**COMMON OBJECTS IN CONTEXT(COCO) PAPER SUMMARY**

**Abstract:-**

Large-scale dataset for object detection, segmentation, and captioning is called COCO. Currently, the most well-liked object detection dataset is Microsoft's Common Objects in Context dataset (COCO). It is frequently employed to evaluate the effectiveness of computer vision techniques.

**Features Of COCO model:-**

* 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

**COCO DATASET:-**

There are 91 common object categories in the Microsoft Common Objects in Context (MS COCO) dataset, with 82 of them having more than 5,000 tagged instances. The collection contains 328,000 photos and 2,500,000 instances that have labels.

COCO includes fewer categories but more cases per category than the well-known ImageNet dataset. Learning intricate object models with accurate 2D localisation can benefit from doing so. The dataset's instance count per category is likewise much higher than that of the PASCAL VOC and SUN datasets.

**COCO DATASET PREPARATION**:-

The COCO dataset classes for object detection and tracking contain the following 80 pre-trained items. The 17 pre-trained critical points in the COCO each have three values labelled on them. The x and y values represent the coordinates, while v denotes whether the key point can be seen. Images of common situations and typical things in their natural contexts have annotations.

These goods are categorised as chairs or bananas, for example. An strategy used frequently in computer vision is labelling, sometimes known as image annotation. Other object recognition datasets concentrate on localising objects within their bounding boxes, classifying images, segmenting objects at the semantic pixel level, and segmenting particular instances of objects. There are many different object category views available.

**CHALLENGES FACED WHILE MS COCO DATASET PREPARATION:-**

Data prejudice. Because more light-skinned persons are represented in the sample, there is some bias in the data. There are twice as many men as women in the images, more light-skinned people than dark-skinned people, and even fewer dark-skinned women. Racial undertones can be found in a few of the image descriptions. This could result in the socially inappropriate use of computer vision to describe individuals. The researchers cite multiple studies that demonstrate fair-skinned people do better than dark-skinned people in tasks like pedestrian detection and facial recognition, and that image caption systems developed using the COCO dataset perform better for fair-skinned people. The context of the photograph contains prejudice. Darker people appear outdoors with moving items in the background, whereas lighter people appear indoors with furniture in the background.

**RECENT DEVELOPMENTS IN THE COCO DATASET AND ITS MODEL:-**

In 2014, the MS COCO dataset's initial version was made public. There are 164K photos total, divided into 83K training images, 41K validation images, and 41K test images. A second test set of 81K photos was released in 2015,including the 40K fresh photographs and all of the test images from the past.