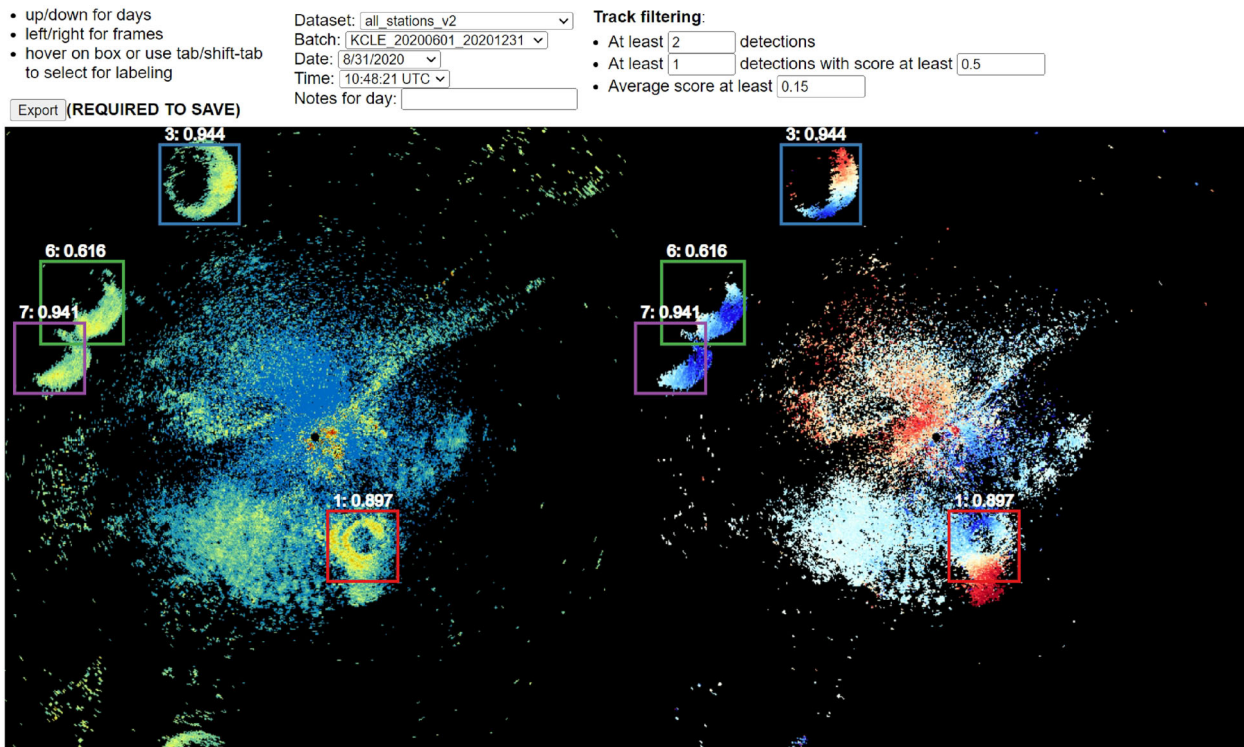


Figure 3. User interface for screening. For each radar scan, reflectivity and velocity at elevation angle 0.5° are displayed side by side. A human annotator can use direction arrows to change the scan and the date and tab and shift-tab to change the bounding box to annotate. Each system predicted bounding box can be verified as a particular type of roost or non-roost.

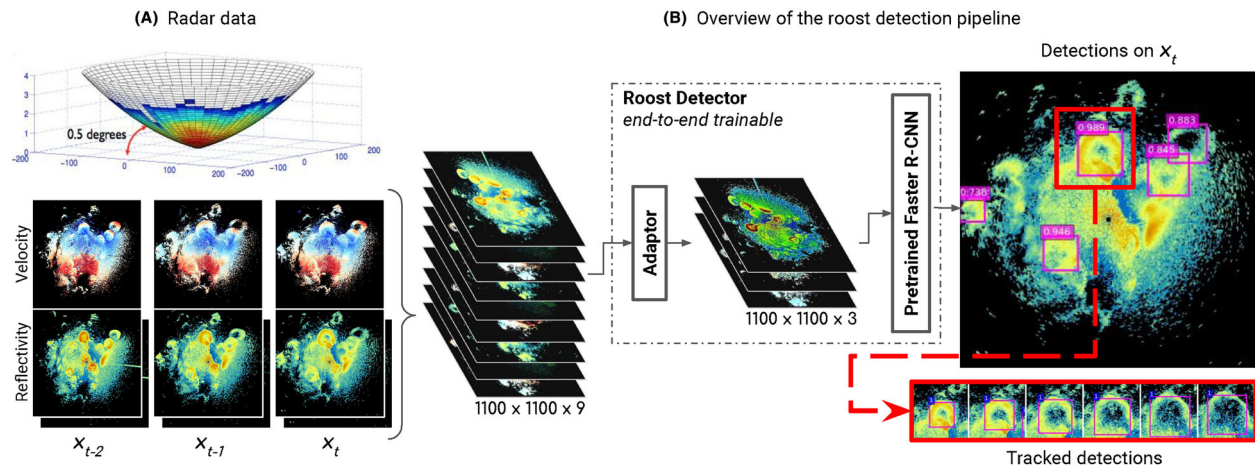


Figure 4. (A) Radar data. We show here a sweep at an elevation of 0.5°. We render top-down view of reflectivity channels at 0.5° and 1.5° and the velocity channel at 0.5° from 3 consecutive time frames x_{t-2} , x_{t-1} , x_t . (B) Overview of the roost detection pipeline. Radar images are concatenated into a single 9-channel image. The adaptor maps 9 channels to 3 channels to match a conventional RGB image for input into a deep network pretrained on color images. A faster R-CNN is trained to detect communal bird roosts for frame x_t . Finally, we associate detections from individual frames into tracks.

of an input with more channels, our preliminary study suggested that linear adaptors are the most effective, see Perez and Maji (2022) for a study on how the architecture of adaptor affects transfer learning. Also, see Section E.2 and Table E.5b for an evaluation of the benefits of pretraining.

Biology case studies

To demonstrate the types of ecological questions that could be answered with our machine learning pipeline, we selected a dense subset of tracks found in the Walpole Island First Nation Reserve. This region is within the ranges of both the KDTX and KCLE stations, which could result in the same roost being detected twice. For the purposes of this paper, we selected tracks captured only by KCLE. After manually screening the detections to remove contaminants, as described above, we used the bounding boxes estimated by the model to extract the raw reflectivity factor values from Level II radar data in polar coordinates (the location of each voxel or sampling volume is defined by its range and azimuth) across all elevation angles of each full scan.

The number of birds within each bounding box can be estimated using the approach established by Chilson et al. (2012). This method relies on two assumptions: (1) that each bounding box only contains biological scatterers and (2) that the birds are uniformly distributed within each radar sampling volume. For our analysis, we filtered only roosts labeled as clear roosts (593 tracks from our case study region), thus minimizing the chance that the two assumptions were violated. We convert equivalent reflectivity factor Z_{e} , originally in dBZ, to linear scale ($\text{mm}^6 \text{mm}^{-3}$), and then transform it into reflectivity (η), which we interpret as the density of scatterers in the atmosphere (in $\text{cm}^2 \text{km}^{-3}$). We can then multiply the reflectivity measurement at each azimuth and interval by the theoretical volume sampled by the radar at that range. Finally, we divide the result by the specific radar cross section (RCS) of the bird species assumed to be found in the roost, thus obtaining an estimate of the number of birds.

