MICROSOFT (MS) COCO: COMMON OBJECTS IN CONTEXT PAPER SUMMARY

MS COCO-IN GENERAL:

COCO is large-scale object detection, segmentation, and captioning dataset. Microsoft's Common Objects in Context dataset (COCO) is the most popular object detection dataset at the moment. It is widely used to benchmark the performance of computer vision methods.

COCO has several features:

- Object segmentation
- Recognition in context
- Super pixel stuff segmentation
- 330K images (>200K labelled)
- 1.5 million object instances
- 80 object categories
- 91 stuff categories
- 5 captions per image
- 250,000 people with key points

THE ARTICLE COVERS:

1. ABOUT MS COCO DATASET:

The Microsoft Common Objects in Context (MS COCO) dataset contains 91 common object categories with 82 of them having more than 5,000 labelled instances. In total the dataset has 2,500,000 labelled instances in 328,000 images.

In contrast to the popular ImageNet dataset, COCO has fewer categories but more instances per category. This can aid in learning detailed object models capable of precise 2D localization. The dataset is also significantly larger in the number of instances per category than the PASCAL VOC and SUN datasets.

2. MS COCO DATASET PREPARATION:

The following pre-trained 80, things are included in the COCO dataset classes for object recognition and tracking. Each of the 17 pre-trained key points in the COCO is annotated with three values. The coordinates are marked by the x and y values, while v indicates whether the key point is visible. There are annotated images of typical scenarios with typical objects in their natural settings. Pre-defined classes like chairs or bananas identify these things. Labelling is a common approach in computer vision, also called image annotation. Other object recognition datasets focus on object bounding-box localisation, image classification, segmenting objects at the semantic pixel level, and segmenting specific object instances. A wide variety of object categories views are available.

3. CHALLENGES FACED WHILE MS COCO DATASET PREPARATION:

Data Biasness. There are some bias in the data because more light-skinned people are represented in the dataset. There are more photographs of light-skinned people compared to dark-skinned people, twice as many men as women, and even fewer dark-skinned women. Several of the image descriptions contain racial connotations. This could lead to computer vision

being used to describe people in a socially unacceptable way. The researchers cite numerous studies that show model performance favours fair-skinned people, and that image caption systems developed using the COCO dataset perform better for fair-skinned people in tasks like pedestrian identification and facial recognition. There is prejudice in the image context. Lighter-skinned people appear indoors with furniture in the background, while darker-skinned people appear outdoors with moving objects in the background.

4. RECENT DEVELOPMENTS IN THE COCO DATASET AND ITS MODEL.

The first version of MS COCO dataset was released in 2014. It contains 164K images split into training (83K), validation (41K) and test (41K) sets. In 2015 additional test set of 81K images was released, including all the previous test images and 40K new images.