Independent Study: Finetuning MistNet for Roost Detection

Videsh Suman
vsuman@cs.umass.edu

Abstract

Weather radar networks can hold detailed information about biological phenomena, particularly bird migration, in the atmosphere. The US radar network covers the entire continental US with archived data since the 1990s. There is congregation of large numbers of roosting birds at nighttime, and their morning retreat from the roosting location is often visible as a distinctive pattern in radar images. There has been recent work on detecting and tracking these roosts using a machine learning system. However, this proposed system is not robust to false positives introduced by weather phenomena, particularly precipitation. My research involves investigating methods to filter out such false positive cases effectively. This can enable the ornithologists to conduct in-depth study on bird migration behavior with historical radar data. In this report, I describe my findings towards leveraging `MistNet`*, a deep learning model for discriminating precipitation from biology, as a post-processing step for the said detection and tracking system. Finally, I present a future plan for addressing the current finetuning issues, and exploring a more consolidated framework for this task.*

1. Introduction

The US weather radar network, WSR-88D[1], established more than 20 years back, is one of the most comprehensive instruments for studying bird migration due to its size, uniformity, and historical data archive. These radars, originally designed to study weather phenomena such as precipitation and severe storms, are also very sensitive to flying animals, like birds, bats, and insects.

A recent work [1] proposes a machine learning pipeline to track the roosting and migration behavior of songbirds from the weather radar data. This pipeline includes solving multiple subtasks — 1) render radar scans as multi-channel images for convenience while training deep recognition models; 2) perform single-frame roost detection leveraging Faster R-CNN [4], an established object detection framework; 3) use the detect-then-track paradigm to assemble roosts into sequences to improve detector performance and measure biological properties; 4) filter out a significant number of false positives using auxiliary geospatial information (introduced later in subsection 2.1). However, this geospatial information is available only for the data collected in the recent times (in and after 2013). Hence, this pipeline can only be used for the inference on recent radar scans. The inability to train such a model using the whole of weather radar archive, has limited the potential of conducting large-scale biological analyses. Analyzing the fine-grained measurements of swallow birds from the entire 20-year radar archive can allow ornithologists to study the demographics and migration behavior with respect to changes in climate and habitat.

The goal of my study has been to eliminate the dependence on this geospatial information for post-processing, which is only available since 2013 [10]. For filtering out the false positive detections arising due to precipitation, a segmentation model can be trained to discriminate between roosts and precipitation. Such a segmentation model, trained on the recent radar scans having geospatial information, can generalize over the archived legacy data, and enable the detection and tracking system to significantly improve its performance on the rest of the historical data.

Recently, a paper [7] proposed a deep segmentation model `MistNet`, to discriminate precipitation from biology in these archived radar scans. Using such a pre-trained MistNet directly on bounding box detections is not quite successful in filtering out the false positives. However, the hope is that a pre-trained MistNet can be fine-tuned to make accurate fine-scaled predictions for discriminating rain from biology in the regions detected by the detection and tracking model. Currently, the results obtained do not conform to this belief, and suggest the need for more training data and longer training time.

As a plausible next step, the detection and tracking model can be extended to generate a segmentation mask, besides the bounding box prediction, for each detected roost. An instance segmentation model like Mask R-CNN can be leveraged for this task. In principle, such a fine-scaled segmentation of roost detections can 1) significantly reduce the number of false positives, 2) raise the extent of

[1] Weather Surveillance Radar, 1988, Doppler.

evaluation to over 20 years of archived data, and 3) possibly be utilized by ornithologists for studies on bird demographics like trends in population size.

2. Method

The success of deep learning frameworks involving CNNs on visual recognition tasks, make them excellent candidates for analyzing other forms of spatio-temporal data. This section deals with the radar specific pre-processing steps, literature on roosts and the training methodology of the `MistNet` model.

