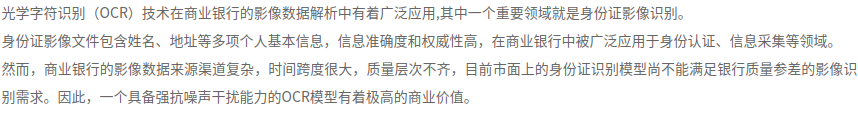
**毕业设计论文进度汇报（2.17-2.23）**

姓名：高尚 学号：20161104492

1. **目前论文的进度**

论文题目为：《基于OCR的身份证要素提取设计与实现》

[题目背景](https://www.datafountain.cn/competitions/346)：



论文正文部分包括：

1. 绪论：光学字符识别技术目前的发展状况，论文采用的技术方案。 完成
2. 图片预处理：原始身份证扫描件中提取有效区域及去除水印。 完成
3. 文本处理：定位图片中的文本和识别文本。
4. 结果纠正：爬取全国行政区和身份证号码的对应关系进行结果纠正
5. 方案展望：总结方案中的不足和后续可提升的地方。
6. **剩余工作量**

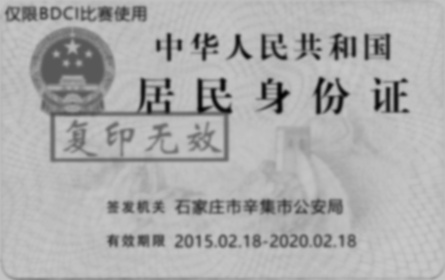
上述的（三）（四）（五）待完成，剩余约60%

1. **遇到的问题**
2. 上周遇到的图像方向颠倒问题已经解决；
3. 采用模型匹配方法去除水印已经出现效果，但仍有水印字样痕迹；
4. 处理(裁切　旋转　去水印)训练集图片时间较长，训练集10000张图片花费约20分钟

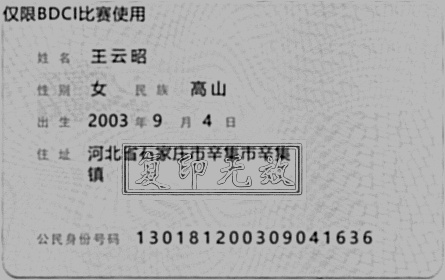
方向调整前：



方向调整后：



水印去除后：



1. **预计解决方案**
2. 去水印部分暂时不进行调整，后续测试对识别结果的影响作用，根据最终识别结果再做调整
3. 将旋转图片和去水印所用到的模板图像保存为.npy文件加快处理速度，[参考文章](https://blog.csdn.net/zsc201825/article/details/90264309)