

一、准备工作

- **/home/zr/data/VOCdevkit/下新建文件夹 MyWork**

1. **MyWork** 文件夹下新建 Annotations 、ImageSets、 JPEGImages 、 Labels 四个文件夹；
其中 JPEGImages 和 Labels 两个文件夹内各自有子文件夹 1 ， 用于 LabelTool 标记时候，输入到 Image Dir: 栏目下，即可导入 JPEGImages/1 文件夹中仍未进行标记的图片，
标记后的标签文件同步保存在 Labels/1 文件夹下，txt 文件命名与图片名称相同。
2. ImageSets 文件夹下包含 Main 子文件夹。

- **/home/caffe/data/下新建文件夹 MyWork**

1. 该文件夹下包含 create_data_list.sh 、 create_data_crew.sh 和 labelmap_crew.prototxt 三个文件，根据需求进行适应性修改。

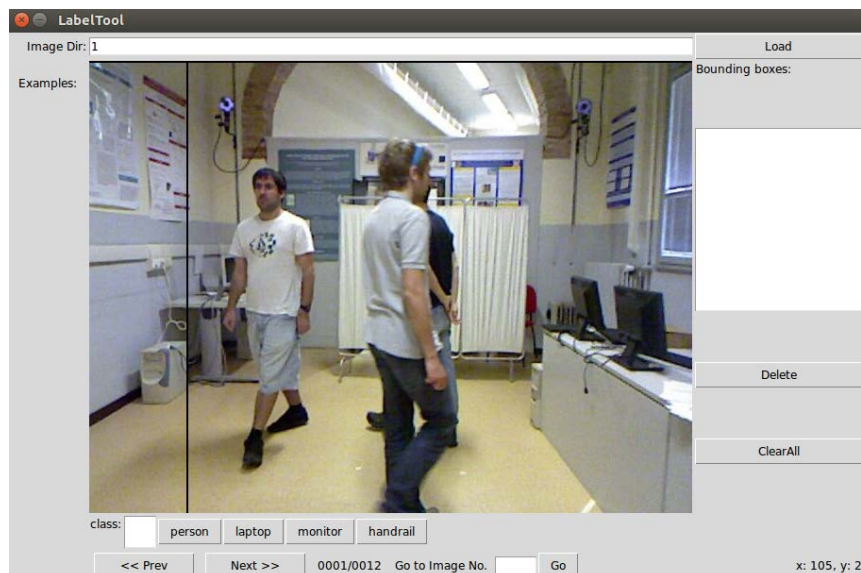
二、数据集的标注

- 1、使用路径/home/zr/data/VOCdevkit/MyWork 目录下的数据集标注工具 SSD_CrewDetect.py 来实现；

- 2、切到所在路径/home/zr/data/VOCdevkit/MyWork 目录下，执行指令

```
python SSD_CrewDetect.py
```

可打开标注软件如下图：



在 Image dir 下输入 1，点击 load，完成未标注的图片的加载。

然后选择 class 为 person，再进行画框，画完后，点击 Next>>则完成保存，继续后续未标注图片的标注。

注意：

(1) 先选择类别，再画框；

(2) 一张图片标注完后，点击 **Next>>**按钮，标注下一张图片，图片 **label** 成功后,会在 **Labels** 对应目录下生成与图片文件名对应的 **label** 文件，如果点击 **Prev>>**则会将之前标注好的图片清空，则需要重新标注；

