Assignment: Bird and Drone Detection & Tracking from UAV Camera

Objective

Develop a computer vision and deep learning model capable of detecting and tracking birds and drones from an onboard camera installed on a UAV. The model should effectively differentiate between birds and drones, even when they appear as small objects in the image (low pixel count). The focus is on building an efficient, explainable, and deployable solution suitable for resource-constrained environments.

Datasets

The following are the links to the datasets used in this project:

- Abdulrahman Eidhah
- Dr. Ali Hilal, Dr. Maher Al-Baghdadi

The datasets contained images of 2 classes: Birds and Drones. It composed of a total 4620 original images and after augmentation composed of about 11000 images. The following augmentations were performed using the Roboflow website to implement changes faster:

- Resize to 640x640 (Fit within)
- Horizontal Flip
- Varied Saturation: +-25%
- Varied Exposure: +-10%
- Varied Brightness: +-15%

Tiling was not used as the images were not of high enough resolution to properly train a model on the tiled image.

Base Model

The model used is the Yolov8 model due to the use of SPPF (Spatial Pyramid Pooling Fusion) in its architecture. The purpose of SPPF is to provide a multi-scale representation of the input feature maps. By pooling at different scales, SPPF allows the model to capture features at various levels of abstraction. This can be particularly useful in object detection, where objects of different sizes may need to be detected. It is an optimized version with the same mathematical functionality of SPP (Spatial Pyramid Pooling) originally used in YOLOv3, but fewer floating-point operations (FLOPs). The specific architecture used in this project is yolov8n for a balance in speed and accuracy.

The selection of Yolov8 is inspired by the positive performance of YOLOv3-SPP used by **Team Alexis [1]** in the **Drone vs. Bird Detection Challenge 2020.** The model proved to be able to achieve good performance in detecting small objects of interest. They were able to detect 89.1% of the objects in the training data. The selection of the Yolov8 model is further supported by the very good support provided by ultralytics.

Optimizations

The following optimizations are possible to be implemented on the model to achieve better performance on edge devices:

- **Pruning:** It is important to note that pruning a model might not always significantly reduce inference time. In some cases, pruning can introduce sparsity patterns that are not favourable for efficient computation, leading to slower inference times. Sparse computations can be more complex to implement and optimize compared to dense computations. This complexity arises from the need to handle irregular data structures, which can make efficient parallel processing more challenging.
- **Static Quantization:** Static quantization converts the weights and activates of a neural network to lower precision (e.g., from 32-bit floating-point to 8-bit integers) during the training or post-training phase. During inference, both the weights and activations are quantized to int8.
- **Model Distillation:** A smaller model, often referred to as a student model, is trained to mimic the behaviour of a larger, more complex model, known as a teacher model. The goal is to transfer the knowledge and performance of the larger model to the smaller one. Probability distributions produced by the teacher model for each input example is used to train student model.
- **SAHI (Slicing Aided Hyper Inference):** Partitioning images into manageable slices, running object detection on each slice, and then stitching the results back together. Allows for improved detection of small objects in large images while reducing computational cost.

Results

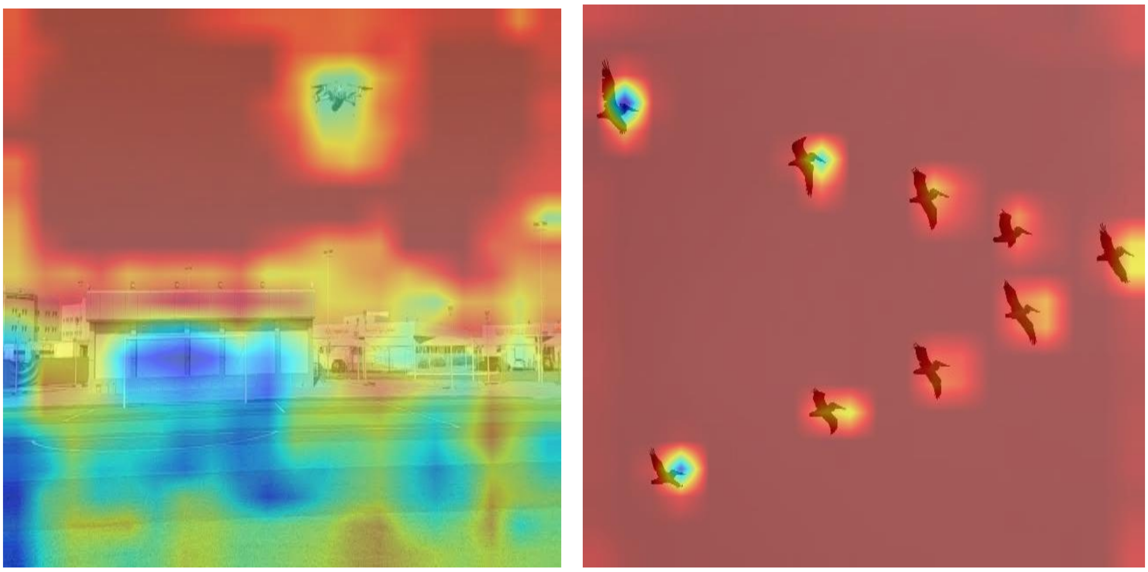
The following table presents the results of evaluation on the validation set containing 924 images with 1102 instances. There are 177 bird instances and 925 drone instances. Furthermore, the evaluation was done using the base model as well as pruned models where 10 and 20% of connections in all 2D-conv layers. It can be seen that there was significant difference in performance of the models.

Model	Box(P)	R	mAP50	mAP50-95
Base Model	0.864	0.864	0.884	0.561
Pruned - 10% Conv	0.539	0.436	0.502	0.255
Pruned - 20% Conv	0.24	0.0585	0.131	0.0582

To test speed of the models, the models were tested of 2 videos containing birds and drones. The following table shows the average inference time of the models on each frame on different videos:

Model	Video-1	Video-2
Base Model	15.65 ms	14.80 ms
Pruned - 10% Conv	15.65 ms	15.52 ms
Pruned - 20% Conv	15.63 ms	15.48 ms

It can be noticed that there is no significant difference in inference time between the models even with pruning, but the difference will increase when deployed on lower performance devices. Furthermore, there is an increase in inference time on video 2 when the pruned model is used. This can be attributed to the increase in sparsity in the model which cause more complex dense computations which in turn increase computation time.



The above images show the activation from the model's second last layer. It highlights the regions of an image most relevant for prediction. The model seems to highlight region of image that are not sky (light blue) and use these regions when making predictions. This might be due to the fact that most of the training data contained birds and drones in the sky.

Resources

[1] Coluccia, Angelo, et al. "Drone vs. bird detection: Deep learning algorithms and results from a grand challenge." Sensors 21.8 (2021): 2824.

[2] Thai, Phat Van, et al. "Small Flying Object Detection and Tracking in Digital Airport Tower through Spatial-Temporal ConvNets." Available at SSRN 4740820 (2023).

[3] Huang, Huadong, et al. "Improved small-object detection using YOLOv8: A comparative study." Appl. Comput. Eng 41.1 (2024): 80-88.

[4] Shandilya, Shishir Kumar, et al. "YOLO-based segmented dataset for drone vs. bird detection for deep and machine learning algorithms." Data in Brief 50 (2023): 109355.