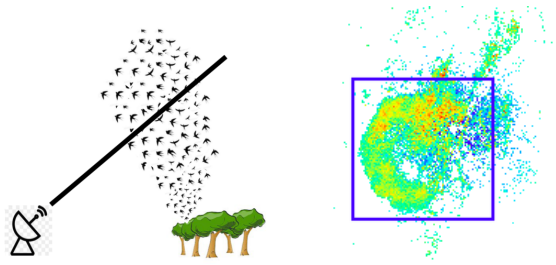
2.1. Radar Data

The WSR-88D (also called NEXRAD) network of over 140 radars is operated by the US National Weather Service [2]. They have ranges of several hundred kilometers and cover nearly the entire US. Radars conduct *volume scans* by rotating the antenna 360° at different elevation angles, thus sampling the cone-shaped slices of the airspace. Each such volume scan includes five sweeps at elevation angles 0.5° to 4.5°, and each sweep generates a set of gridded data products summarizing the radar signals. Of these data products, three legacy products — reflectivity factor (Z), radial velocity (v_r), and spectrum width (σ_w) — are being collected since the installation of these radars in 1990s, and three dual polarization (or "dual-pol") products [10] — differential reflectivity, differential phase, and correlation coefficient — have become available after 2013.

The dual-pol correlation coefficient (ρ_{HV}) is useful for discriminating precipitation from biology [2][10], but it is only available for the scans from 2013. The roost detection and tracking model [1] utilizes this information in the post-processing step to eliminate the false positive detections that are actually rain particles instead roosts. On the other hand, `MistNet` [7] utilizes it to create weak training labels.

A standard radar scan consists of a collection sweeps, one for each of the 3 legacy data products at each of the 5 elevation angles. Each of the 15 sweeps is a two-dimensional data array corresponding to a polar grid indexed by radius and azimuth. After standardizing the data and discarding the higher sweeps (which are unavailable in certain operating modes) [7], each sweep is resampled onto a Cartesian grid of size 600×600 centred at the radar station (see Figure 2a). This form of two-dimensional representation of a sweep resembles an image channel in terms of spatial coherence, and can serve as a reasonable input channel for deep convolutional networks.

[2]The biology class refers to all non-hydrometeor scatterers, including insects, dust, debris, etc. The goal is to eliminate precipitation, and not to make fine-grained distinctions among non-hydrometeor scatterers.



(a) (b)

Figure 1: (a) An illustration of a roost exodus (Source: Cheng et al. [1]) (b) Reflectivity map of a radar scan with a roost annotation.

