



# 基于在线 OCR 技术的核酸报告 信息录入系统

上课时间：周四 8、9 节

姓名：俞辰杰

指导教师：王长波 教授

学号：10192100571

## 目录

摘要 .....	3
I. 引言 .....	3
II. 模型框架 .....	3
A.中文场景文字数据集	
B.CRNN	
C.模型训练	
III. 图像分割 .....	5
D.文字区域分割	
E.文字逻辑关系	
F.结果分析	
IV.交互设计 .....	7
G.Tkinter 前端	
H.socket 通讯	
I.Linux 云服务器	
V.总结 .....	8
参考文献.....	8

## 摘要

新冠疫情爆发常态化以来,需要检验的核酸检验报告大量增加,采用人工方法进行核酸报告的核查难免会导致错误率和效率低下的问题。因此,本论文就这一问题提供了一种在线的基于 OCR(光学字符识别)的核酸检测报告录入系统。实现由核酸检测的批量输入,到统计判断结果的 CSV 输出。在本论文中,对比了在传统 CRNN 网络产出的模型与通用中文场景 OCR 模型的效果差异;同时,为了减少在本地部署的难度和规模,便于推广使用,本模型采用了云计算方法,将系统部署在服务器上,本地只需要上传图片文件就可以得到结果。结果表明,此系统可以高效准确的识别核酸检测报告中的关键信息。

## I. 引言

自 2019 年新型冠状病毒爆发以来,居民的日常出行生活受到了巨大的影响,核酸检测逐渐常态化。2022 年上海新冠 BA.2.12.1 变种奥密克戎爆发至 2022 年 5 月 31 号为止,上海单日最大核酸检测 850 万管以上,已设置超过 1.8 万个常态化免费核酸采样点[1]。而根据上海市卫健委的要求,出入公共场所,乘坐公共交通上海市居民出入需携带 72 小时内有效核酸检验证明。在核酸检测常态化的当下,核酸检测报告已成为全国市民出入所必备的道具。

在如此巨大的检测量之下,如果仅靠少数志愿者和核酸检疫人员通过肉眼进行核查报告,将难免导致核酸检测的漏检误检。这种疏漏在人口高度密集的海,无疑将导致新冠病毒的大规模传播,其造成的经济损失,生命财产损失难以估量。此外,人为核实检测报告的效率相比计算机也更加低下。通过“图像输入到文字输出”的转换可以有效的提升效率。

光学字符识别技术(OCR)是近年来计算机视觉分析领域的研究方向之一,其在自然语言处理方面也起到的重要的作用。在实际应用中,由于中文场景中存在背景干扰、字体相近、残缺等多种复杂情况。因此国内学者对于中文场景文字识别(STR)的研究也是在光学字符识别技术的重要研究方向之一。本系统是基于辜双佳、栗智 [2] 提出了一种基于 CRNN 模型的中文场景文字识别方法,通过 CNN [3] 神经网络提取图像的卷积特征,然后使用 RNN [4] 进行序列特征预测,并以 BiGRU 和 Dropout [5] 抑制梯度爆炸、消失和过拟合的问题,获得了良好的检测精度。

## II. 模型框架

实验用电脑搭载 Intel i5-10210U 处理器,一个 NVIDIA 2060 显卡;深度学习框架采用 Pytorch-1.7.1,以及 Nvidia 公司的 CUDA 10.2 GPU 运算平台以及 CUDNN 深度学习 GPU 加速库。

### A. 中文场景文字数据集

如前文所说,中文场景的背景非常复杂,为了简化数据学习和实际预测的复杂程度,本论文是基于黑白图像进行文字识别,所要完成的检测仅是从核酸报告的白底黑字中提取重要信息,因此,用于深度学习的数据集也只要单通道下的中文印刷书籍的照片。对于待学习的数据集,建立由汉字与数字之间的对应关系表,将学习得到的映射关系和数据标注保存成两份 txt 文件,共计约 300 万余张。

## B. CRNN

模型卷积循环神经网络(CRNN,Convolutional Recurrent Neural Network)是一种将 CNN、RNN 和 CTC 三个模块结合起来的新型深度神经网络模型，相对于技术较为成熟的 OCR 技术。