To obtain conservative estimates, we used Purple Martins as our benchmark average radar cross section, since they are the largest species of Hirundinidae found in North America. The average RCS of Purple Martins can be obtained from their mass – 51 g, see (Dunning, 2008) – adopting the relationship between mass and RCS proposed by Horton *et al.* (Horton et al., 2019): $\log(\text{RCS}) = 0.699 \times \log(\text{mass})$. Bank Swallows (*Riparia riparia*), possibly the smallest species to participate in such aggregations, would yield count estimates approximately 2.6 times higher, since their mass is 13 g according to Dunning (2008). The volume sampled by the radar is assumed to be shaped like a truncated cone with axis

aligned with the antenna's peak power axis, cut by two parallel planes at each range gate. For data before the Super Resolution upgrade, the cone's apex angle was assumed to be 1° . After the 2007–2008 upgrade, we assumed an elliptical cone of 1° vertical beam width and 0.5° horizontal beam width (Torres, 2007).

We extracted the number of birds from all sweeps available in each radar scan. In a post-processing stage, we filtered the sweeps that had height lower than 5000 m within the 150 km radius of each station. In order to capture the entire 3D structure of the roost departure, grouped the sweeps in bins of 1° interval according to their elevation angle, and we calculated the mean count within each bin. Finally, we summed the estimates from each elevation bin to get the bird count of each roost. This procedure aims to avoid double counting birds due to sweeps where the radar beam of 1° beamwidth would sample the same area of the airspace twice (that is, if the variation in elevation between two consecutive sweeps is lower than 1°).

To obtain a single estimate per roost dispersal, we further summarized the bird counts across detections by taking the mean estimate per track. We then explored phenological trends at the spatial scale of this roost, which consistently occurs from 2000 to 2020. Data were considered as missing on days when scans had more than 50% weather contamination, intense anomalous propagation, or when the sampling window was shorter than 100 min. Days without detections were considered as true zeros. To calculate the daily bird count within the roost for each day, we derived the maximum number of birds for each roost track. We then fit a generalized additive model (GAM) to each roost-year to model the roosting activity throughout a roosting season. We constructed GAMs with daily estimates of the number of birds as the response variable and ordinal date as the independent variable with the smoothing parameter k set to 5 using a quasi-Poisson distribution. We used this model construction to predict estimates throughout the season and selected the 50% passage date, i.e., the first date in which the cumulative predicted estimates of number of birds exceeded half of the yearly total, as our phenology estimate for that year.

Results

We report the results of our deployed AI system (§ 3.1), the improvements of our detection model with temporal information (§ 3.2), and biological case studies (§ 3.3).

System deployment

Here we report the results of the deployment experiment described in [AI-Assisted System: Development and Deployment](#). Table 1 shows statistics for the results of the

Table 1. Statistics for rendered scans, system predictions, and time needed to screen the predictions. System-predicted detections and tracks are either of high or of low confidence and displayed as *roost* or *non-roost* by default in the screening interface; see [AI-Assisted System: Development and Deployment](#) for details. The bounding boxes are shown with high and low opacity, respectively, to ease screening.

Station	Years	Scans	Detections		Tracks		Screening time
			High conf	Low conf	High conf	Low conf	
KAPX	21	51 390	3124	16 125	738	9502	8h57m
KBUF	21	55 402	15 433	42 754	3172	26 092	17h09m
KCLE	21	56 197	22 459	42 774	4792	25 735	20h55m
KDLH	20	49 599	4817	18 483	1129	11 617	9h45m
KDTX	21	51 024	19 349	35 309	3745	19 707	18h45m
KGRB	20	47 249	14 106	35 806	3259	20 599	18h54m
KGRR	21	52 060	10 463	29 320	2538	18 624	16h53m
KIWX	21	53 534	16 447	45 410	4138	29 832	20h15m
KLOT	21	52 761	9113	41 442	2462	28 083	16h29m
KMKX	21	50 476	14 957	42 049	3356	26 047	19h51m
KMQT	21	49 611	1737	9766	369	6042	5h41m
KTYX	17	43 483	8031	13 356	1615	8208	10h02m
Total	246	612 786	140 036	372 594	31 313	230 088	183h36m

automated steps. There were six station-years for which data were not available or could not be rendered: the year 2000 for KDLH and KGRB and the years 2001–2004 for KTYX. For the remaining station-years, 612 786 scans were successfully rendered in total. The automated annotation steps predicted 31 313 “high-confidence tracks” assembled from 140 036 single-scan detections and 230 088 “low-confidence tracks” with 372 594 single-scans detections.

Table 2 shows the statistics for human-screened results. After human screening of machine predictions, we identified 13 860 clean roost tracks, 477 roost tracks with minor weather contamination, 100 roost tracks with minor

anomalous propagation, and 1191 with minor unknown noise. These four categories together produced 15 628 roost tracks (64 620 detections) that can be used for ornithology research. Other machine-predicted tracks are considered false-positive predictions.

Among the “high-confidence” machine-predicted tracks, 12 025 tracks (56 354 detections) were marked as one of the four roost categories. Among the “low-confidence” tracks that are by default displayed as “non-roost”, only 3603 tracks (8266 detections) out of 230 088 tracks (372 594 detections) needed to be marked as roosts; the remaining do not require screeners to change their labels.

Table 2. Statistics for screened data. Entries show numbers of detections/tracks assigned each of the 7 labels: clean roost (cln), weather contaminated roost (wth), anomalous propagation contaminated roost (ap), unknown noise contaminated roost (nse), duplicate (dup), bad track (bad), and non-roost track (non). The first four categories can be used for aeroecology research.

Station	cln	wth	ap	nse	dup	bad	non
	Number of detections/number of tracks						
KAPX	1040/311	14/1	18/4	22/6	5/3	32/7	18118/9908
KBUF	6669/1628	384/79	72/11	1129/156	90/30	334/45	49509/27315
KCLE	12863/3017	800/151	53/12	903/155	46/16	585/82	49983/27094
KDLH	2249/679	52/15	54/13	136/30	13/6	49/10	20747/11993
KDTX	9568/1990	729/117	81/12	578/95	74/28	659/75	42969/21135
KGRB	3616/998	94/20	51/12	1364/292	15/6	725/116	44047/22414
KGRR	4827/1340	68/13	31/9	572/138	23/7	346/44	33916/19611
KIWX	7017/1890	118/25	98/21	632/146	4/1	449/65	53539/31822
KLOT	1841/610	39/11	1/1	186/42	2/2	50/9	48436/29870
KMKX	1911/510	90/19	10/2	520/91	5/2	95/14	54375/28765
KMQT	9/7	0/0	0/0	0/0	0/0	0/0	11494/6404
KTYX	3752/880	146/26	11/3	202/40	24/11	1116/147	16136/8716
Total	55362/13860	2534/477	480/100	6244/1191	301/112	4440/614	443269/245047

Savings in human labeling efforts

Our AI system reduces human labor to screening system predictions and enables tractable annotation to obtain high-quality roost data. The screening amounts to 183.6 annotator hours and an average of 17.6 seconds per radar-station day (see Table 1). If the annotators were to find roosts, draw bounding boxes for all roosts, and assemble them into tracks completely manually, the annotation process would be significantly more time consuming and error prone. For instance, we estimate a reduction from 31.4 weeks to 4.6 weeks of full-time work at 40 hours/week of total manual annotation (1254.6 \rightarrow 183.6 h, a 6.8 \times saving) considering the same amount of screened days (from June to October) and the 246 season-years from the Great Lakes (See Table 1). We calculate the total time of the annotation process from scratch (i.e., 1254.6 h) using a measured annotation time of 1 hour per station-month (120 seconds per day) and extrapolating to the complete screening period of our deployed system (153 days of 246 station-years).

