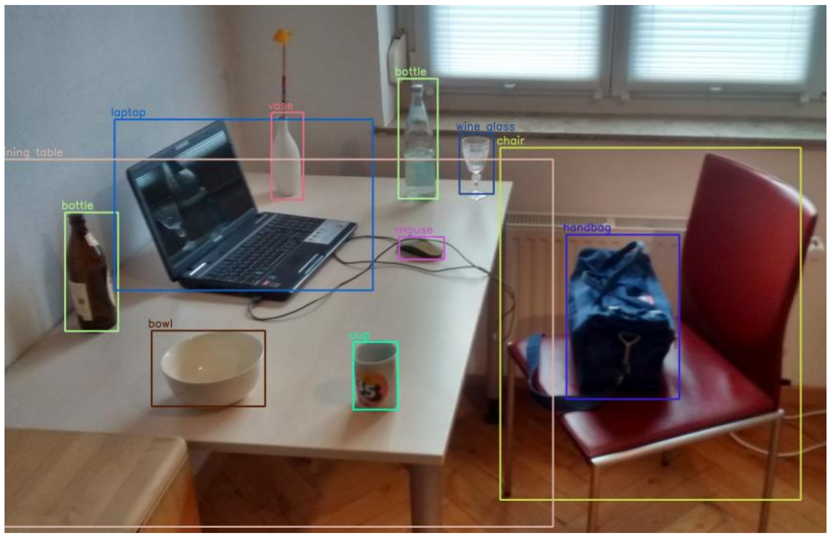
Object Detection

Overview

In the object detection problem, an algorithm locates objects in images or videos and classifies those objects. For example, objects in the following image[1] were detected using the You Only Look Once (YOLO) v3 model trained on the Microsoft Common Object in Context (COCO) data set[2].



Deep learning object detection models can be divided into one- or two-stage methods. As the name implies, two-stage methods such as Region Based Convolutional Neural Networks (R-CNN)[3] perform detection in two steps. First, the image is analyzed to identify candidate regions. Secondly, a classification network is applied to each candidate region to determine if the image corresponds to a known object and, if so, that object's class. A one-stage method performs both the localization and classification with a single network in a single pass. The YOLO model[4] is one of the most well-known and widely used single-stage methods. Two-stage methods tend to be more accurate but require significantly more computation; single-stage methods like YOLO offer the alternative set of tradeoffs.

Sophisticated models like YOLO are often quite large. These models require large amounts of data and significant compute resources to train. The training process described in the YOLO

[1] https://en.wikipedia.org/wiki/Object_detection#/media/File:Detected-with-YOLO--Schreibtisch-mit-Objekten.jpg

[2] https://cocodataset.org/

[3] Ross Girshick, Jeff Donahue, Trevor Darrell, Jitendra Malik. "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* 2014, pp. 580-587

[4] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi. "You Only Look Once: Unified, Real-Time Object Detection." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 779-788

paper involved one week of training for image classification on the large ImageNet data set[5], which contains over 1.2 million labeled training images, followed by further training specifically for object detection.

The computational resources required to train these large models is prohibitive for many researchers. Thankfully, trained versions of many of these models can be repurposed without having to retrain them from scratch. This has led to a change in how machine learning research is disseminated. Researchers no longer just publish descriptions of their models; they often publish the trained models as well.

In this project, you are going to learn how to apply a pre-trained YOLOv3 model to your own images.

Instructions

1. Review the provided notebook and associated files. The weights are saved in the same folder (yolov3.weights).
2. Run the notebook on the included images. Note that you may need to install the Python bindings for OpenCV. They are available through the opencv-python pip package.
3. Find some images of your own. These can either be your personal images or images found on the web. Run the notebook on these.
4. Read the provided YOLO paper.
5. Bonus (+10 points): run one or more other versions of YOLO and compare the results with YOLOv3.

[5] ImageNet

Lab Report
Write up your analysis in a separate report consisting of maximum 7 pages. Your report should address:

1. Evaluate the performance of the predictions on the included images. Were the objects detected? Were they correctly classified? Did the bounding boxes accurately localize the objects? (Include examples in your report as figures)

2. Evaluate the performance of the predictions on your images. Were the objects detected? Were they correctly classified? Did the bounding boxes accurately localize the objects? (Include examples in your report as figures)

3. Based on your reading of the YOLO paper, what are some of the competing models? How does YOLO differ from these models?

4. Answer the following questions about potential ethical issues that can arise.

  a. How might the large data sets violate privacy? Even if the data set is not made public, how can sharing a trained model violate privacy?

  b. Researchers have found that models trained on the ImageNet data set perform worse on people of color, and in some cases, produce offensive classifications[6]. For example, ImageNet (and similar data sets) do not include an equal number of people of different racial and ethnic backgrounds which biases the resulting model. Develop and describe some criteria you might evaluate data sets and the resulting models to check for bias and strategies you could employ to avoid or reduce that bias.

Submission Instructions
Submit your lab report as a PDF in Canvas.

[6] Kaiyu Yang, Klint Qinami, Li Fei-Fei, Jia Deng, Olga Russakovsky. "Towards Fairer Datasets: Filtering and Balancing the Distribution of the People Subtree in the ImageNet Hierarchy." *Conference on Fairness, Accountability, and Transparency (FAT*),* 2020.