**Project Report**

**A Deep Learning model for Bird Habitat detection using Earth Observation Satellite Images**

**Team Members:**

1. Akash Anand ([aanand37@wisc.edu](mailto:aanand37@wisc.edu) ), Department of Forestry and Wildlife Ecology (Graduate)
2. Zi Hern Wong ([zwong4@wisc.edu](mailto:zwong4@wisc.edu)), Department of Computer Science. (Undergraduate)

**GitHub: https://github.com/zihernwong/ECE539**

**ABSTRACT**

Habitat detection is a major part of wildlife ecology, and it has been done since hundreds of years. But the conventional approaches require huge manpower and resources which is not feasible. Furthermore, it is impossible to map millions of species around the world. This has prompted us to try to train a CNN model on different layers of satellite images to map the habitats of various species of birds. Satellite images are used because they are accessible around the globe and are updated frequently. In this project, we performed classifications on different birds from published data (CUB-200 bird dataset) using CNN to compare the approach with the habitat detection models and compare the accuracy of both models.

**2. Introduction**

Species habitat detection is one of the mainstays of ecology, in theory, species are expected to occur where conditions permit population growth to be stable or positive. Although there are millions of species present on this planet and mapping their distribution by conventional sampling methods is impossible. Therefore, as an alternative we can use satellite images as a proxy for different environmental and climatic parameters for two good reasons, first being the limitation for humans to do sampling in distinct locations around the world and secondly it can take a lot of time and human resources. Satellite images provide full coverage of the globe and are available every week or so. Different environmental and climatic parameters influence different species, and we can use their corresponding satellite images to train CNN architecture which can learn from the available ground datasets to map the distribution of species for a broader range.

The recent advancements in artificial intelligence for computer vision hold immense potential to study several aspects of ecological systems. Till now, several works have been done to identify the bird species (physically) (Weinstein, Garner et al. 2021), but not a lot of work has been done to identify their habitat. As bird habitat depends on multiple factors like land type, climate, latitude etc., there is a need to build a model that can take these parameters as an input and accurately predict their habitat. Our project is going to focus on a single bird, Northern Parula, which is only found in the eastern US and has a very selective habitat (as shown in Fig 1), and we will try to predict its habitat.

In this project we will also train a model to classify different bird that are present in CUB-200 dataset to identify a bird based on their RGB photographs. This is a typical neural network example which has been perform in past studies as well. But we are performing the bird classification to compare it with the CNN model we will be developing for the habitat detection. Further, for the habitat detection model, we trained the CNNs based on two criteria, the first one is using only the RGB satellite image, and the second approach is to use 20 satellite bands. The purpose of using two approaches is to understand the effect of different layers on the model.

Present work can have a significant impact on identifying species habitat around the world with limited ground datasets. The classification will help in tracking the change in distribution of species and ultimately help in the conservation of species that are threatened and endangered.

Map

Description automatically generated

*Figure 1: Training and testing data for Northern Parula. Red dots (#916) denote the presence and yellow dot (#2662) denotes absence of the bird.*

**3. Method**

**3.1. Data**

**3.1.1. Caltech-UCSD Birds**

The Caltech-UCSD Birds 200 (CUB-200) dataset is one of the most popular and challenging image dataset annotated with 200 birds species (Welinder et al., 2010). There are 200 different birds in the data with total 11,788 images. For present study, due to the limitation of processing power and time we have only used 20 random birds for the classification. The number of images used for training and testing of each bird is shown in Fig. 2

A picture containing text, bird

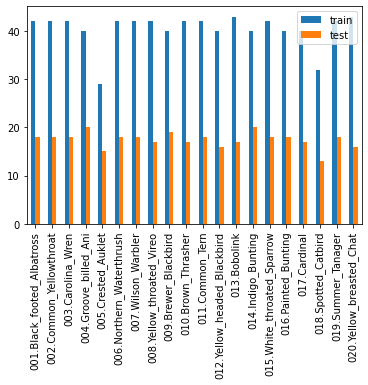
Description automatically generated

Figure 2: Number of training and testing data used for each of the 20 birds

**3.1.2. Bird Location Data**

As shown in Fig. 1, we have used almost 3578 birds’ locations for training and testing our habitat detection model. These location data came from the Breeding Bird Survey (BBS) which provides ground data each year for thousands of bird’s species (Bystrak, 1981). But we were only focusing our study on Northern Parula, so all the location where this bird was recorded, we used it as our present only points and others we used as absent point locations.