Roost detection using temporal information

A significant technical advance was to incorporate temporal information from past frames to capture roost dynamics. When viewed in a size-constrained window rain can take many different shapes, including the appearance of a ring-shaped roost. However, the movement of weather is different from roosts – rain often moves in a straight trajectory, while roosts diverge from a point. Temporal information reduces false positives due to small patches of rain with a roost-like appearance (see Fig. F.3 in Appendix F) and improves the detector performance from 48% to 56% $AP^{IoU} = .50$.

The best-performing temporal model overall included the adaptor layer and used as input 3 consecutive time frames, x_{t-2} , x_{t-1} , and x_t , with three input channels per frame: reflectivity at 0.5°, reflectivity at 1.5°, and radial velocity at 0.5° (see Fig. 4). These 9 total input channels were rendered at size 1100 \times 1100 as the inputs to a Faster R-CNN detection model with linear adaptor and a pretrained ResNet101-FPN backbone, then trained for more than 40 k iterations with a batch size of 4 samples, see Appendix E for ablation of the roost detector experiments and Appendix F for qualitative results.

Further savings in human efforts

We perform the manual screening of the station-year KTYX 2020 as a pilot study to calculate the time savings when using the predictions of the temporal detector. We were able to reduce the time needed to extract phenology information from 1.17 h to 30 min, suggesting a further labeling effort reduction of 2.3 \times with the inclusion of temporal information to the pipeline.

Biology case studies results

Figure 5A shows the locations of the first detection in each of the 702 tracks, colored by year, in the study area at Walpole Island around Lake Saint Claire. Martins and swallows gathered in this region to roost every year from 2000 to 2021, with an average of approximately 33 days of detections per year. The yearly maximum number of days when birds were detected occurred in 2019, when birds congregated around the lake on 42 days between July 23 and September 8.

The peak timing (50% passage date) for roosting activity in the study area was 20 August 2000, and 21 August

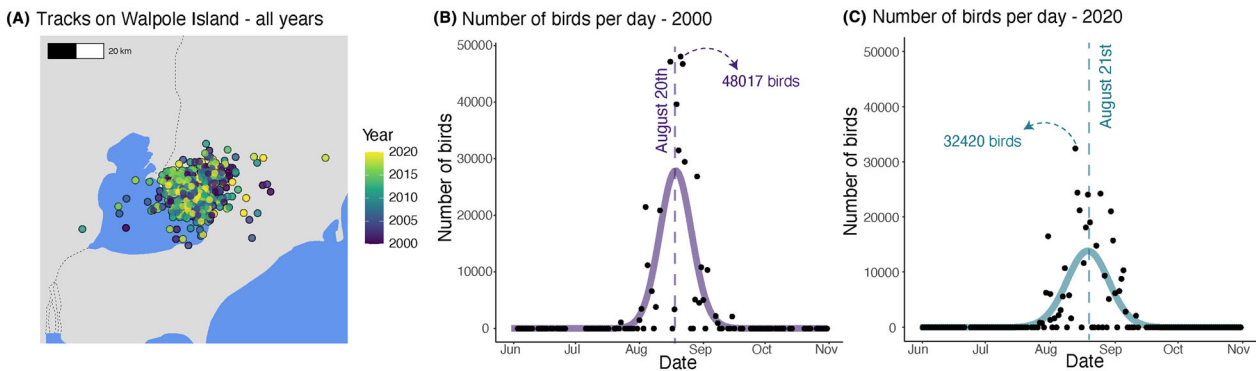


Figure 5. Case study: roosts at the Walpole Island detected by the KCLE station. (A) Map where points, colored by year, represent the first detection of each track captured by the KCLE radar station. Dispersals over water are from roosts in the margins of the lake. (B) Raw estimates of number of birds (represented as points) and the GAM prediction (continuous line) of roosting activity from June to October, 2000. (C) Raw estimates of number of birds (represented as points) and the GAM prediction (continuous line) of roosting activity from June to October 2020.

2020 (Fig. 5B and C). The average peak number of birds per year detected in the region throughout our study period was 73 925 birds (SD = 37 743). The year when the roost received the highest number of birds occurred on 2010, when our estimates of number of birds reached 156 864 on August 8th. In contrast, the year with lowest peak estimate was 2013, when we detected at most 23 949 birds on August 9.

Discussion

False positives due to noisy evaluation data

Future research can focus on improving an AI system through careful error analysis. Figure 6B shows the distribution of top-scoring false positives (detection with a score of 0.5 or higher) for the temporal detector. Localization errors ($0 < \text{IoU} < 0.5$) account for nearly 44% of the false positives, while detections on background regions ($\text{IoU} = 0$), frequently due to weather, static structures, or anomalous propagation, account for the remaining 56%. Figure 6A shows the precision-recall curve by varying the overlap threshold at which a detection is considered a true positive. By reducing the IoU, the performance of the detector improves from 56.3% $\text{AP}^{\text{IoU}=0.50}$ to

66.4% $\text{AP}^{\text{IoU}=0.40}$ to 73.5% $\text{AP}^{\text{IoU}=0.20}$, while still producing useful roost detections. See examples of localization errors in Figure 7A–C, missed roosts in Figure 7D, and detections on background regions in Figure 7E.

Another source of false positives are missing roost annotations. We perform a manual inspection of all non-overlapping false positives ($\text{IoU} = 0$ with any annotations) at a recall of 20% and find 94% of these to be actual roosts, thus increasing the precision of our detector from 88% to 92%.

Sources of false negatives

We found several sources of false negatives, which are roosts that were annotated by researchers who originally created our training dataset (Laughlin et al., 2016) but missed by our system. The first source is radar scans that contained roosts together with large amounts of other noise such as anomalous propagation, ground clutter, or other biological scatterers. Figure 8A shows an example of three consecutive frames with a single annotated roost (green) amidst anomalous propagation and ground clutter with reflectivity values so high that the model is unable to discriminate the roost from the background. In this case, the human labeler likely relied on contextual

