COCO (Common Objects in Context) là một tập dữ liệu lớn cho việc phát hiện và nhận dạng vật thể trong ảnh. COCO bao gồm hơn 330.000 hình ảnh được gán nhãn với hơn 2.5 triệu đối tượng thuộc 80 loại khác nhau. Bên cạnh đó, COCO còn có thông tin về mối liên hệ giữa các đối tượng và định vị vị trí của chúng trong không gian 2D.

COCO có 5 loại Annotation được sử dụng cho các bài toán:

* Object detection.
* Keypoint detection.
* Stuff segmentation.
* Panoptic segmentation.
* Image captioning.

COCO lưu trữ các dữ liệu annotatión trong 1 file có định dạng JSON. Với 1 file dữ liệu JSON của COCO sẽ gồm những block chính như sau:

* **info**: Trường dữ liệu này chứa thông tin về tập dataset.
* **licenses**: Trường dữ liệu này chứa thông tin về giấy phép của tập dataset
* **categories**: Trường dữ liệu này chứa một danh sách tên các đối tượng
* **images**: Trường dữ liệu này lưu trữ thông tin của các hình ảnh như đường dẫn đến ảnh, tên hình ảnh, id của ảnh ( mỗi ảnh có 1 id duy nhất).
* **annotations**: Trường dữ liệu này chứa thông tin về các bouding boxes của các đối tượng trong 1 ảnh đối với bài toán object detection, thông tin về segmentation cho bài toán segment hoặc thông tin về caption cho bài toán image captioning.

{

"info": info,

"licenses":[lisenses],

"categories":[category],

"images":[image],

"annotations":[annotation]

}

**info**

Mô tả thông tin của tập datasets, Dưới đây là một ví dụ:

"info": {

"description": "COCO 2017 Dataset",

"url": "http://cocodataset.org",

"version": "1.0",

"year": 2017,

"contributor": "COCO Consortium",

"date\_created": "2017/09/01"

}

#### licenses

Phần này bao gồm các thông tin liên quan đến giấy phép sử dụng của tập datasets. Thông thường thì mọi người download và sử dụng thường thì sẽ không quan tâm nhiều lắm đến mục này. Tuy nhiên nếu sử dụng cho các sản phẩm thương mại thì nên chú ý mục này khi sử dụng tập datasets.

"licenses": [

{

"url": "http://creativecommons.org/licenses/by-nc-sa/2.0/",

"id": 1,

"name": "Attribution-NonCommercial-ShareAlike License"

},

{

"url": "http://creativecommons.org/licenses/by-nc/2.0/",

"id": 2,

"name": "Attribution-NonCommercial License"

},

...

]

#### categories

Trường dữ liệu chứa danh sách tên và id của các đối tượng xuất hiện trong tập dữ liệu. Mỗi đối tượng gồm các thông tin là id, tên và có thể có thêm tên của đối tượng cha của nó:

"categories":

[

{"id": 1, "name": "cat" , "supercategory": "animal"},

{"id": 2, "name": "dog" , "supercategory": "animal"},

{"id": 3, "name": "apple" , "supercategory": "fruit"}

]

#### images

Trường dữ liệu này chứa các thông tin của ảnh như: id (mỗi ảnh chỉ có 1 id duy nhất), tên ảnh, chiều cao, chiều rộng...

"images":

[

{

"id": 297,

"width": 1366,

"height": 768,

"filename": "chien.png",

"license": 1,

"date\_captured": "2021-05-07 16:24:30"

},

{

"id": 305,

"width": 1920,

"height": 1080,

"filename": "yoona.png",

"license": 2,

"date\_captured": "2021-05-07 16:25:30"

}

]

#### annotations

Trường dữ liệu này chứa thông tin nhãn của từng đối tượng trong ảnh, gồm các thông tin như sau:

annotation{

"id": int,

"image\_id": int,

"category\_id": int,

"segmentation": [polygon],

"area": float,

"bbox": [x, y, width, height],

"iscrowd": 0 or 1,

}

Trong đó:

* **id**: id của annotation.
* **image\_id**: id của ảnh chứa object .
* **category\_id**: id của đối tượng (danh sách các id đã được liệt kê ở trường dữ liệu **categories**.
* **segmentation**: chứa thông tin về tọa độ x, y cho các đa giác bao quanh đối tượng như đề cập ở phần trên.
* **area**: diện tích bounding box của đối tượng
* **bbox**: tọa độ của bounding box tương ứng với đối tượng ( x\_top left, y-top left, width, height).
* **iscrowd**: là giá trị thể hiện thông tin ở trường segmentation là của một đối tượng riêng lẻ hay là của cả 1 nhóm đối tượng.