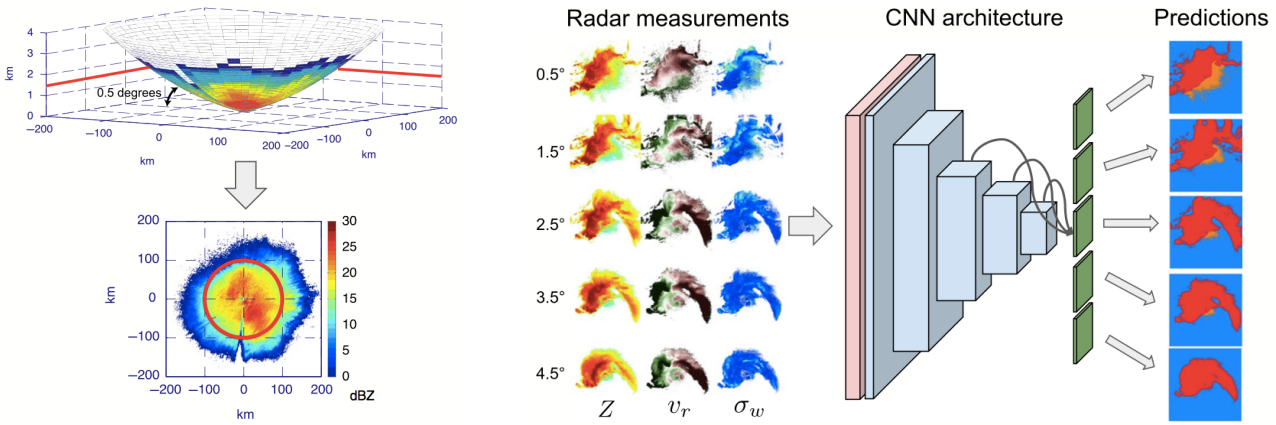
2.2. Roosts

A roost is the congregation of migratory birds at a particular location during the nighttime (see Figure 1a). The roost exodus refers to their mass departure 15-30 minutes before sunrise, and is visible on radar when the birds fly upward and outward within a radar airspace. The detection and tracking model [1] is trained with the following three input channels — reflectivity at 0.5°, radial velocity at $0.5^\circ$and reflectivity at 1.5°— the most discriminative radar products for humans. An important post-processing step is to leverage the correlation coefficient (ρ_{HV}) dual-pol data product to eliminate the false positive rain dominated regions while making inference on scans from 2013 and later. Biological targets have much lower ρ_{HV} values than precipitation due to their high variance in orientation, position and shape over time. A common rule is to classify pixels as rain if $\rho_{HV} > 0.95$ [3]. Cheng et al. classify a roost detection as a false positive (rain region) if a majority of pixels inside its bounding box have $\rho_{HV} > 0.95$. To make such an evaluation on the radar scans before 2013, an automatic precipitation segmentation model like `MistNet` needs to be utilized.

The data of manually annotated roosts [6] was used by the detection and tracking model [1]. These roost annotations (Figure 1b shows an example) are available only for the lowest elevation angle, where they are easily interpretable by humans. In their work, Cheng et al. restricted to seven stations in the eastern US and to month-long periods that were exhaustively labeled. They only used scans from 30 minutes before to 90 minutes after sunrise. For my experiment of finetuning, I restrict to a subset of the dataset used by Cheng et al., that contain the dual-pol data products.

2.3. The `MistNet` Model

For `MistNet` model, the input has dimension $15 \times 600 \times 600$ (a 15 channel 'image') and the output has dimension $3 \times 5 \times 600 \times 600$, which corresponds to the class probability for each of 3 classes (precipitation, biology, back-



(a) Radar geometry

(b) Overview of the CNN framework used in `MistNet`

Figure 2: (a) A sweep, at elevation angle of $0.5°$, traces out a conical slice of the airspace, and is rendered as a top-down image in Cartesian coordinates (Source: Lin et al. [7]). (b) A radar measurement is rendered as a 15-channel (3 products $\times$ 5 elevations) image, which acts as an input to the CNN. The pink layer depicts the adapter network, responsible for mapping 15 channels to 3 channels, resembling an RGB image. The intermediate blue layers, comprising of VGG-16 backbone and FCN8, process the 3-channel image to output five segmentation maps (red: rain, orange: biology, blue: background), one for each elevation. Activations into the green boxes are functions of several preceding layers. (Source: Lin et al. [7])

ground) at each pixel position in five 600×600 images, one for each elevation angle. On prediction, only the highest probable class for each pixel position is obtained. The input of 15 channels is reduced to 3 channels using a layer of $\times 1$ convolutions, to match a conventional RGB image. This 3-channeled 'image' acts as an input for the VGG-16 backbone [9] (pre-trained on ImageNet). The segmentation module of `MistNet` is based on the FCN8 (fully convolutional network with predictions at a spatial granularity of 8 pixels) architecture [8], and is designed to make five predictions (one per elevation) for each input scan. This is accomplished by creating five separate branches with input as the activations from several preceding convolutional layers, and output class probabilities. Originally, this model used pixel-level segmentation labels from the thresholding criterion of dual-pol correlation coefficient ($\rho_{HV} > 0.95$: precipitation, $\rho_{HV} \leq 0.95$: biology), as weak training labels. According to [7], a fully trained `MistNet` exceeds the performance of dual-pol thresholding on discriminating precipitation from biology.