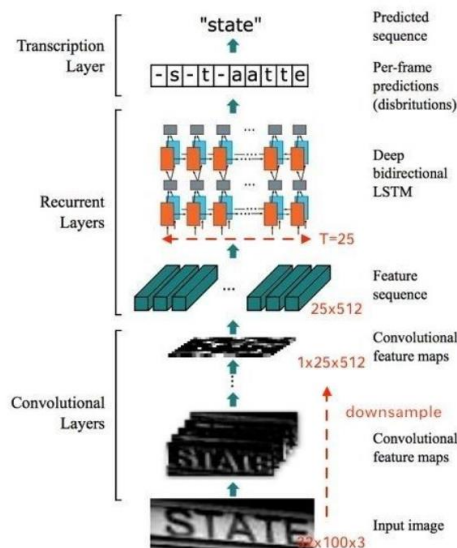


图 2-1 CRNN 模型示意图

其中，卷积层 CNN 针对输入图像，将原图转换成一系列的特征图，这些特征图保留了原始图片的视觉特征信息。循环层 RNN 提取其中的语义关联，本模型中使用双层 LSTM 在卷积特征的基础上继续提取文字序列特征。最后经过损失函数 CTC 进行 softmax 获取最大概率，对齐字符并输出标签序列。

## C. 模型训练

命令行打开项目文件夹，键入命令 `pythontrain.py--cglib/config/360CC_config.yaml` 开始训练。期间使用 `tensorboard` 可以查看训练的 Loss 曲线。

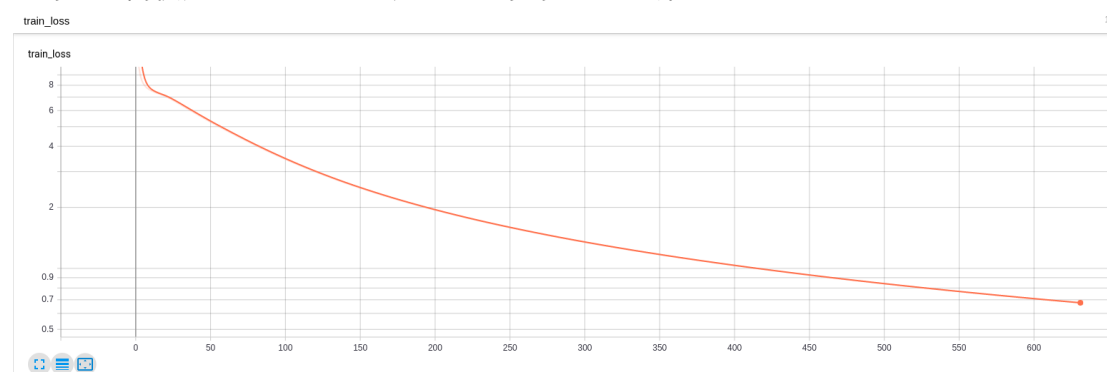


图 2-2 模型训练 Loss 曲线

### III. 图像分割

由于一份核酸检测报告中存在着的大量的文字信息, 所以需要将原图像中需要进行摘取的文字进行区域分割。在现有的文字图像提取的方法中, 以深度学习完成的图像分割类模型, 例如 Google 实验室的 Liang-Chieh Chen 等人通过 DeepLab v3+[6]在图像分割上取得了显著的精确程度。然而本论文的测试图像有着高度相似的文字结构和内容, 用数字图像处理的办法可以对图像进行统一预处理进行提取, 可以更加准确的提取出所需要的内容。本项目利用 OpenCV 的一些库函数对原图像进行调整, 从而分割出文字所在的区块。

#### D. 文字区域分割

对于若干张核酸报告截图, 首先将图像尺寸统一为固定大小, 然后将图片转为单通道图像减少数据规模。因为图像原来存在着有存在样式的装饰性文字, 灰度图的文字特征还需要使用 OTSU 方法将其转化为二值图后取反。在取反过程中为了消除图像尺度的变化所导致的变化, 用像素级操作消除横向的线和大块区域, 对图像进行若干次腐蚀-膨胀操作, 直到需要的文字清晰可见。

经过参数调整后得到的最终效果对比图如下。依照分割效果图的联通区块, 由于多个文字横向拼接, 区块一定是长大于宽的长方形, 因此去掉偏正方形的区域和宽大于长的区域。根据最终的分割结果, 将原图像分为若干小块, 并对小块按照从上到下, 从左到右的顺序排序。



