

主讲人：正点原子团队

硬件平台：正点原子ATK-DLRV1126开发板

版权所有：广州市星翼电子科技有限公司

资料下载：[www.openedv.com/docs/index.html](http://www.openedv.com/docs/index.html)

教学平台：[www.yuanzige.com](http://www.yuanzige.com)

天猫店铺：[zhengdianyuanzi.tmall.com](http://zhengdianyuanzi.tmall.com)

技术论坛：[www.openedv.com/forum.php](http://www.openedv.com/forum.php)

公众平台：正点原子



## ■ 数据集标注的格式

- 1、常见数据集标注格式
- 2、COCO格式转化为YOLO格式
- 3、VOC格式转化为YOLO格式

## 1、常见数据集标注格式

标签包括：边界框信息 和 类别标签信息

边界框信息告诉模型目标在图像中的什么位置；类别标签信息告诉模型图像中的某个位置所对应的目标是什么类别物体。

常见数据集格式	说明
YOLO	<p>txt文件，标注的信息为：</p> <div><div><p>&lt;class_id&gt; 第一列，物体的类别ID</p><p>&lt;center_x&gt; 第二列，物体边界框的中心点X坐标（归一化到图像宽度）</p><p>&lt;center_y&gt; 第三列，物体边界框的中心点Y坐标（归一化到图像高度）</p><p>&lt;width&gt; 第四列，物体边界框的宽W（归一化到图像宽度）</p><p>&lt;height&gt; 第五列，物体边界框的高H（归一化到图像高度）</p></div><div>}</div><div>ground truth boxes（简写gtbox）</div></div>
PASCAL VOC	<p>xml文件，20类物体，链接：<a href="http://host.robots.ox.ac.uk/pascal/VOC">http://host.robots.ox.ac.uk/pascal/VOC</a></p> <p>标注的位置信息<b>bndbox</b>： 物体的类别名、物体边界框左上角X坐标、物体边界框左上角Y坐标、物体边界框右下角X坐标、物体边界框右下角Y坐标：</p> <p>&lt;class_name&gt;、&lt;xmin&gt;、&lt;ymin&gt;、&lt;xmax&gt;、&lt;ymax&gt;</p>
MS COCO	<p>json文件，80类物体，链接：<a href="https://cocodataset.org">https://cocodataset.org</a></p> <p>标注的位置信息<b>bbox</b>： 物体边界框左上角X坐标、物体边界框左上角Y坐标、物体边界框的宽、物体边界框的高</p>

## 1、常见数据集标注格式--YOLO格式

标签包括：边界框信息 和 类别标签信息

边界框信息告诉模型目标在图像中的什么位置；类别标签信息告诉模型图像中的某个位置所对应的目标是什么类别物体。



图像

zidane.txt - 记事本				
文件(F)	编辑(E)	格式(O)	查看(V)	帮助(H)
0	0.481719	0.634028	0.690625	0.713278
0	0.741094	0.524306	0.314750	0.933389
27	0.364844	0.795833	0.078125	0.400000

标签文件的内容

训练的数据集 = 图像 + 图像对应的标签文件（1张图像对应1个标签文件）

## 1、常见数据集标注格式--YOLO格式

zidane.txt - 记事本				
文件(F)	编辑(E)	格式(O)	查看(V)	帮助(H)
0	0.481719	0.634028	0.690625	0.713278
0	0.741094	0.524306	0.314750	0.933389
27	0.364844	0.795833	0.078125	0.400000

标签文件的内容

标签文件中的数据：

第一列：物体的类别ID（在COCO数据集中0表示person，27表示tie）

第二列：边界框所对应的 中心点X坐标

第三列：边界框所对应的 中心点Y坐标

第四列：边界框所对应的 宽W

第五列：边界框所对应的 高H



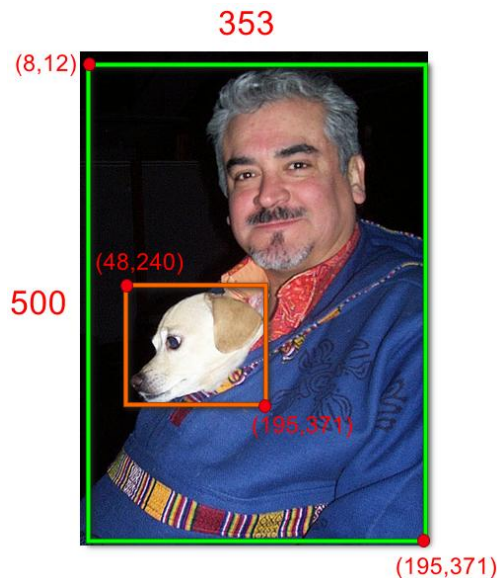
标签标识的含义

训练集上标注的边界框（bounding box，简写bbox），通常叫做ground truth boxes（简写gtbox），检测得到的框，通常叫做prior boxes