**3.1.2. Satellite Images**

For habitat detection we have prepared a stack of 20 different satellite datasets which correspond to different environmental and climatic parameters (such as RBG image, temperature, precipitation, elevation etc.). Before building the CNN models, we performed some data preprocessing steps. First, we resampled all the layers to a common resolution because originally spatial resolution of different satellite images varies from 10m to 1000 m. Second, as the locations of Northern Parula indicate only the presence only data, which is insufficient for training the models as we also require absence data. To get the absence data we will generate pseudo-absence data (Barbet‐Massin, Jiguet et al. 2012). After having both presence and absence data, the third step was to extract small image chips from the full image. We intend to make training and testing data images of size 50 x 50 pixels (with 1 pixel in image covering 250 m on ground). Therefore, a single training or testing image has a dimension 50 x 50 x 20 (an example is shown in Fig 3), where 50 x 50 is the image size and 20 is the total number of layers. After generating the sample image, we labeled those images as 0 for absence and 1 for presence.

A picture containing text

Description automatically generated

Figure 3: A sample of all the 20 bands used as input for a single point location.

**3.2. Methods**

**3.2.1. CNN model for Bird Identification**

For the bird identification model, we used 20 birds from the CUB-200 dataset. The image size was different for each image, so we have to fix the image input size to 100x100. Also, we have performed data format related changes to make dataset compatible with keras. The training architecture we used for the classification is shown in Fig. 4.

Graphical user interface, table

Description automatically generated with medium confidence

Figure 4: CNN architecture for Bird Identification

To prevent overfitting, we applied early stopping and stop the learning if val\_loss value has not decreased for 10 epochs. We also applied leaning rate reduction and data augmentation on the trained images. An example of data augmentation is shown in Fig 5.

A group of birds flying

Description automatically generated with medium confidence

Figure 5: Example showing data augmentation

**3.2.2. CNN model for Bird Habitat Detection**

For bird habitat detection we trained two CNN models, the first one was only using the RGB satellite image as input and the second was using all the 20 satellite image bands as input. We performed data format related changes to make the input layers compatible with keras. The architecture used for both the models are relatively not too deep, because the models was learning very quickly, therefore deeper CNN architecture would have led to overfitting. The architectures are shown in Fig. 6(a) and 6(b).

Figure 6(b): CNN Architecture for all 20 satellite bands as input

Table

Description automatically generatedGraphical user interface, text

Description automatically generated

Figure 6(b): CNN Architecture for 20 satellite image data as input

Figure 6(a): CNN Architecture for RGB satellite data as input

The hyperparameters used in this analysis had SGD as optimizers with a learning rate of 0.005 and the losses were estimated using sparse categorical cross entropy. The epoch was also not set high because the CNN was learning very quickly.

**3.2.3. Validation Metrics**

To validate the results, we used different validation metrices in this study. For the bird identification model, after making predictions we plotted confusion matrix to visualize how many of the classification went well and how many of them got misclassified. We also printed the classification report with precision, recall, f-1 score and support in it. The performance of CNN architectures was analyzed by plotting loss and accuracy of training and validation data.

1. **Result and Discussion**
   1. **Results**

For the bird identification model, we achieved a decent accuracy of about 0.6 which is good because of the small number of images and too many classes that too of same species (i.e., bird). The loss and accuracy of our model is shown in Fig. 7, where it can be seen that after 50 epochs the overall learning stabilizes.

