Lab 09 - Object Detection using Transfer Learning and Pascal VOC 2007 Dataset

Student: Ruben Valenzuela

Course: ITAI 1378

# 1. Conceptual Understanding

• The main difference between image classification and object detection is that image classification assigns a single label to an entire image, while object detection identifies and localizes multiple objects within an image using bounding boxes. In this exercise, the difference is evident as the model predicts not just what objects are present but also where they are located.

• I used SSD MobileNet V2 for this task because it offers a good trade-off between speed and accuracy. It's lightweight and suitable for environments with limited computational resources. However, its accuracy is generally lower compared to heavier models like Faster R-CNN or EfficientDet.

# 2. Code Interpretation

• The function `find\_images\_with\_classes` helps filter out images containing specific object classes. This is useful when working with large datasets like COCO, where selecting relevant images speeds up training and evaluation.

• In the `plot\_detections` function, the threshold value (e.g., 0.5) controls which detections are shown. Only predictions with confidence scores above the threshold are displayed, which reduces noise and false positives.

• Heatmap visualization provides insight into the model’s confidence and helps identify how certain the model is about its predictions. It aids in debugging and understanding detection behavior.

# 3. Observing Results and Limitations

• The model tends to detect larger and more distinguishable objects like people and cars more accurately. Small, occluded, or overlapping objects are more challenging due to resolution and model limitations.

• Bounding boxes may be inaccurate or miss certain objects entirely, especially small ones or those in cluttered environments. Low confidence, class confusion, and dataset bias can contribute to these errors.

• Using the full Pascal VOC 2007 dataset would likely improve accuracy by exposing the model to more training examples and a wider variety of object contexts, helping it generalize better.

# 4. Critical Thinking

• To detect only animals or vehicles, we could filter the output classes or fine-tune the model with a dataset containing only those categories.

• To train our own object detection model, we would need a labeled dataset (with bounding boxes), choose a base model, and run training with tools like TensorFlow Object Detection API. Challenges include time, computing power, and annotation quality.

• Despite its limitations, this model is still useful in real-world scenarios like mobile apps, basic surveillance systems, or embedded devices, where lightweight models are preferred.

# 5. Going Further (Optional)

• I researched other models from TensorFlow Hub such as EfficientDet and Faster R-CNN. Compared to SSD MobileNet V2, they offer higher accuracy but require significantly more resources and longer inference times.

• I also ran a few test images through another pre-trained model (EfficientDet-D0). While the accuracy was slightly higher, inference was slower, confirming the trade-off between speed and precision.