3. Experimental Setup

Cheng et al. [1] used the dual-pol thresholding (pixels with $\rho_{HV} > 0.95$: classified as precipitation) to remove a significant number of false roost detections in the post-processing step. However, that was only possible for the recent scans. In this work, my experiments are aimed at using an automatic segmentation model for filtering out the false positive detections, and extend this post-processing step for legacy scans too. In this section, I describe the dataset, the finetuning and visualization experiments in detail.

3.1. Dataset

From the dataset used in [1], I extracted only the newer scans having the dual-pol data products, a total of 7261 scans. Out of these, 3346 contain one or more roosts, whose bounding box annotations are available [6]. For the fine-tuning experiment, I use the split of 4394/830/2037 as train/validation/test sets. The pixel-wise segmentation labels for training are obtained from dual-pol thresholding of the correlation coefficient ρ_{HV}.

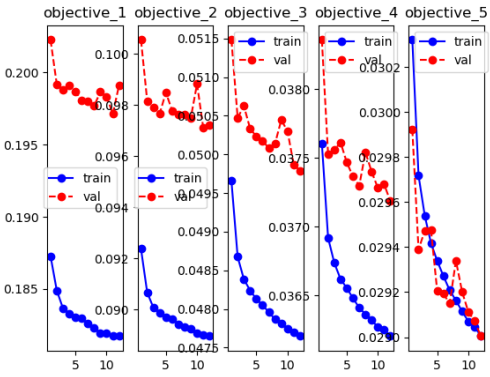
3.2. Experiments

For baseline performance, the trained `MistNet` model was directly used to classify the roost regions based on their pixel-wise labels of precipitation, biology or background. Roost annotations are available only for the lowest elevation angle, hence prediction for only the corresponding channel should be considered. A roost annotation can be classified as a roost by `MistNet` if the labels in its corresponding segmentation map follows this rule,

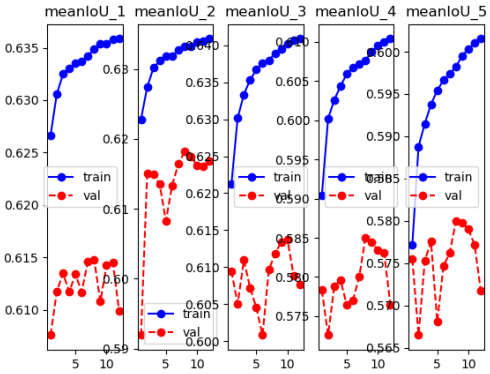
$$\frac{\#biology}{\#biology + \#rain} \geq threshold \tag{1}$$

where $\#biology$ and $\#rain$ are numbers of biology and precipitation classified pixels respectively, while $threshold$ can be set to 0.5 for reasonable confidence. A perfect segmentation model should be able to classify all the roost annotations as roosts with a reasonable confidence.

In the hope of better performance, the `MistNet` model is fine-tuned using the data and labels as mentioned in the subsection 3.1. The model parameters are initialized with weights from the pre-trained `MistNet`. The weights are updated by stochastic gradient descent, while the gradients are computed using backpropagation. The fine-tuning is performed in almost the same configuration as described in [7]. The training progress (Figure 3) is observed through the metrics of mean *intersection over union* (IoU) and the objective i.e. cross-entropy loss. In my experiment, I stopped the training when the validation objective for the lowest elevation angle (Figure 3a) had convereged.



