# VOC2COCO: VOC数据集转COCO数据集

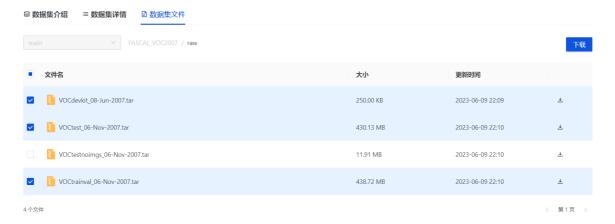
### 参考资料

https://mmdetection.readthedocs.io/zh-cn/latest/user\_guides/useful\_tools.html#id17

#### 数据集下载

https://opendatalab.com/OpenDataLab/PASCAL\_VOC2007/tree/main?source=OpenMMLab%20Git Hub

1.把raw文件夹下的这三个文件夹下载下来



## 2.解压三个文件夹

```
tar -xvf VOCdevkit_08-Jun-2007.tar
tar -xvf VOCtest_06-Nov-2007.tar
tar -xvf VOCtrainval_06-Nov-2007.tar
```

然后三个压缩包里的东西都会被解压到一个 vocdevkit 文件夹中

#### 3.运行代码

```
python tools/dataset_converters/pascal_voc.py ${DEVKIT_PATH} [-h] [-o
${OUT_DIR}]
```

例:

```
python tools/dataset_converters/pascal_voc.py /data/bzx_yjy/VOC2007/VOCdevkit -o
.
```

默认的转换格式是pkl,改成coco或者在命令行里加一个 --out-format coco

```
def parse_args():
    parser = argparse.ArgumentParser(
        description='Convert PASCAL VOC annotations to mmdetection format')
    parser.add_argument('devkit_path', help='pascal voc devkit path')
    parser.add_argument('-o', '--out-dir', help='output path')
    parser.add_argument(
        '--out-format',
        default='coco",
        choices=('pkl', 'coco'),
        help='output format, "coco" indicates coco annotation format')
    args = parser.parse_args()
    return args
```

DEVKIT\_PATH换成VOCdevkit文件夹所在地址

4.生成的voc07\_train.json等json文件放在一个annotations里

图片在 VOCdevkit/VOC2007/JPEGImages 中

可以再用一些脚本把图片放到train、test、val里,也可以改一下读取图片的地址,统一换到 [] PEGI mages 中即可。