## 1、常见数据集标注格式--VOC 格式

PASCAL VOC 2007: <http://host.robots.ox.ac.uk/pascal/VOC/voc2007/index.html>

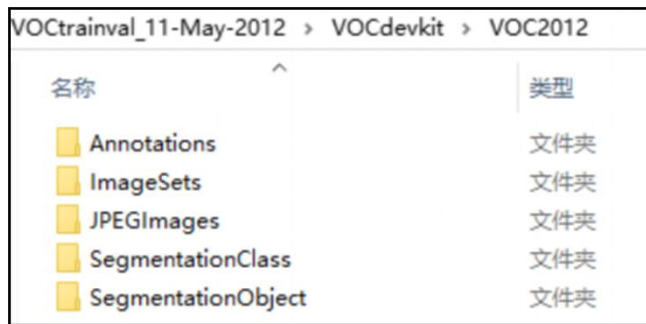


```
<object>
  <name>dog</name>
  <pose>Left</pose>
  <truncated>1</truncated>
  <difficult>0</difficult>
  <bndbox>
    <xmin>48</xmin>
    <ymin>240</ymin>
    <xmax>195</xmax>
    <ymax>371</ymax>
  </bndbox>
</object>
<object>
  <name>person</name>
  <pose>Left</pose>
  <truncated>1</truncated>
  <difficult>0</difficult>
  <bndbox>
    <xmin>8</xmin>
    <ymin>12</ymin>
    <xmax>352</xmax>
    <ymax>498</ymax>
  </bndbox>
</object>
```

PASCAL VOC数据集标注，坐标点：左上角坐标点 + 右下角坐标点

## 1、常见数据集标注格式--VOC 格式

PASCAL VOC 2012数据集: <http://host.robots.ox.ac.uk/pascal/VOC/>  
<http://host.robots.ox.ac.uk/pascal/VOC/voc2012/index.html>  
[http://host.robots.ox.ac.uk/pascal/VOC/voc2012/VOCtrainval\\_11-May-2012.tar](http://host.robots.ox.ac.uk/pascal/VOC/voc2012/VOCtrainval_11-May-2012.tar)



```
VOCdevkit
├── VOC2012
│   ├── Annotations                # 所有的图片的标注信息(.xml文件)
│   ├── ImageSets
│   │   ├── Action                # 人的行为动作图像信息
│   │   ├── Layout                # 人的各个部位图像信息
│   │   ├── Main
│   │   │   ├── train.txt         # 目标检测分类图片信息
│   │   │   ├── val.txt          # 训练集
│   │   │   └── trainval.txt      # 验证集
│   │   │                           # 训练集+验证集
│   │   └── Segmentation
│   │       ├── train.txt         # 目标分割图像信息
│   │       ├── val.txt          # 训练集
│   │       └── trainval.txt      # 验证集
│   │                               # 训练集+验证集
│   ├── JPEGImages                # 所有图像文件
│   ├── SegmentationClass          # 语义分割png图 (基于类别)
│   └── SegmentationObject         # 实例分割png图 (基于目标)
```



## 1、常见数据集标注格式--COCO格式

MS COCO数据集：<https://cocodataset.org>

训练集：<http://images.cocodataset.org/zips/train2017.zip>

验证集：<http://images.cocodataset.org/zips/val2017.zip>

测试集：<http://images.cocodataset.org/zips/test2017.zip>

标注信息：[http://images.cocodataset.org/annotations/annotations\\_trainval2017.zip](http://images.cocodataset.org/annotations/annotations_trainval2017.zip)

常见数据集格式	说明
image captions (看图说话)	描述信息通常用于图像字幕 (image captioning) 任务，即生成对图像内容的自然语言描述
object instances (目标实例)	描述信息包括目标的类别 (如人、车、动物等) 以及每个目标在图像中的位置信息，通常以边界框坐标的形式表示，这些信息对于目标检测任务至关重要
object keypoints (目标上的关键点)	描述信息包括人体关键点的位置信息，如鼻子、眼睛、手等关键点的坐标，常用于人体姿态估计任务中

## 1、常见数据集标注格式--COCO格式

```
{
  "info": {info},           # 数据集基本信息（年份、版本、作者、日期等）
  "licenses": [license],    # license
  "images": [image],        # 图片名字和图片路径（文件名、图像id、图像的宽高、图像的url地址）
  "annotations": [annataton], # 物体分割或目标检测标注数据
  "categories": [category]  # 物体的类别名和对应的类别ID
}
```