(a) cross-entropy loss



(b) mean intersection over union (IoU)

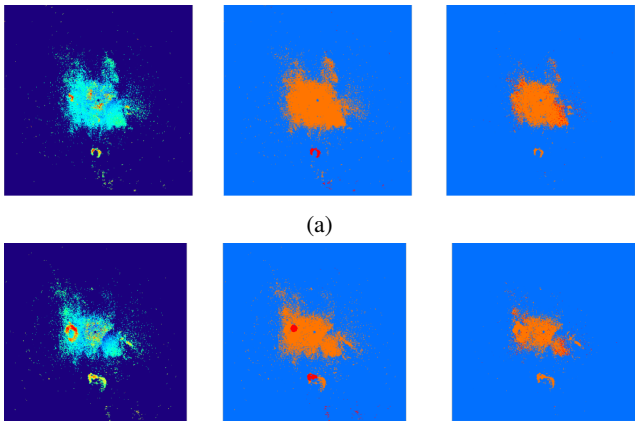
Figure 3: Training and validation plots of the objective and the mean IoU across the fine-tuning epochs for all 5 elevations.

4. Results

This section includes the quantitative as well as the qualitative results of the conducted experiments, along with some interpretations and insights for these results.

4.1. Baseline

The pre-trained `MistNet` fails to segment a significant number of roosts as biology. Figure 4 illustrates one such example. The pre-trained `MistNet` performs much worse on comparing with the performance by dual-pol thresholding in Table 1. After observing a number of such visualizations qualitatively, it seems that the regions surrounded by background from all sides (Figure 4a), or those with very high reflectivity (Z) signal, are usually segmented as precipitation. In Figure 4b, the pre-trained `MistNet` identifies a small sub-region of the left roost as precipitation, which is also the location of very high reflectivity signal.



(a)

(b)

Figure 4: Left to right: reflectivity (Z) projection; segmentation map (red: rain, orange: biology, blue: background) obtained using pre-trained `MistNet`; segmentation map obtained from dual-pol thresholding. These two scans have been collected on the same date, by the same radar station within a small time interval.

4.2. Fine-tuning

The pre-trained `MistNet` was fine-tuned for 13 epochs, the training progress of which is shown in Figure 3. The fine-tuning was performed only on the newer radar scans, having the dual-pol data products. The segmentation masks obtained from ρ_{HV} thresholding, served as labels for fine-tuning.

Contrary to the expectation, the fine-tuned model performs even worse than the pre-trained model as presented in Table 1. With a careful observation at visualizations of several radar scans, it can be inferred that the fine-tuned model learns to perform well on roost regions with high reflectivity (Figure 5c). However, it becomes worse at predicting biology for some of the low reflectivity regions (Figure 5a, Figure 5c) compared to its pre-trained version. Segmentation masks from the fine-tuned model in Figure 5b, with precipitation (black) all around the predicted biology regions,

