

Course Project: Birdwatching

1 Project Description

In this project, you will tackle a common challenge of computer vision: object detection. We will use a birds dataset for this project. While standard classification tasks (e.g., Cat vs. Dog) rely on distinct structural differences, this project requires your model to locate the bird in an input image (bounding box regression) and to classify the bird into one of the many bird species (classification). You are free to treat them as two independent tasks, or you can look up models designed for object detection that implement both tasks simultaneously.

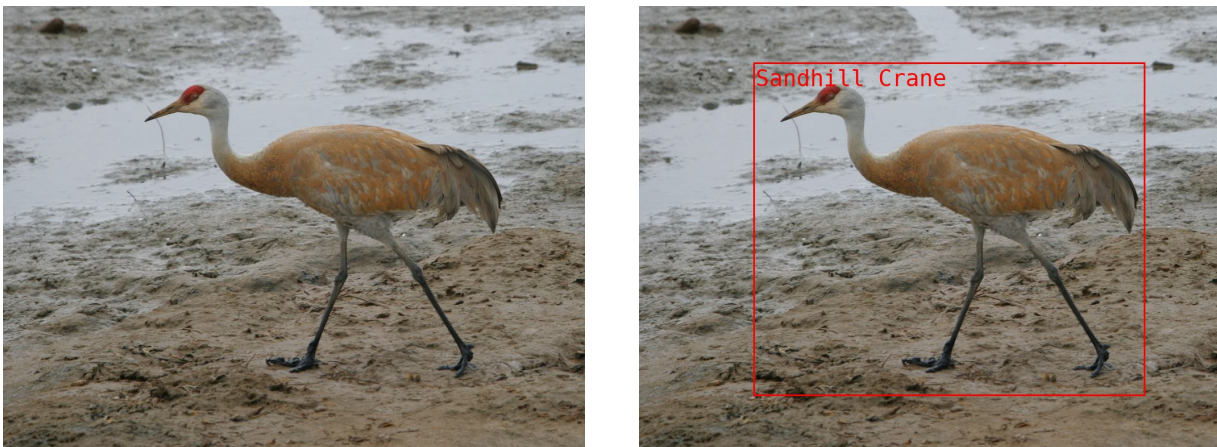


Fig. 1: An example from the dataset.

The training and validation datasets can be downloaded from the course website. You will use these to train and evaluate your models. You may use the validation dataset to train the model, but we recommend first using it for hyperparameter optimizations. The models will be scored on Kaggle on a test dataset. You can use the `calculate_AP` method from `topic_1/utis/metrics.py` to evaluate your models. Note that this method supports multiple objects in an image, so make sure to `det_boxes`, `det_labels`, `det_scores`, `true_boxes`, and `true_labels` include a leading dimension representing the number of objects (which will be 1 for all examples in this project). `n_classes` is the total number of classes.

2 Kaggle Submission Format

To ensure effective evaluation, the header and content of the `prediction.csv` file must exactly match the format, and text case matters. We recommend using the Python `csv` library to format the file correctly. Custom code may result in errors. As Table 1 shows, each row corresponds to one image.

You can use the “Submit Prediction” button at the upper right corner of the Kaggle page to submit. Remember to write the names of all group members in the description. After submitting, you need to go to the “Submissions” tag and press the “Select” box after your best submission, so your submission will be shown on the leaderboard.

Id	Predictions
...	...
...	...
...	...
...	...
43cc1fa0-2eaa-11f1-8427-c84bd6ab7bb0	{"label": "Merlin", "x_min": 0.69140625, "y_min": 0.1953125, "x_max": 0.919921875, "y_max": 0.23046875, "score": 0.80742108}
...	...

Table 1: Format of `prediction.csv`

3 Deliverables

The following are the deliverables of this project:

- A model that outputs the bounding box (normalized) and species of the bird, along with the score.
- The number of parameters in the model must not exceed 10 million.
- You are free to use any publicly available model (pre-trained or otherwise) with or without augmentation, but it is not a requirement.

As a starting point, check out some of the object detection models in torchvision and the classic object detection tutorial.

4 Submission

1. Submit all your code, including training and evaluation, as a `.zip` file.
2. Submit `prediction.csv` file to Kaggle for scoring and evaluation
3. Submit a 2-page report (1-inch margin, 12-point font) and include
 - Your approach, model, and any other design choice.
 - Hyperparameters that you used for training.
 - Training and test results.
 - Any other interesting details about the approach or model.