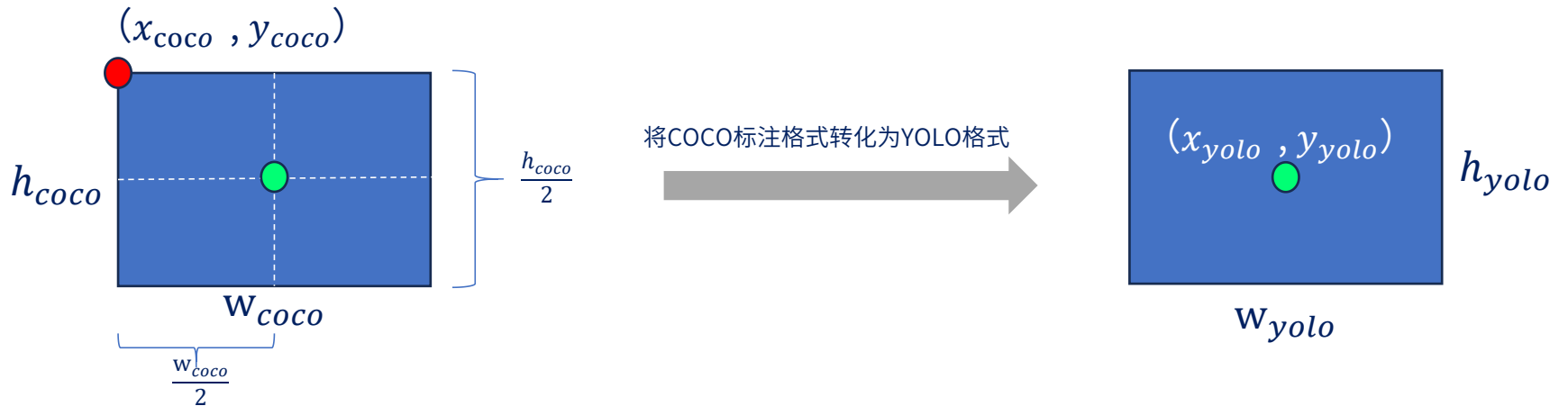
```
"images": [
{ "license": 3,                # license
  "file_name": "000000391895.jpg", # 图片文件名
  "coco_url": "http://images.cocodataset.org/train2017/000000391895.jpg", # coco图片链接url地址
  "height": 360,              # 图片的宽
  "width": 640,               # 图片的高
  "date_captured": "2013-11-14 11:18:45", "flickr_url": # 获取日期
"http://farm9.staticflickr.com/8186/8119368305_4e622c8349_z.jpg", "id": 391895 # flick图片链接url
}]
```

## 1、常见数据集标注格式--COCO格式

```
"annotations": [  
  { "segmentation": [[239.97,260.24,222.04,270.49.....]], # 物体分割位置信息  
    "area": 2765.1486500000005, # 被检测物体的面积  
    "iscrowd": 0, # 物体是否被遮盖（0表示否，1表示被遮盖）  
    "image_id": 558840, # 该物体所在图片的编号  
  
    # 目标检测框的坐标信息（[x, y, width, height]，其中x表示左上角x坐标，y表示左上角y坐标）  
    "bbox": [199.84,200.46,77.71,70.88],  
  
    "category_id": 58, # 被标记物体的类别id编号  
    "id": 156 # 图片中每个被标记物体的id编号  
  }  
]
```

```
"categories": [  
  { "supercategory": "person", # 类别所属的大类，如卡车和轿车都属于机动车这个类  
    "id": 1, # 类别id号  
    "name": "person", # 物体类别名  
  },.....]
```

## 2、COCO格式转化为YOLO格式



$$x_{coco} = \text{box}[0]$$

$$y_{coco} = \text{box}[1]$$

$$h_{coco} = \text{box}[2]$$

$$w_{coco} = \text{box}[3]$$

转化公式

$$x_{yolo} = x_{coco} + \frac{w_{coco}}{2}$$

$$y_{yolo} = y_{coco} + \frac{h_{coco}}{2}$$

$$h_{yolo} = h_{coco}$$

$$w_{yolo} = w_{coco}$$

归一化

$$x_{yolo} = \frac{(x_{coco} + \frac{w_{coco}}{2})}{img\_width}$$

$$y_{yolo} = \frac{y_{coco} + \frac{h_{coco}}{2}}{img\_width}$$

$$h_{yolo} = \frac{h_{coco}}{img\_width}$$

$$w_{yolo} = \frac{w_{coco}}{img\_width}$$

参考课程中的atk\_coco\_to\_yolo\_step1.py文件

## 3、VOC格式转化为YOLO格式



$$x_{\min} = \text{box}[0]$$

$$y_{\min} = \text{box}[2]$$

$$x_{\max} = \text{box}[1]$$

$$y_{\max} = \text{box}[3]$$

转化公式

$$x_{yolo} = \frac{x_{\min} + x_{\max}}{2} - 1$$

$$y_{yolo} = \frac{y_{\min} + y_{\max}}{2} - 1$$

$$h_{yolo} = y_{\max} - y_{\min}$$

$$w_{yolo} = x_{\max} - x_{\min}$$

归一化

$$x_{yolo} = \frac{(\frac{x_{\min} + x_{\max}}{2} - 1)}{\text{img\_width}}$$

$$y_{yolo} = \frac{\frac{y_{\min} + y_{\max}}{2} - 1}{\text{img\_width}}$$

$$h_{yolo} = \frac{y_{\max} - y_{\min}}{\text{img\_width}}$$

$$w_{yolo} = \frac{x_{\max} - x_{\min}}{\text{img\_width}}$$

参考课程中的atk\_voc\_to\_yolo\_step2.py文件

## 3、VOC格式转化为YOLO格式







版权所有：广州市星翼电子科技有限公司  
天猫店铺：<https://zhengdianyuanyi.tmall.com>