发票识别

SVN项目名称: ml-invoice-recognition-v2

A800目录: /data/machine\_learning/szb/ml-invoice-recognition-v2

现有环境：A800---->invoice

python>=3.8

如需自己安装环境

1.安装paddlepaddle

https://www.paddlepaddle.org.cn/install/quick?docurl=/documentation/docs/zh/install/pip/linux-pip.html

cpu版本:

python -m pip install paddlepaddle==2.6.1 -i https://pypi.tuna.tsinghua.edu.cn/simple

gpu版本:

1查看cuda版本

2安装 CUDA11.x 包含 cuDNN 动态链接库的 PaddlePaddle（目前安装的11.6）

python3 -m pip install paddlepaddle-gpu==2.6.1.post116 -f https://www.paddlepaddle.org.cn/whl/linux/cudnnin/stable.html

2.安装paddleocr

pip install paddleocr

3.安装pandas（运行ocr.py脚本，如有报错，自行补全其他包）

修改ocr.py脚本中main函数的root\_dir并运行

例如把root\_dir = '../data/202403账期'

修改为root\_dir = '../data/202404账期'

后面月份以此类推

输出目录: '../data/202403账期识别后'

程序运行完，查看'../data/202403账期识别后/202403账期.xlsx'，筛选有问题的发票

查看'../data/202403账期识别后/problem'

把缺失信息补全到'../data/202403账期识别后/202403账期.xlsx'

最后把'../data/202403账期识别后'这个目录打包发给会计