

AI图像数据处理岗位小测试

请在GitHub里创建一个Repository，将答案写在里面，提交只要提交此Repository的链接即可。答案的每次新增、删除或修改都记得commit到GitHub里。

第一题：data/目录下的包含了若干张图片以及其对应的数据文件，请用Python脚本批量处理这些数据，实现以下需求：

- (1) 解析数据文件里的json数据，获取人脸矩形框key为"left"、"top"、"right"和"bottom"的值，将图片按照此人脸矩形框的数值进行截图，并将截图保存到新的目录clip/下
- (2) 依旧是解析该json数据，获取里面的key为"pointsArray"的值，该值包含了106个人脸关键点的位置信息，按照这些点的位置信息，将点画在原图上并保存到新的目录data_106/下
- (3) 继续第(2)小题，将那些106个点画在第(1)题截图后的图片上，图片的宽和高放大到原来的两倍，并保存在新的目录clip_106/下

第二题：编写Python脚本遍历解析data/目录下的所有xml文件，统计标签信息。

已知每个xml文件包含了一个或者多个标签矩形框位置信息，每个标签信息记录在"<object>"中，" <name>"中的值为该标签的名字，"<xmin>"、"<ymin>"为该标签矩形框的左上角点坐标，" <xmax>"、"<ymax>"为该标签矩形框的右下角点坐标。

现在需要按照标签名统计每个标签的信息包括（单位可忽略）：

- 总数
- 平均面积
- 最大面积
- 面积的方差

将统计的标签信息记录在一个txt文件里。

第三题：请在文件夹里创建一个Django项目，此web app能够满足以下需求：

- 每天9:00定时爬取豆瓣厦门正在上映的电影信息 <https://movie.douban.com/cinema/nowplaying/xiamen/>。
- 爬取信息包括：片名、导演、编剧、主演、类型、制片国家/地区、语言、上映日期、片长、又名、IMDb链接、评分、购票链接。保存到数据库中。
- 用户可以在web app 主页里查看正在上映的电影信息。可以点击购票按钮进行购票。
- 将此web app打包成docker镜像。
- 用户可以切换不同地区。（选做）

