Maria Jose Viveros Riquelme

Professor Patricia McManus

ITAI 1378

October 15, 2025

Lab 06. Object Detection with Transfer Learning

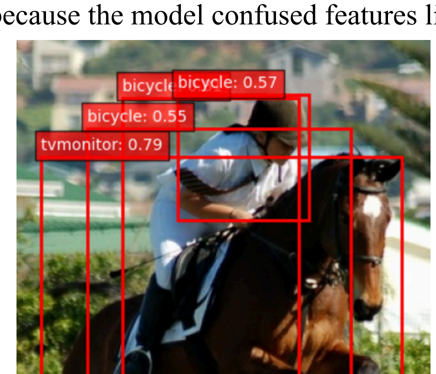
Image classification assigns a single label to an entire image, indicating the main object or category present. Object detection, however, locates all objects in an image, specifying both their classes and positions with bounding boxes. So, while **classification** answers “what is in the image?” **detection** answers “what is in the image and where is it?” We use fractions (0-1) for bounding box coordinates instead of pixels because they **normalize the coordinates**. This makes the model more flexible: it can handle images of different sizes without recalculating coordinates, and predictions are easier to generalize across datasets with varying resolutions. IoU, or **Intersection over Union**, measures how much the predicted bounding box overlaps with the ground truth box. **An IoU of 0.5 is commonly used as a threshold because it strikes a balance: it considers a prediction correct if the box overlaps sufficiently with the ground truth, but it does not require perfect alignment.**

During the Implementation Experience, in the **Visualization function** section, the most challenging part of implementing *plot_detection* was ensuring that the bounding boxes and labels were correctly drawn over the image, especially when the coordinates were normalized. I resolved this by multiplying the normalized coordinates by the image's width and height before plotting. This ensured the boxes aligned with the actual image pixels. However, if I forgot *max(0, ...)*, the width or height could become negative when boxes do not overlap, leading to a negative intersection area. Using *max(0,..)* ensures the intersection area is never negative.

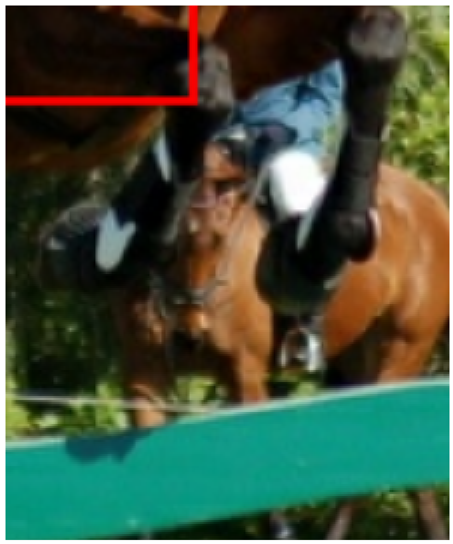
In the **Results Analysis section**, during the confidence threshold experiment, a threshold of 0.3 detected many more boxes, including several low-confidence predictions, while a threshold of 0.7 resulted in fewer boxes that were more likely to be correct. I prefer the higher threshold because it reduces false positives and highlights predictions where the model is more confident, even if it misses some true objects. In the IoU threshold experiment, increasing the IoU threshold caused recall to decrease because predicted boxes that only slightly deviated from the ground truth were counted as incorrect. Conversely, precision slightly improved because only predictions with significant overlap were considered true positives. This behavior occurs because higher IoU thresholds require greater localization accuracy for predictions to be classified as correct.

Model errors

- **False positive:** The model detected a bicycle in the image, even though there was only a horse. This likely happened because the model confused features like fur color and size.



- ***False negative:*** The model didn't detect a horse in the corner of the image. This might be because the horse was partially hidden, making it harder to identify.



When evaluating the trade-off between precision and recall, the focus depends on the application. For a self-driving car detecting pedestrians, high recall is more important because it ensures almost all pedestrians are identified, lowering the chance of missing someone and causing an accident, even if it occasionally results in misidentifying non-pedestrians. On the other hand, a photo app that tags objects in pictures should prioritize high precision, as users prefer accurate labels over capturing every possible object, and false tags can be confusing or irritating. Using pre-trained models provides several advantages, such as shorter training times, improved performance on small datasets, and quicker experimentation, since the model already learns general features

However, they also have drawbacks, such as limited flexibility for specialized tasks, inherited biases from the original training data, and potentially large sizes that demand significant computational resources. Object detection is vital in real-world situations where knowing the exact location of objects matters. In autonomous vehicles, detecting pedestrians, vehicles, and traffic signs enables the system to navigate safely, not just recognize what is present. In retail and inventory management, detecting products on shelves allows for precise counting and placement, which classification alone cannot provide. Similarly, in medical imaging, locating tumors accurately is essential for diagnosis and treatment, while simple classification only indicates the presence of a tumor without its exact location.