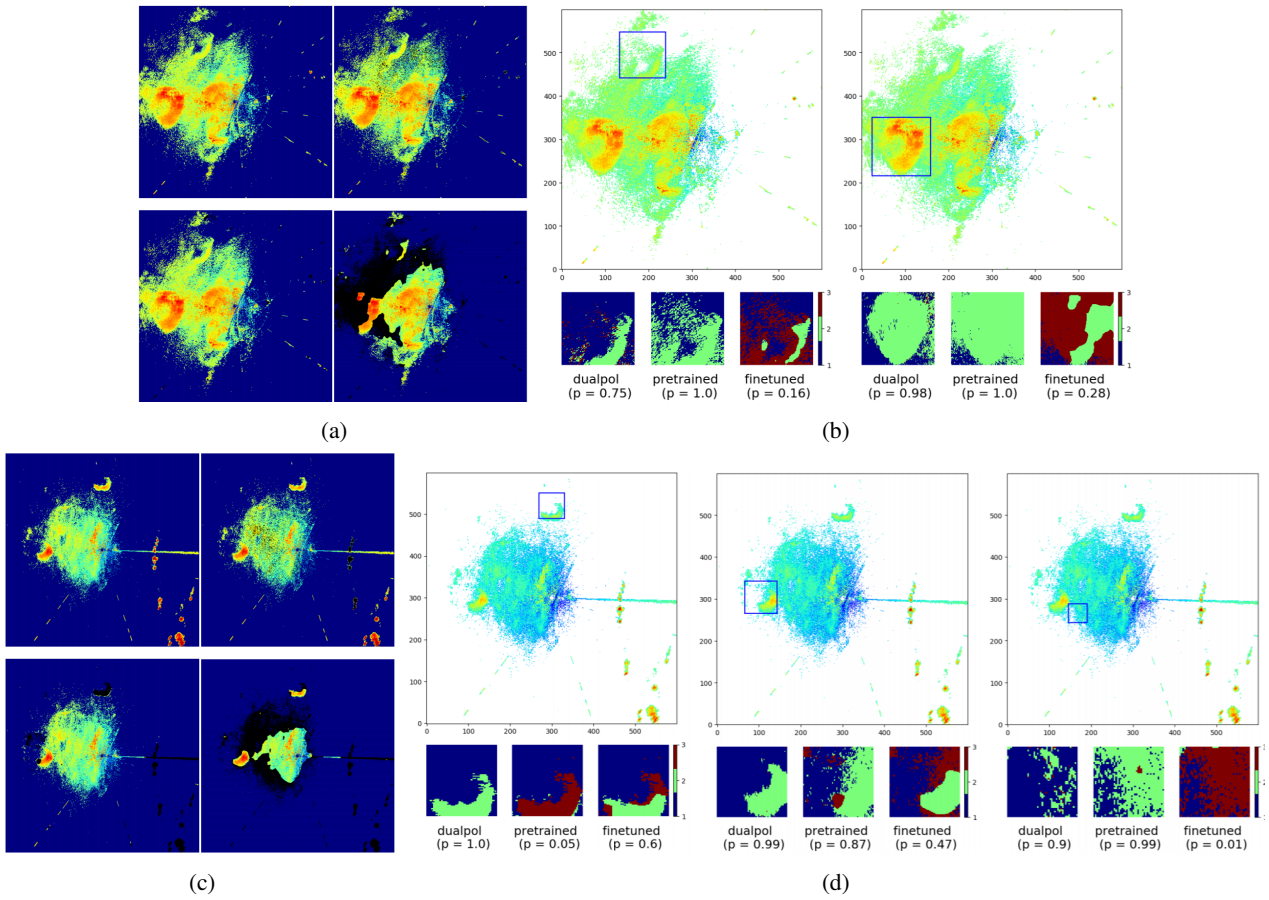


Figure 5: (a) Top: reflectivity (Z) projection for the lowest elevation angle $0.5°$ and the precipitation mask (black color over the reflectivity signal) obtained from dual-pol thresholding; bottom: precipitation masks obtained from pre-trained `MistNet` and fine-tuned `MistNet` respectively. (b) Bounding box annotations of roosts, along with the respective segmentation masks (brown: rain, green: biology, blue: background) obtained by the three techniques (p is the confidence measure). Figures (c) and (d) represent similar visualizations for a different radar scan.

Segmentation Technique	Dual-pol Thresholding	Pre-trained `MistNet`	Fine-tuned `MistNet`
Performance	1165/1170	870/1170	786/1170

Table 1: Accuracy of classifying the annotated roost regions as roosts based on the pixel-wise segmentation maps obtained by different segmentation techniques. A fraction refers to the number of roost regions classified correctly with respect to the total number of ground truth annotations in the test set.

indicates that the model has memorized to classify the high reflectivity ring-like regions as biology.

The incorrectly segmented black masked regions by the finetuned model in Figure 5a and Figure 5c, are not exhibited by the pre-trained model. As a result, many roost regions with lower reflectivity (3rd roost annotation in Figure 5d) are misclassified as precipitation. There are more examples like these which indicate the *un-learning* of the model upon fine-tuning, and that contributes to the bulk of the poor performance represented in Table 1.

