

Questions for Reflection and Analysis

Conceptual Understanding:

What do you understand about the concept and main difference between image classification and object detection?

Preferably, image classification provides only one label for the entire image while object detection labels and also outlines several objects contained in an image. This difference is well illustrated in this exercise by bounding boxes that capture multiple objects rather than a single label.

Why not the proposed SSD MobileNet V2 model is used for this task of objects detection? SSD MobileNet V2 is lightweight and optimized with the ability to operate more significantly with fewer resources than its SSD MobileNet counterpart. However, this is accompanied by the tradeoff of having lower number of reliable connections compared to bulkier models especially in cases of small or the ones that overlap.

Code Interpretation:

Explain what the `find_images_with_classes` function is supposed to do.

This function makes it easier to gradually select images containing specific classes from a huge database such as COCO by tagging relevant classes.

What effect does the threshold value have in the `plot_detections` function (`threshold=0.5`)? The first level filters detections based on scores obtained from confidence evaluations. There is more true negative case since the higher threshold implies the only correct object that will be shown is those that the model is very positive about, hence some correct but low confidence detection may be left out. Less stringency implies higher number of objects but it may contain several impostors as well.

Observing Results and Limitations:

Ideally, run the exercise multiple times with different actors starting the lesson. Which objects does the model perform well; and which ones give it a hard time to identify?

The model has good performance with large and complex shaped objects such as cars and persons and low accuracy with small, partially occluded, or similar objects. This difference

is due to the relatively low feature-resolution of the model and due to the overall structure of the model, which in order to be fast enough necessarily has to make approximations.

Are the shapes of the bounding boxes ever wrong? What do you think exacerbate these inconsistencies?

Indeed, bounding boxes can either completely miss an object or contain extraneous regions that have no relevance to the object of interest. Object size, position, illumination and occlusion are crucial parameters as the model might fail to distinguish between overlapping objects as well as in photo with low contrast.

What would happen if we take not only 100 images from Pascal VOC for validation but the whole Pascal VOC 2007 dataset instead?

It finds that a bigger data will allow for better learning of robust features that will boost the accuracy of the model. The generalization is done due to the fact that in registering different objects and scenarios involved in a large dataset minimizes over-fitting.

Critical Thinking:

How could you change the code to recognize only a definite type of objects such as just animals or just vehicles?

Doing this, detections for specific categories can be filtered the `plot_detections` can be adjusted to define check if the detected object belongs to the particular class or classes we are interested in and plot only the objects that correspond to them.

What would you have to do if you wanted to train your own object detection model?

Object detection model training involves compiling a dataset and labeling it, choosing an architecture, parameter tuning, and resources. Issues involved include data collection, computation, and control of overemphasizing some factors or underemphasizing others.

Knowing the drawbacks of this model, in which practical usage will it still be applicable?

A case of real-time usage of SSD MobileNet V2 is suitable for the use in real-time applications on edge devices such as mobile phones or embedded systems where computational resources are a concern and use cases include basic surveillance or lightweight object detection applications.