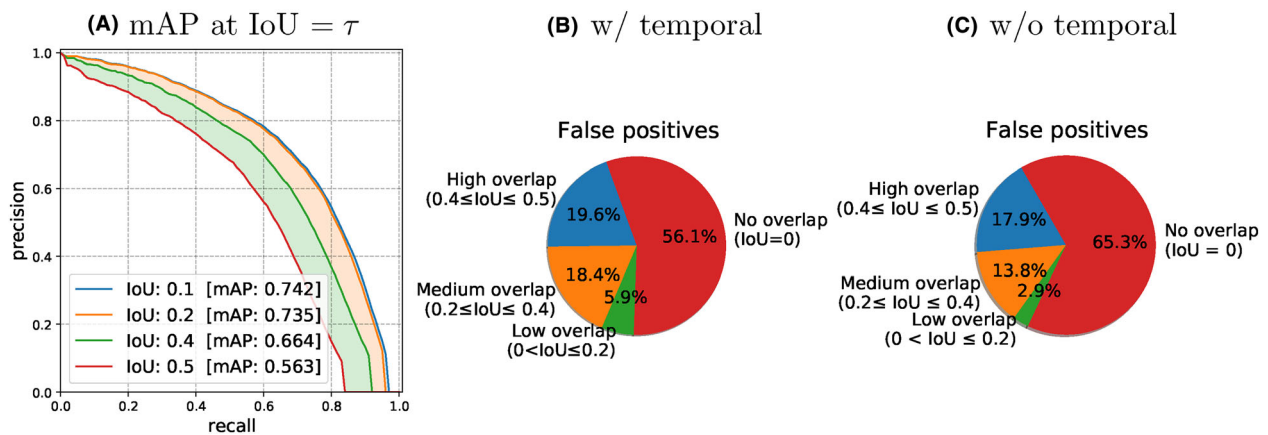
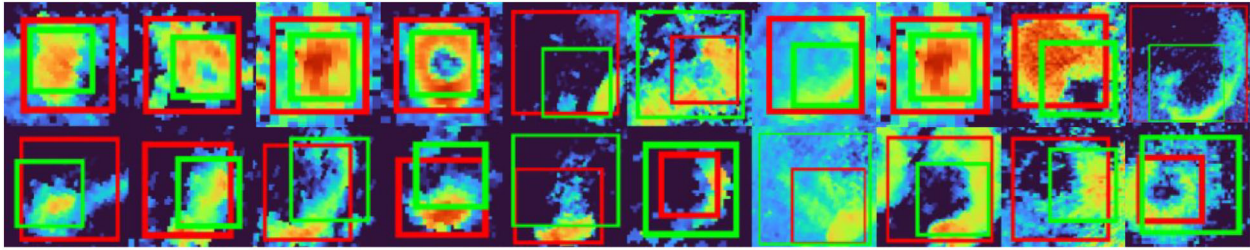
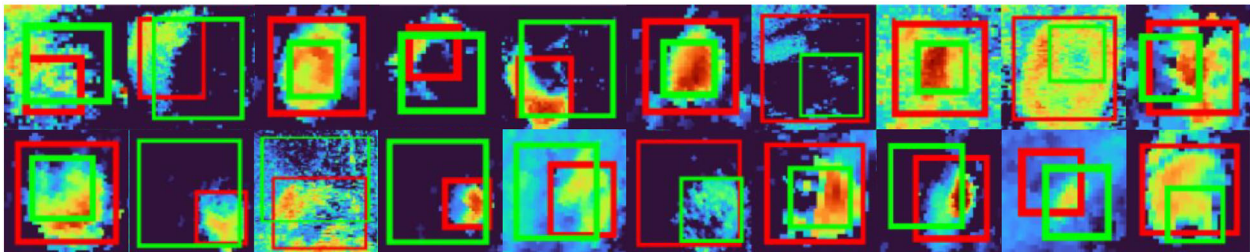


Figure 6. (A) Precision-recall curves at different IoU thresholds (higher values reflect a stricter requirement). The mAP improves from 56.3% to 66.4% when a threshold $\text{IoU} > 0.4$ is used compared to $\text{IoU} > 0.5$ suggesting that a large number of top-scoring detections are “near misses.” (B) The distribution of top scoring false positives for the model with temporal information. Localization errors ($\text{IoU} > 0$) account for nearly 44% of the data, while detections on background regions, frequently due to weather, account for the remaining 56.1%. (C) The distribution of top scoring false positives without using temporal information shows a greater fraction of false positives on the background regions.

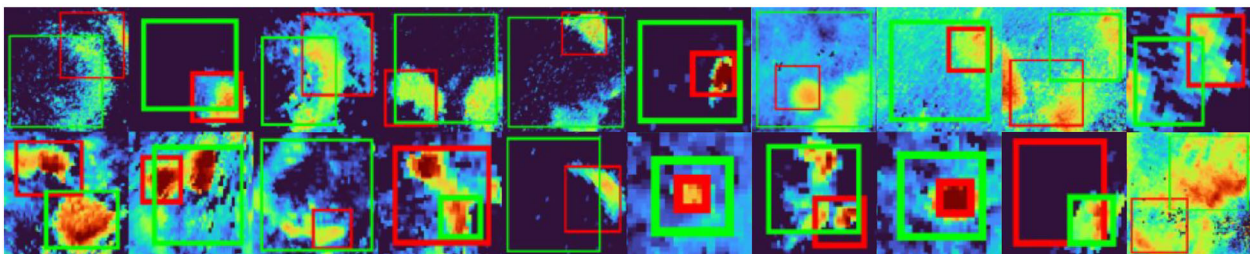
Figure 7. Detection errors. A detection is considered false positive if the overlap (IoU) between the annotation and prediction bounding boxes is ≤ 0.5 . (A–C) Detection with different overlap values below the threshold. (D) Human-annotated roosts that the model does not detect. (E) High-scoring detections on rain show a similar morphology as roosts in a single frame, but the temporal model is able to reduce many of these false positives. We show human annotations in green and predictions by the model in red.



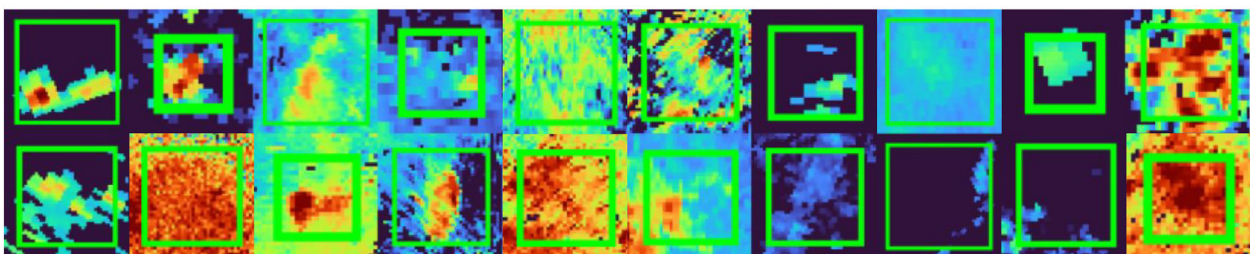
(A) High-overlap localization errors ($0.4 \leq \text{IoU} \leq 0.5$).



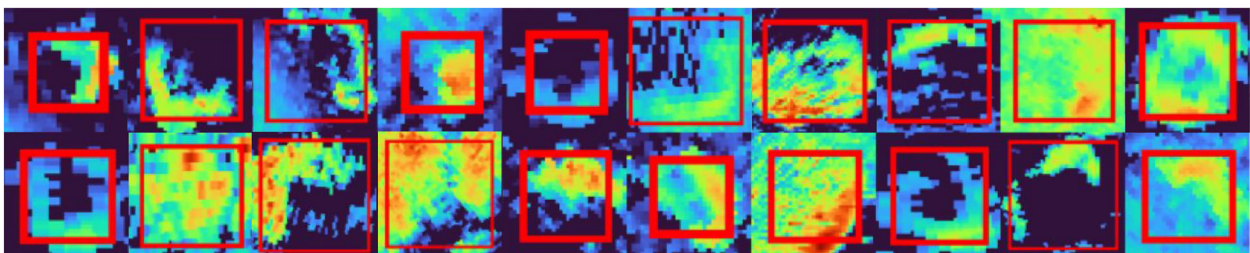
(B) Medium-overlap localization errors ($0.2 \leq \text{IoU} \leq 0.4$).



(C) Low-overlap localization errors ($0 < \text{IoU} \leq 0.2$).



(D) Roosts with no overlap with a high-scoring detection (false negatives).



(E) High-scoring detections with no overlap with an annotated roost (false positives).

