1、图片重命名和json文件生成。

图片重命名：3位日期编码+4位序号编码共七位，统一为jpg格式。

识别来自Eagle的json标注文件，将tags和color信息写入其他信息。

json中的信息包括：

def get\_name\_str():

# time\_str : 2002\_10\_20 | 可以使用 time.time() 获取

#

# json

\* 解析

\* 核对图片信息，长宽

\* json 信息相加，有问题主动报错

\*

2、其他后处理功能：导出不同格式的数据集。

导出(1个或多个)固定标签的数据，输出格式支持：xml，json，coco，voc，分类数据。

\* 格式的支持存在问题，xml 是不支分割的,在导出分割相关的数据集的时候，需要进行处理

\*

3、合并标注文件

导入文件名为唯一编码的xml或json文件，将其中的标注，写入对应的唯一json文件中。若编码不对，则报错。标注框基于编号。

\* 需要考虑两个问题

\* 核对

\* 去重

4、小图导出

根据方框标签，导出截出的小图，输出格式同样需支持上述5种格式，自动完成坐标转换。小图重命名，命名规则为：父类唯一编码\_标注框序号。并在唯一json中的bbox下，新增一个子节点：<导出>：{类别，工程}

\*

5、训练情况记录

需保存每一条导出数据的操作做出记录。

若有版本号，则工程师将访问导出这批数据的记录，将参与此次工程的数据们补写参与训练的信息：键值对：训练：工程名-日期。工程师需提供工程名+版本号即可。

同时留下一个训练信息：日期，工程名，数据索引列表。

6、整合标注的检测功能，目前数据组已在使用中，可整合进平台，上传数据后则自动检测。

7、统计每日导出的数据个数。唯一编码每天是有限额的。预计4W6张-167万张。

8、重复图片检测功能，逐步剔除原始数据中的不当存在。

9、缓存池机制。导入的数据，暂不并入数据库。需经审核后才可正式入库。防止导入导出发生错误，支持导入数据的撤回操作。注意被撤回的图片，其编码也不会被回收，保证编码的唯一性。

10、数据集传递功能。在平台上，可分享已导出的数据集的唯一编码列表。目的是接受同一项目的工程师，可实时同步数据集变更情况。

11、导出数据，即可下载到本地也可在后台自动传输至所需服务器的文件夹上：例如：101，132，80的某目录内。