5. Conclusion

A robust automatic roost segmentation model is necessary for a significant improvement in the detection and tracking inference on the legacy scans. Considering the performance of the pre-trained model as a baseline on the roost classification task, the fine-tuned model performs even worse. The model seems to memorize the high reflectivity ring-shaped regions as biology, while tending to predict some of the lower reflectivity regions (including some roost regions) as rain, and contributing to *un-learning*. This

should be alleviated by a larger dataset for fine-tuning; a dataset of only 7261 scans is probably not enough. The strategy of freezing some of the pre-trained weights of the CNN framework can mitigate the *un-learning* issue while fine-tuning.

6. Future Work

The current results are contrary to what the goal of this work is. Nevertheless, there is quite some room for improvement. In this section, I present two major directions for future work that are aimed at improving the performance of the detection and tracking model on historical scans.

6.1. Improving the fine-tuning performance

The fine-tuned model performs worse compared to the pre-trained MistNet. With relevant visualizations, some inference (subsection 4.2) can be drawn on what the model is learning. A very natural step is to gather more training data having dual-pol products, from the roosting season. Currently, I only used radar scans from years 2012 and 2013. More scans till 2019 can be gathered for fine-tuning the model which can increase the size of the entire dataset to 30000 scans. Freezing some layers of the MistNet framework can help reduce the parameter space, and might address the issue of *un-learning*. After sufficient improvement on the dual-pol scans, this model should be tested on the legacy scans to filter out the false positive detections and improve the results of the roost detection model.

6.2. Extending to Mask R-CNN

The detection and tracking model can be extended to predict segmentation masks for each candidate region of interest. The recent success of Mask R-CNN [5] on instance segmentation tasks make it an appropriate candidate framework for improving the detection performance on historical radar scans. Mask R-CNN includes an additional branch, compared to the Faster R-CNN framework, for generating the segmentation masks over candidate RoIs generated by the Region Proposal Network. A suitable baseline for this task could be the performance of the roost detection model coupled with the false positive filtering by dual-pol thresholding.

Acknowledgements: I would like to acknowledge the guidance of Prof. Daniel Sheldon and Prof. Subhransu Maji, and the active support of Zezhou Cheng and Tsung-Yu Lin throughout this independent study.

References

[1] Z. Cheng, S. Gabriel, P. Bhambhani, D. Sheldon, S. Maji, A. Laughlin, and D. Winkler. Detecting and tracking communal bird roosts in weather radar data supplementary materials.

[2] T. D. Crum and R. L. Alberty. The wsr-88d and the wsr-88d operational support facility. *Bulletin of the American Meteorological Society*, 74(9):1669–1688, 1993.

[3] A. M. Dokter, P. Desmet, J. H. Spaaks, S. van Hoey, L. Veen, L. Verlinden, C. Nilsson, G. Haase, H. Leijnse, A. Farnsworth, et al. biorad: biological analysis and visualization of weather radar data. *Ecography*, 42(5):852–860, 2019.

[4] R. Girshick. Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 1440–1448, 2015.

[5] K. He, G. Gkioxari, P. Dollár, and R. Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017.

[6] A. J. Laughlin, D. R. Sheldon, D. W. Winkler, and C. M. Taylor. Quantifying non-breeding season occupancy patterns and the timing and drivers of autumn migration for a migratory songbird using doppler radar. *Ecography*, 39(10):1017–1024, 2016.

[7] T.-Y. Lin, K. Winner, G. Bernstein, A. Mittal, A. M. Dokter, K. G. Horton, C. Nilsson, B. M. Van Doren, A. Farnsworth, F. A. La Sorte, et al. Mistnet: Measuring historical bird migration in the us using archived weather radar data and convolutional neural networks. *Methods in Ecology and Evolution*, 2019.

[8] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.

[9] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[10] P. M. Stepanian, K. G. Horton, V. M. Melnikov, D. S. Zrnić, and S. A. Gauthreaux Jr. Dual-polarization radar products for biological applications. *Ecosphere*, 7(11):e01539, 2016.