Report on COCO dataset

Microsoft Common Objects in Context (COCO) dataset is the for evaluating the performance of state-of-the-art computer vision models. COCO contains over 330,000 images, of which more than 200,000 are labelled, across dozens of categories of objects. COCO is a collaborative project maintained by computer vision professionals from numerous prestigious institutions, including Google, Caltech, and Georgia Tech.

The COCO dataset is designed to represent a vast array of things that we regularly encounter in everyday life, from vehicles like bikes to animals like dogs to people.

The COCO dataset contains images from over 80 "object" and 91 generic "stuff" categories, which means the dataset can be used for benchmarking general-purpose models more effectively than small-scale datasets.

In addition, the COCO dataset contains:

- 121,408 images
- 883,331 object annotations
- 80 classes of data
- The median image ratio is 640 x 480

There are multiple COCO datasets, each one made for a specific machine learning task, with additional data. The 3 most popular tasks are:

1. object detection — model should get bounding boxes for objects, i.e., return list of object classes and coordinates of rectangles around them; objects (also called “things”) are discrete, separate objects, often with parts, like humans and cars; the official dataset for this task also contains additional data for object segmentation
2. object/instance segmentation — model should get not only bounding boxes for objects (instances/ “things”), but also segmentation masks, i. e. coordinates of polygon closely around the object
3. stuff segmentation — model should do object segmentation, but not on separate objects (“things”), but on background continuous patterns like grass or sky

In computer vision, those tasks have tremendous usage, e. g. for self-driving vehicles (detection of people and other vehicles), AI-based security (human detection and/or segmentation) and object re-identification (object segmentation or removing background with stuff segmentation helps with checking object identity).