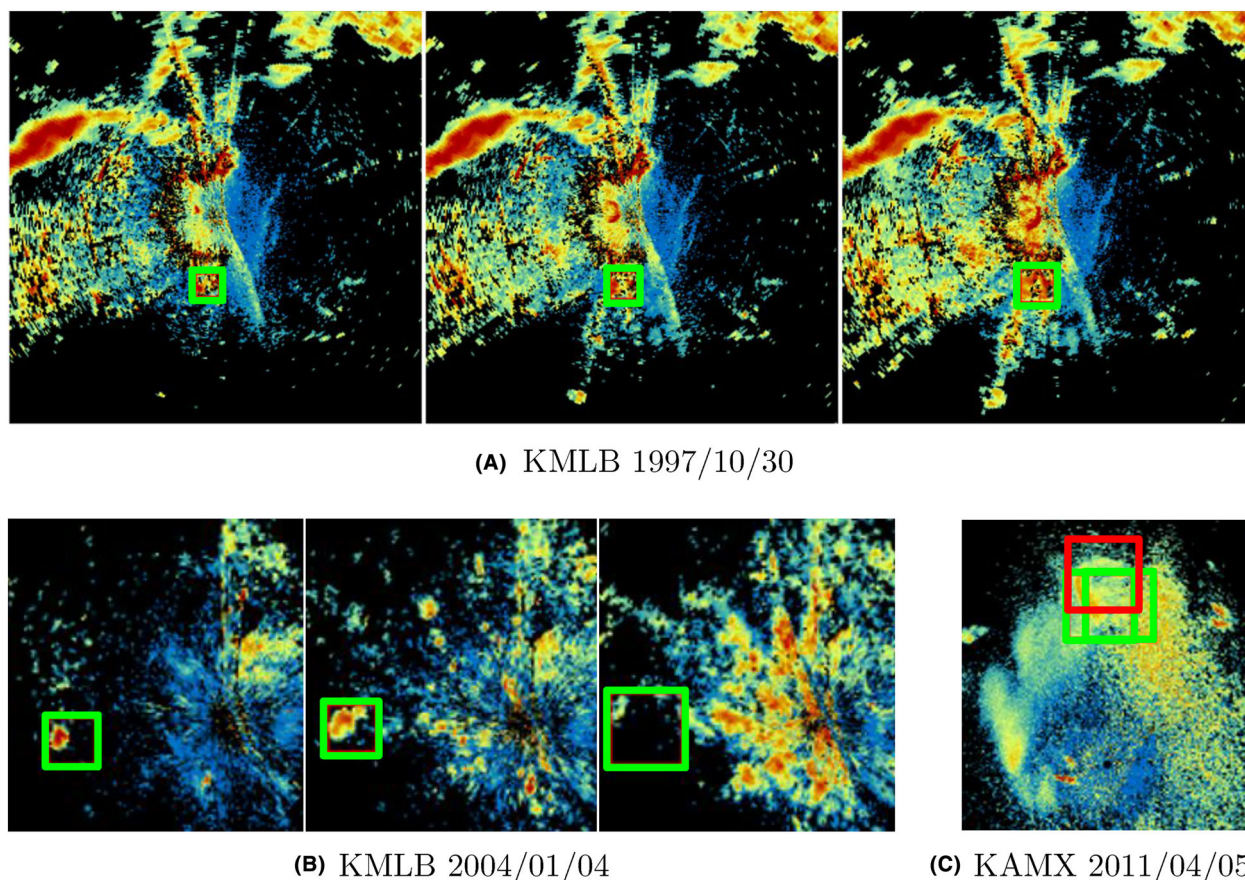


Figure 8. False-negative sources. (A) Three consecutive radar scans with a large amount of anomalous propagation and other biology. (B) Three consecutive scans with a roost that lacks a well-defined ring shape. (C) Single scan with two roosts overlapping – the model only detects one. We show human annotations in green and predictions if produced by the model in red. Images correspond to the reflectivity at 0.5° elevation.

information such as the presence of roosts in the same location across many frames and days (see more examples of noisy roosts in Figure 7D). Also note that the model successfully detects roosts in radar scans with anomalous propagation when the reflectivity values are not so high (see examples in Fig. F.2a in the Appendix F).

Another source of false negatives is roosts that lacked the prototypical ring shape. Figure 8B shows annotations for a non-ring-shaped roost in three consecutive frames. These are encountered rarely in the dataset and the morphology can easily be confused with weather and other biology (see more examples of roosts without ring shape in Fig. 7D). Figure F.2c shows an example where the model *is* able to detect a roost without the usual ring shape; however, there is a higher correspondence between consecutive frames and a more consistent diverging pattern compared to the example in Figure 8B. Detection of non-ring-shaped roosts could be complicated by the fact that annotators were required to use a circle to annotate roosts, and often adopted different labeling styles,

especially for how to label a roost that was not ring shaped (Cheng et al., 2020).

One more source of false negatives is bird roosts that appear too close to each other in the radar scan. As shown in Figure 8C, the model detects only one of the two annotated roosts. Again, the human annotators may have used complex contextual information to interpret these overlapping patterns as two different roosts.

Future research for AI-assisted roost detection

There are a number of future directions to improve an AI-assisted system for roost measurement. Iterative refinements to an existing model can often improve performance significantly over time, especially when driven by a careful error analysis. Adding new sources of information, including new input features, training signals, or training labels, together with data curation efforts to improve the quality of training data, are among the most promising ways to