Chart

Description automatically generatedChart, histogram

Description automatically generated

Figure 7: Loss and Accuracy of training and validation dataset

The confusion matrix is presented in Fig 8, which shows that the model is predicting some of the species very well like Cardinal and Painted Bunting (one of the reason is because they can be physically differentiated from other common birds) whereas a few species are being predicted incorrectly because of being quite common physically.

Graphical user interface

Description automatically generated with medium confidence A picture containing text, bird, oscine

Description automatically generated

Figure 8:Confusion matrix and an example of misclassified image

The classification report in Fig 9 also indicates that some species have a higher recall rate and f1-score than others whereas some of the species like Black Footed Albatross has extremely low recall (0.06) and f1-score (0.08).

Table

Description automatically generated

Figure 9: Classification Report

Chart

Description automatically generatedA picture containing shape

Description automatically generatedFor the bird habitat detection model, the training went well for both RBG input and 20 bands input with an accuracy of 0.89 and 0.80 respectively. The loss and training graph for both the models are shown in Fig 10 & 11.

Figure 10: Accuracy and loss of CNN model with RGB satellite image as input

Graphical user interface

Description automatically generated with medium confidence Chart, line chart

Description automatically generated

Figure 11: Accuracy and loss of CNN model with 20 satellite image bands as input

A picture containing text, monitor, indoor, screen

Description automatically generatedAs per Fig 10, the training went well for the RGB satellite image with 50 epochs whereas the model got saturated from the beginning in 20 band model especially the accuracy, the good part was the loss function as it was decreasing in every epoch which means the model was learning something in every epoch. As we archived a good training for the RGB satellite image, we went on to see the prediction on Madison’s RGB image, shown in Fig 12.

Map

Description automatically generated

Figure 13: RGB satellite image of Madison (left) and pixel-wise probability of Northern Parula (target bird) habitat

The prediction came out well, visually we can interpret that the habitat probability is high where there are more vegetation and water. Also, it is clearly eliminating the urban buildings, especially in downtown Madison, the probability is near to 0.

* 1. **Discussion**

Based on the result, it is concluded that the bird classification model is performing well for the 10 birds of CUB-200 dataset (with good recall and f1-score for those birds that are physically different from each other). Whereas there are few birds which were misclassified that is mostly due to the similarity in their structure and the dataset was very complex and needed more preprocessing. We achieved an accuracy of 0.6 which is close to the accuracy achieved on Kaggle which is 0.69 (but this is for all 200 birds, but presently we used 20 random birds to reduce the processing time). To improve our model, we can try different optimizers, learning rates, improve data augmentation parameters. Also, we can train the model in deeper architecture to get a better accuracy.

For the habitat detection, model performed well for the RGB satellite input, and the prediction also looked great in Fig 12. To further improve this model, we can try data augmentation and different optimizers. The model that ran on 20 satellite bands saturated in its early epoch, we can try decreasing the input shape, maybe 100x100 is becoming homogenous for all the samples and reducing the shape will increase the heterogeneity in the data.

Overall, the objectives of this study have been achieved, we successfully predicted the habitat of single bird species. Although our result opens up new challenges including mapping of species on a regional or global scale and designing similar architectures for other species.

**References**

Barbet‐Massin, M., et al. (2012). "Selecting pseudo‐absences for species distribution models: how, where and how many?" Methods in ecology and evolution **3**(2): 327-338.

Deneu, B., et al. (2021). "Convolutional neural networks improve species distribution modelling by capturing the spatial structure of the environment." PLoS computational biology **17**(4): e1008856.

Weinstein, B. G., et al. (2021). "A general deep learning model for bird detection in high resolution airborne imagery." Ecological Applications: e2694.

Bystrak, D. (1981). The North American breeding bird survey.

Welinder, P., Branson, S., Mita, T., Wah, C., Schroff, F., Belongie, S., & Perona, P. (2010). Caltech-UCSD birds 200.