(3) 如果目录中的图片已经标注过,点击 **load** 时不会被重新加载进来；方便中断，下次可继续标注未标注图片。

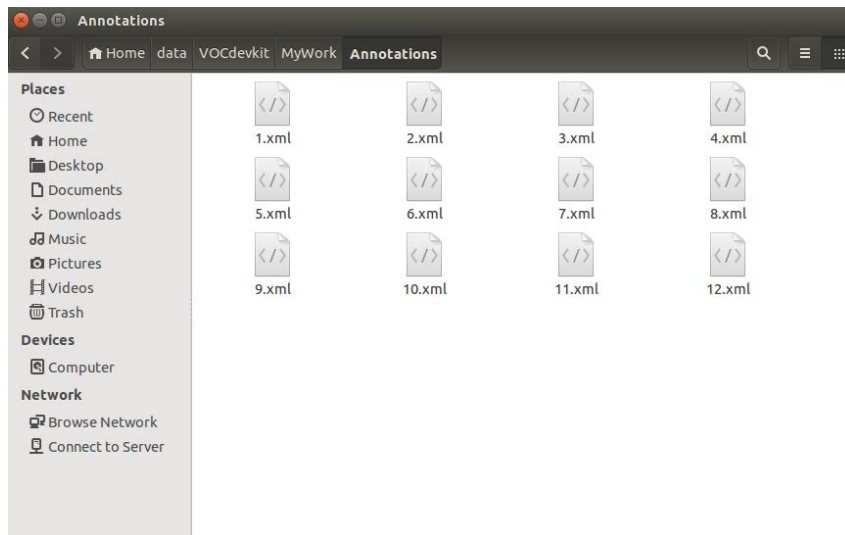
生成的 **txt** 文档存储的标签内容格式为：

```
object_number  
className xmin ymin xmax ymax  
classname x2min y2min x2max y2max  
...
```

三、数据集转换为 xml 格式

1、使用路径/home/zr/data/VOCdevkit/MyWork 目录下的 createXml.py 程序来实现；

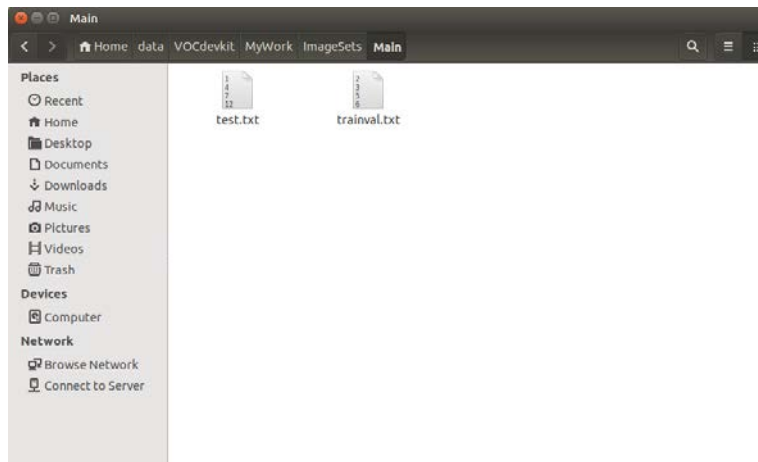
2、切换到/home/zr/data/VOCdevkit/MyWork 目录下，执行指令 `./createXml.py` 实现数据转化为 xml 格式，可见在/home/zr/data/VOCdevkit/MyWork/Annotations 目录下生成了对应的.xml 格式文件。



四、生成训练集和测试集标识文件

1、使用路径/home/zr/data/VOCdevkit/MyWork 目录下 createTest.py 程序来实现；

2、切换到/home/zr/data/VOCdevkit/MyWork 目录下，执行指令 `./createTest.py 1 3250 800` 实现训练、测试数据集标识文件生成，可见在/home/zr/data/VOCdevkit/MyWork/ImageSets/Main 目录下生成了标识文件。命令参数中，1 表示起始图片序号，3250 表示结束图片序号，800 为测试集的图像数目。