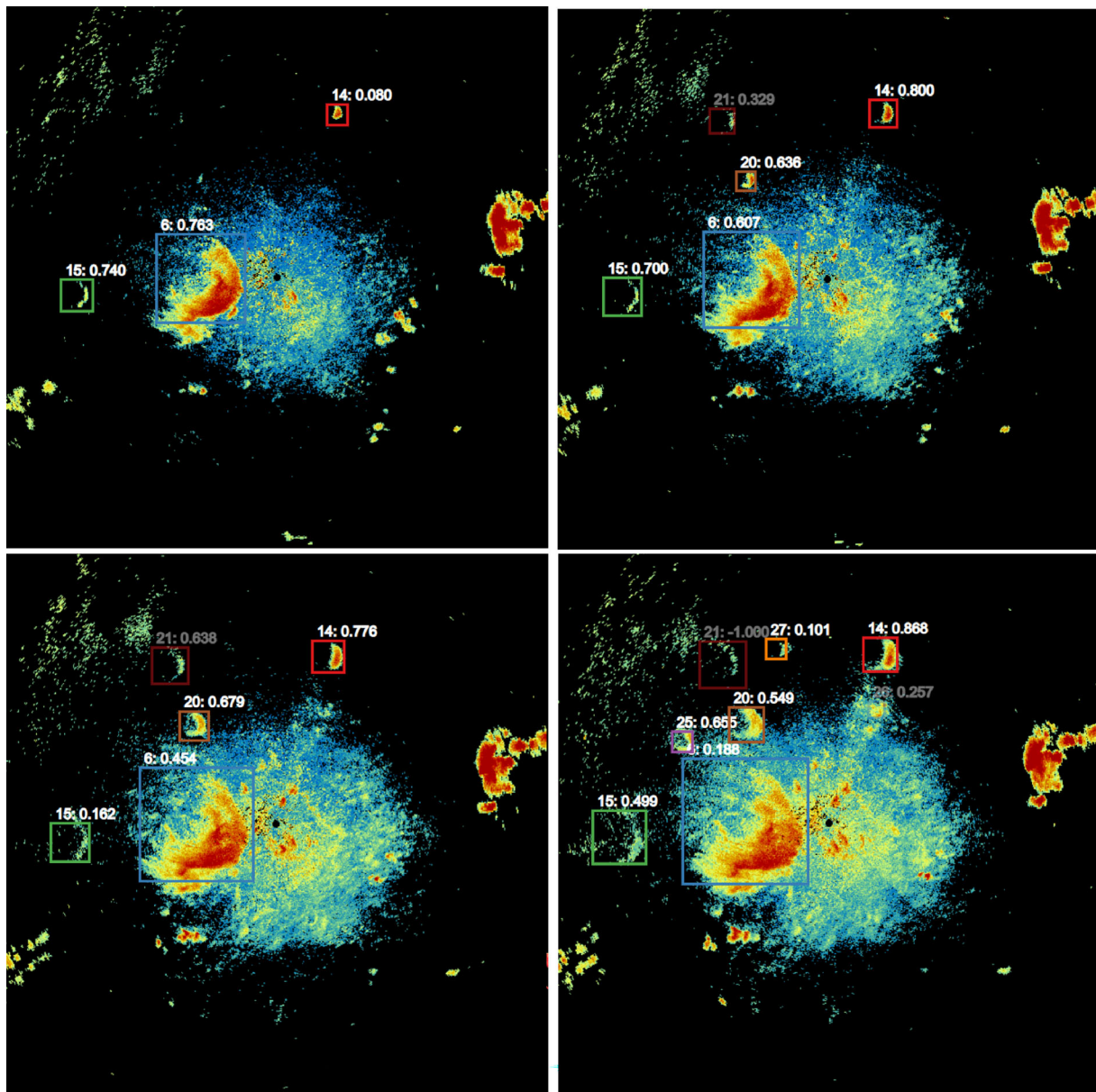


Figure 9. System detection and tracking of bat roosts from the KEWX radar station on 2012/6/30. The scans are temporally consecutive in the order of top-left, top-right, bottom-left, and bottom-right.

improve performance. In the specific context of our work, adding the results of human-screened predictions back into the model as training data may provide a significant boost in model performance to scale our model to larger geographical regions with less human effort.

Future research can incorporate additional sources of information related to the contextual cues that humans use to detect roosts and discriminate them from other patterns in radar images. One such cue is the persistence of roosts in the same or similar locations across days and years, which may allow humans to confidently detect

roosts in noisy radar scans including high-reflectivity anomalous propagation, weather, or other biological scatterers. A model could be given input features from radar scans in previous days or years, similar to the way we added within-day temporal information in this paper; or, it could be given historical *detections* of a simpler roost detection model (Zhou et al., 2020). We found that rain is a persistent source of false positives, even though humans have a relatively easy time distinguishing rain from roosts using the full context of a radar image sequence, and AI models can successfully discriminate

rain from broad-scale bird migration (Lin et al., 2019). There may be a mismatch between the “local” scale of an object detection model and the broader scope of contextual information needed to discriminate rain. A possible remedy is to train a multihed model to jointly perform roost detection together with another task such as rain segmentation (He et al., 2017; Kirillov et al., 2018; Li et al., 2022; Shen et al., 2021), or provide the predictions of a rain segmentation model (Lin et al., 2019) as input to the roost detection model.

Ultimately, the goal is to collect roost measurements from a very large but finite set of images. Future research should focus on effective ways to combine human effort, computational effort, AI training, and statistical estimation to achieve the desired scientific outcomes. For example, what is the tolerance of measurements of swallow phenology or population declines to an AI system with a certain performance level? Is human screening of outputs required? What are the most effective strategies for interleaving human annotation and model training to analyze a very large image data set? An interesting research direction is to pose the scientific question (e.g., how many roosts) as a statistical estimation problem and to consider statistical estimators that can give high confidence bounds after examining only a subset of the images (Meng et al., 2021; Perez et al., 2023).

Broader directions for biological recognition

There are several broader directions in the recognition of biological patterns in radar data that can be informed by our work. Our roost-detection model can serve as a starting point for AI models to detect and track related taxa-specific biological phenomena in radar data, including bat roosts (Stepanian & Wainwright, 2018), roosts of non-swallow bird species (Van Den Broeke, 2019), mayfly hatches (Stepanian et al., 2020), and waterfowl (O’Neal et al., 2010). Of these, bat roosts are most similar to swallow roosts and may require the least adaptation to the model. As a proof of concept, we deploy our best model found in Section E.2 on bat roosts around the KEWX radar station in Texas. We process scans from 90 min before local sunset to 150 min after the sunset in June 2012. We observe that our system developed using bird roost data performs reasonably well at detecting and tracking bats without any customization (Fig. 9). Roosts of other bird species including robins, blackbirds, starlings, and waterfowl are also visible on radar (Russell et al., 1998), but are often less obvious for humans to discern and usually lack the distinct “expanding ring” pattern of swallow roosts, probably due to differences in roost emergence behavior. Mayfly hatches have a distinct and rather different appearance than bird roosts. The automatic detection of these phenomena is an

interesting frontier for AI methods in radar aeroecology. Investigating whether it is possible to distinguish among roosts of different swallow species, e.g., Purple Martins and Tree Swallows, from fine-grained radar characteristics is an interesting open question that could have important biological implications. This task could be potentially accomplished, for example, by pairing radar detections from our system with records from large-scale citizen science datasets. Finally, extending these models to radar networks outside the US could provide information to track bird species beyond national borders.

Acknowledgments

This material is based upon work supported by the US National Science Foundation under Grant Nos. 2017554 to Colorado State University, 1749833 and 2017756 to the University of Massachusetts Amherst, and 2017582 to the University of Oklahoma.

Author Contributions

DS and SM conceived the study. WZ, ZC and SM created training data sets. WZ and ZC wrote code for model training, evaluation, and deployment. GP, WZ, ZC, and SM developed neural network models. GP developed final models and ran experiments. GP and SM analyzed results. WZ deployed model on Great Lakes data. DS wrote code to visualize and screen results. MB and YD designed screening protocol. MB, YD, and VS executed screening protocol. MB and YD analyzed data for biology case study. GP, DS, WZ, SM, MB, YD, and ZC wrote the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

References

- Ansari, S., Del Greco, S., Kearns, E., Brown, O., Wilkins, S., Ramamurthy, M. et al. (2018) Unlocking the potential of NEXRAD data through NOAA’s big data partnership. *Bulletin of the American Meteorological Society*, **99**, 189–204.
- Belotti, M.C.T., Deng, Y., Zhao, W., Simons, V.F., Cheng, Z., Perez, G. et al. (2023) Long-term analysis of persistence and size of swallow and martin roosts in the us great lakes. *Remote Sensing in Ecology and Conservation*, **9**, 469–482.
- Bridge, E.S., Pletschet, S.M., Fagin, T., Chilson, P.B., Horton, K.G., Broadfoot, K.R. et al. (2015) Persistence and habitat associations of purple Martin roosts quantified via weather surveillance radar. *Landscape Ecology*, **31**, 43–53.
- Bruderer, B. (1997) The study of bird migration by radar part 1: the technical basis. *Naturwissenschaften*, **84**, 1–8.
- Buler, J.J. & Dawson, D.K. (2014) Radar analysis of fall bird migration stopover sites in the northeastern US. *The Condor*, **116**, 357–370.

- Buler, J.J. & Diehl, R.H. (2009) Quantifying bird density during migratory stopover using weather surveillance radar. *IEEE Transactions on Geoscience and Remote Sensing*, **47**, 2741–2751.
- Cheng, Z., Gabriel, S., Bhambhani, P., Sheldon, D., Maji, S., Laughlin, A. et al. (2020) Detecting and tracking communal bird roosts in weather radar data. *Association for the Advancement of Artificial Intelligence*, **34**, 378–385.
- Chilson, P.B., Frick, W.F., Stepanian, P.M., Shipley, J.R., Kunz, T.H. & Kelly, J.F. (2012) Estimating animal densities in the aerosphere using weather radar: to Z or not to Z? *Ecosphere*, **3**, 1–19.
- Cohen, E.B., Horton, K.G., Marra, P.P., Clipp, H.L., Farnsworth, A., Smolinsky, J.A. et al. (2021) A place to land: spatiotemporal drivers of stopover habitat use by migrating birds. *Ecology Letters*, **24**, 38–49.
- Crum, T.D. & Alberty, R.L. (1993) The WSR-88D and the WSR-88D operational support facility. *Bulletin of the American Meteorological Society*, **74**, 1669–1687.
- Dempster, A.P., Laird, N.M. & Rubin, D.B. (1977) Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B: Methodological*, **39**, 1–38.
- Deng, Y., Belotti, M.C.T., Zhao, W., Cheng, Z., Perez, G., Tielens, E. et al. (2023) Quantifying long-term phenological patterns of aerial insectivores roosting in the great lakes region using weather surveillance radar. *Global Change Biology*, **29**, 1407–1419.
- Dokter, A.M., Desmet, P., Spaaks, J.H., van Hoey, S., Veen, L., Verlinden, L. et al. (2018) bioRad: biological analysis and visualization of weather radar data. *Ecography*, **42**, 852–860.
- Dokter, A.M., Farnsworth, A., Fink, D., Ruiz-Gutierrez, V., Hochachka, W.M., La Sorte, F.A. et al. (2018) Seasonal abundance and survival of North America's migratory avifauna determined by weather radar. *Nature Ecology & Evolution*, **2**, 1603–1609.
- Dokter, A.M., Liechti, F., Stark, H., Delobbe, L., Tabary, P. & Holleman, I. (2011) Bird migration flight altitudes studied by a network of operational weather radars. *Journal of the Royal Society Interface*, **8**, 30–43.
- Dunning, J. (2008) *CRC handbook of avian body masses*, Vol. 384, 2nd edition. Boca Raton, FL: CRC Press.
- Farnsworth, A., Van Doren, B.M., Hochachka, W.M., Sheldon, D., Winner, K., Irvine, J. et al. (2016) A characterization of autumn nocturnal migration detected by weather surveillance radars in the northeastern USA. *Ecological Applications*, **26**, 752–770.
- Fraser, K.C., Stutchbury, B.J.M., Silverio, C., Kramer, P.M., Barrow, J., Newstead, D. et al. (2012) Continent-wide tracking to determine migratory connectivity and tropical habitat associations of a declining aerial insectivore. *Proceedings of the Royal Society of London B: Biological Sciences*, **279**, 4901–4906.
- Gauthreaux, S.A., Jr. (1970) Weather radar quantification of bird migration. *Bioscience*, **20**, 17–19.
- Gauthreaux, S.A., Jr. & Belsler, C.G. (1998) Displays of bird movements on the WSR-88D: patterns and quantification. *Weather and Forecasting*, **13**, 453–464.
- Gauthreaux, S.A., Belsler, C.G. & van Blaricom, D. (2003) Using a network of WSR-88D weather surveillance radars to define patterns of bird migration at large spatial scales. In: Berthold, P., Gwinner, E. & Sonnenschein, E. (Eds.) *Avian migration*. Berlin, Heidelberg: Springer. Available from: https://doi.org/10.1007/978-3-662-05957-9_23
- He, K., Gkioxari, G., Dollár, P. & Girshick, R. (2017) Mask r-cnn. Cite arxiv:1703.06870Comment: open source; appendix on more results.
- Hoen, B., Diffendorfer, J., Rand, J., Kramer, L., Garrity, C. & Hunt, H. (2019) United States wind turbine database. US Geological Survey, American wind energy association, and lawrence berkeley national laboratory data release: Uswtodb v2. 1. United States Wind Turbine Database.
- Horn, J.W. & Kunz, T.H. (2008) Analyzing NEXRAD doppler radar images to assess nightly dispersal patterns and population trends in Brazilian free-tailed bats (*Tadarida brasiliensis*). *Integrative and Comparative Biology*, **48**, 24–39.
- Horton, K.G., Doren, B.M.V., Sorte, F.A.L., Cohen, E.B., Clipp, H.L., Buler, J.J. et al. (2019) Holding steady: little change in intensity or timing of bird migration over the Gulf of Mexico. *Global Change Biology*, **25**, 1106–1118.
- Horton, K.G., La Sorte, F.A., Sheldon, D., Lin, T.Y., Winner, K., Bernstein, G. et al. (2020) Phenology of nocturnal avian migration has shifted at the continental scale. *Nature Climate Change*, **10**, 63–68.
- Horton, K.G., Van Doren, B.M., La Sorte, F.A., Fink, D., Sheldon, D., Farnsworth, A. et al. (2018) Navigating north: how body mass and winds shape avian flight behaviours across a north American migratory flyway. *Ecology Letters*, **21**, 1055–1064.
- Kalman, R.E. (1960) A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, **82**, 35–45.
- Kelly, J.F. & Pletschet, S.M. (2017) Accuracy of swallow roost locations assigned using weather surveillance radar. *Remote Sensing in Ecology and Conservation*, **4**, 166–172.
- Kirillov, A., He, K., Girshick, R., Rother, C. & Dollár, P. (2018) *Panoptic segmentation*.
- Kunz, T.H., Gauthreaux, S.A., Jr., Hristov, N.I., Horn, J.W., Jones, G., Kalko, E.K. et al. (2008) Aeroecology: probing and modeling the aerosphere. *Integrative and Comparative Biology*, **48**, 1–11.
- Laughlin, A.J., Sheldon, D.R., Winkler, D.W. & Taylor, C.M. (2016) Quantifying non-breeding season occupancy patterns and the timing and drivers of autumn migration for a migratory songbird using doppler radar. *Ecography*, **39**, 1017–1024.
- Laughlin, A.J., Taylor, C.M., Bradley, D.W., Leclair, D., Clark, R.G., Dawson, R.D. et al. (2013) Integrating information