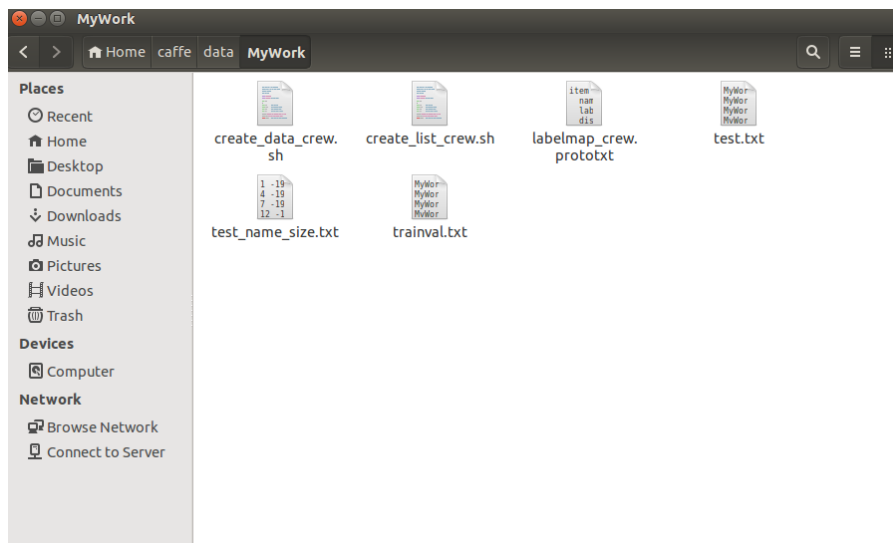


五、生成 Imdb 数据集

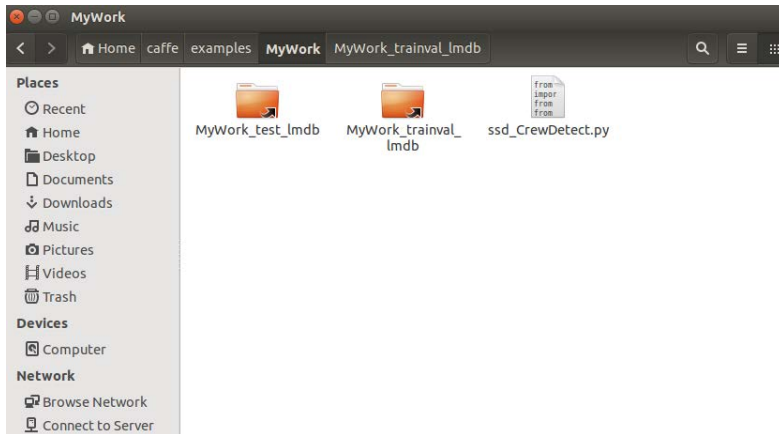
1、使用/home/zr/caffe/data /MyWork 下的 create_list_crew.sh 和 create_data_crew.sh 程序实现；

2、使用路径/home/zr/data/VOCdevkit/MyWork 路径：

(1) 先运行 `./create_list_crew.sh` 指令，在 `/home/zr/caffe/data/MyWork` 目录下生成了 `trainval.txt` 、 `test.txt` 、 `test_name_size.txt` 三个文件。



(2) 再运行 `./create_data_crew.sh` 指令，在 `/home/zr/data/VOCdevkit/MyWork/lmdb` 目录下生成了 `MyWork_trainval_lmdb` 文件夹，里面包含了 `data.mdb` 、 `lock.mdb` 两个文件；生成了 `MyWork_test_lmdb` 文件夹，里面同样包含了 `data.mdb` 、 `lock.mdb` 两个文件；同时在 `/home/zr/caffe/examples/MyWork` 下生成了 Imdb 数据集的链接（不一定需要，因为训练网络中数据集可直接地址定位到 `/home/zr/data/VOCdevkit/MyWork/lmdb` 地址下）。



六、修改 labelmap_crew.prototxt 文件

修改后为：

```
item {  
    name: "none_of_the_above"  
    label: 0  
    display_name: "background"  
}  
  
item {  
    name: "person"  
    label: 1  
    display_name: "person"  
}
```

注意: labelmap_crew.prototxt 文件中使用的名称, 应该和数据集标注工具 SSD_CrewDetect.py 使用的标签名称完全一致, 包括大小写!

至此, 完成自己数据集制作, 可用其进行网络的训练应用!