- from geolocators, weather radar and citizen science to uncover a key stopover area for an aerial insectivore. *The Auk*, **130**, 230–239.
- Li, F., Zhang, H., Xu, H., Liu, S., Zhang, L., Ni, L.M. et al. (2022) *Mask dino: towards a unified transformer-based framework for object detection and segmentation*.
- Lin, T.Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J. et al. (2014) *Microsoft coco: common objects in context*. Cite arxiv:1405.0312. Comment: (1) updated annotation pipeline description and figures; (2) added new section describing datasets splits; (3) updated author list.
- Lin, T.Y., Winner, K., Bernstein, G., Mittal, A., Dokter, A.M., Horton, K.G. et al. (2019) MistNet: measuring historical bird migration in the US using archived weather radar data and convolutional neural networks. *Methods in Ecology and Evolution*, **10**, 1908–1922.
- McLaren, J.D., Buler, J.J., Schreckengost, T., Smolinsky, J.A., Boone, M., van Loon, E. et al. (2018) Artificial light at night confounds broad-scale habitat use by migrating birds. *Ecology Letters*, **21**, 356–364.
- Meng, C., Liu, E., Neiswanger, W., Song, J., Burke, M., Lobell, D. et al. (2021) Is-count: large-scale object counting from satellite images with covariate-based importance sampling. *arXiv preprint arXiv:2112.09126*.
- Nebel, S., Mills, A., McCracken, J.D. & Taylor, P.D. (2010) Declines of aerial insectivores in North America follow a geographic gradient. *Avian Conservation and Ecology*, **5**, 1.
- Nilsson, C., Dokter, A.M., Verlinden, L., Shamoun-Baranes, J., Schmid, B., Desmet, P. et al. (2019) Revealing patterns of nocturnal migration using the European weather radar network. *Ecography*, **42**, 876–886.
- Nussbaumer, R., Benoit, L., Mariethoz, G., Liechti, F., Bauer, S. & Schmid, B. (2019) A geostatistical approach to estimate high resolution nocturnal bird migration densities from a weather radar network. *Remote Sensing*, **11**, 2233.
- O'Neal, B.J., Stafford, J.D. & Larkin, R.P. (2010) Waterfowl on weather radar: applying ground-truth to classify and quantify bird movements. *Journal of Field Ornithology*, **81**, 71–82.
- Parker, J.A., Kenyon, R.V. & Troxel, D.E. (1983) Comparison of interpolating methods for image resampling. *IEEE Transactions on Medical Imaging*, **2**, 31–39.
- Perez, G. & Maji, S. (2022) Domain adaptors for hyperspectral images. In: *26TH International conference on pattern recognition (ICPR), 2022*. Los Alamitos, CA, USA: IEEE Computer Society, pp. 3048–3055. Available from: <https://www.computer.org/csdl/proceedings-article/icpr/2022/09956407/1IHpCXQTsPe>
- Perez, G., Maji, S. & Sheldon, D. (2023) DISCount: counting in large image collections with detector-based importance sampling. *arXiv:2306.03151*.
- Ren, X. (2008) Finding people in archive films through tracking. In: *2008 IEEE Conference on Computer Vision and Pattern Recognition*. Anchorage, AK, USA, pp. 1–8. Available from: <https://doi.org/10.1109/CVPR.2008.4587533>
- Rhodes, B.C. (2011) *Pyephem: astronomical ephemeris for python*. Astrophysics Source Code Library, pp. ascl-1112. Available from: <https://ui.adsabs.harvard.edu/abs/2011ascl.soft12014R/exportcitation>
- Ribani, R. & Marengoni, M. (2019) A survey of transfer learning for convolutional neural networks. In: *2019 32nd SIBGRAPI conference on graphics, patterns and images tutorials (SIBGRAPI-T)*. Rio de Janeiro, Brazil, pp. 47–57. Available from: <https://doi.org/10.1109/SIBGRAPI-T.2019.00010>
- Rosenberg, K.V., Dokter, A.M., Blancher, P.J., Sauer, J.R., Smith, A.C., Smith, P.A. et al. (2019) Decline of the north American avifauna. *Science*, **366**, 120–124.
- RoyChowdhury, A., Sheldon, D., Maji, S. & Learned-Miller, E. (2016) *Distinguishing weather phenomena from bird migration patterns in radar imagery*. CVPR workshop on Perception Beyond the Visual Spectrum (PBVS), pp. 1–8.
- Russell, K.R. & Gauthreaux, S.A., Jr. (1998) Use of weather radar to characterize movements of roosting purple martins. *Wildlife Society Bulletin (1973–2006)*, **26**(1), 5–16.
- Russell, K.R., Mizrahi, D.S. & Gauthreaux, S.A., Jr. (1998) Large-scale mapping of purple martin pre-migratory roosts using wsr-88d weather surveillance radar. *Journal of Field Ornithology*, **69**(2), 316–325.
- Shamoun-Baranes, J., Alves, J.A., Bauer, S., Dokter, A.M., Hüppop, O., Koistinen, J. et al. (2014) Continental-scale radar monitoring of the aerial movements of animals. *Movement Ecology*, **2**, 1–6.
- Shen, Y., Cao, L., Chen, Z., Zhang, B., Su, C., Wu, Y. et al. (2021) Parallel detection-and-segmentation learning for weakly supervised instance segmentation. In: *2021 IEEE/CVF International conference on computer vision (ICCV)*. Montreal, QC, Canada, pp. 8178–8188. Available from: <https://doi.org/10.1109/ICCV48922.2021.00809>
- Stepanian, P.M. (2015) *Radar polarimetry for biological applications*. Ph.D. thesis. Oklahoma, USA: University of Oklahoma Norman.
- Stepanian, P.M., Entekin, S.A., Wainwright, C.E., Mirkovic, D., Tank, J.L. & Kelly, J.F. (2020) Declines in an abundant aquatic insect, the burrowing mayfly, across major north American waterways. *Proceedings of the National Academy of Sciences*, **117**, 2987–2992.
- Stepanian, P.M., Horton, K.G., Melnikov, V.M., Zrnić, D.S. & Gauthreaux, S.A., Jr. (2016) Dual-polarization radar products for biological applications. *Ecosphere*, **7**, e01539.
- Stepanian, P.M. & Wainwright, C.E. (2018) Ongoing changes in migration phenology and winter residency at bracken bat cave. *Global Change Biology*, **24**, 3266–3275.
- Torres, S. (2007) *Initial implementation of super-resolution data on the NEXRAD network*. Texas: San Antonio.
- Van Den Broeke, M.S. (2019) Radar quantification, temporal analysis and influence of atmospheric conditions on a roost

- of American robins (*Turdus Migratorius*) in Oklahoma. *Remote Sensing in Ecology and Conservation*, **5**, 193–204.
- Van Doren, B.M. & Horton, K.G. (2018) A continental system for forecasting bird migration. *Science*, **361**, 1115–1118.
- Van Doren, B.M., Horton, K.G., Dokter, A.M., Klinck, H., Elbin, S.B. & Farnsworth, A. (2017) High-intensity urban light installation dramatically alters nocturnal bird migration. *Proceedings of the National Academy of Sciences of the United States of America*, **114**, 11175–11180.
- Van Doren, B.M., Willard, D.E., Hennen, M., Horton, K.G., Stuber, E.F., Sheldon, D. et al. (2021) Drivers of fatal bird collisions in an urban center. *Proceedings of the National Academy of Sciences*, **118**, e2101666118.
- Winkler, D.W. (2006) Roosts and migrations of swallows. *Hornero*, **21**, 85–97.
- Yosinski, J., Clune, J., Bengio, Y. & Lipson, H. (2014) How transferable are features in deep neural networks? In: Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N.D. & Weinberger, K.Q. (Eds.) *Advances in neural information processing systems*, Vol. **27**. Curran Associates, Inc, pp. 3320–3328. Available from: https://papers.nips.cc/paper_files/paper/2014/hash/375c71349b295fbc2dcdca9206f20a06-Abstract.html
- Zhou, X., Koltun, V. & Krähenbühl, P. (2020) Tracking objects as points. In: Vedaldi, A., Bischof, H., Brox, T. & Frahm, J.M. (Eds.) *Computer vision – ECCV 2020. ECCV 2020. Lecture notes in computer science()*, Vol. **12349**. Cham: Springer. Available from: https://doi.org/10.1007/978-3-030-58548-8_28
- Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H. et al. (2019) A comprehensive survey on transfer learning. *CoRR*, abs/1911.02685.
- Zrnić, D.S. & Ryzhkov, A.V. (1998) Observations of insects and birds with a polarimetric radar. *IEEE Transactions on Geoscience and Remote Sensing*, **36**, 661–668.

Supporting Information

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

Appendix S1. Supplementary material. (A) Note about height bias on the radar sampling scheme. (B–D) Object detection preliminaries and implementation details. (E) Detector ablation experiments. (F) Supporting figures.