图 3-1(a)(b)(c) 图像分割处理



图 3-2 文字区块提取结果

## E. 文字逻辑关系

处理将分割好的文字块输入训练好的模型中，得到按顺序排列的文字信息。对文字信息的前后文逻辑关联进行梳理之后，总结表项得到表格。

8	个人信息
9	姓名
10	金子龙
11	****1614
12	证件号码
13	姓名
14	金子龙
15	2022052809:57
16	采样时间
17	检测机构
18	上海解码医学检验所
19	(气膜实

(a) 文字原信息

1	姓名,检测时间,检测机构,结果
2	0810,-05-,上海世和医学检验实验室,阴
3	霞,:520,上海交通大学医学院附属瑞金医院,阴
4	俞辰杰,-053,上海交通大学医学附属瑞金医院,阴
5	金子龙,0528,上海解码医学检验所(气膜实验室),阴
6	

(b) CSV 文件

图 3-3(a)(b) 文字映射结果

## F. 结果分析

批量处理若干张图像后得到最终表格，可以发现图中的检测的准确率实际小于 80%，由此可以判断基于 CRNN 的中文场景文字识别效果在处理此类图像的能力比较一般。在识别文字区块时容易产生错字现象，并且对符号，数字的抗干扰能力较差。因此处理语言逻辑时需要构建类似字符的字母表，根据上下文关联进行筛选。并且，由于本系统对于图像还需要进行使用分割-检测-处理的多层级图像处理，分析一张图片需要耗时大约 1 分钟。可以预见到，这种方法在实际中运用中，不论是准确率和效率都远不如人工识别。



图 3-4 CRNN 训练识别结果

而第二种通用场景下的 OCR 模型在对于相同输入的情况，得到了准确率更高的判断结果，并且对符号的识别也比较精确。通用模型的耗时约 2-3 秒。因此在模型的选择上，还是选择通用中文场景 OCR 模型。模型的效果如下图。图 2-6 通用中文场景 OCR 模型

1: 18:11 天?的	0.829
2: 张型ID 08.182%	0.506
3: 检测时间	0.973
4: 上海解码医学	0.982
5: 接种记录	0.995
6: 个人信息	0.837
7: 2022-05-31 18:11:28	0.978
8: 姓名	0.999
9: 张*鑫	0.657
10: 证件号码	0.997
11: 卡***0810	0.884
12: ②老幼接种券	0.906
13: ②境外人士接种	0.906
14: 姓名:	0.991
15: 张*鑫	0.505
16: 采样时间:	0.979
17: 2022-05-30 10:02:31	0.952
18: 检测机构:	0.980
19: 上海世和医学检验实验室	0.953
20: 检测时间:	0.994
21: 2022-05-31 00:10:00	0.963
22: 检测项目:	0.995
23: 核酸	0.999
24: 检测结果:	0.976
25: 【阴性】	0.931
26: 姓名:	0.676
27: 采样时间:	0.979
28: 2022-05-27 10:10:09	0.952
29: 检测机构:	0.986
30: 华康康华医学检验实验室	0.948
31: 检测时间:	0.992
32: 2022-05-27 20:40:25	0.965
33: 检测项目:	0.994
34: 核酸	1.000
35: 检测结果:	0.986
36: 【阴性】	0.944
37: 姓名:	0.893
38: 张*鑫	0.604

图 3-5 通用中文 OCR 预测结果

## IV.交互设计

对于程序而言，为了更好的实现人机交互，我们搭建了应用程序窗口界面。通过可视化的可视化的交互式前端和服务器通信，可以高效，准确，便捷的实现交互设计。而前端和服务器之间采用云计算的方法。云计算（cloud computing）是分布式计算的一种，指的是通过网络“云”将巨大的数据计算处理程序分解成无数个小程序，然后，通过多部服务器组成的系统进行处理和分析这些小程序得到结果并返回给用户。将项目部署在云端可以避免使用者在不同设备上重复安装环境、部署项目，可以较为简易的被大众所使用，减少部署项目，更新项目所导致的时间开销和成本开销，同时也可以利用云服务器的高效计算特征来提升效率。

### G. Tkinter 前端。

Tkinter 模块(“Tk 接口”)是 Python 的标准 TkGUI 工具包的接口。Tkinter 模块适合搭建简洁的前端。对于我们的软件而言，只显示了我们的项目名称，以及选择图像文件，选择保存目录，以及将数据上传至服务器进行文本识别这三个功能。

整体前端窗口简介，采用深蓝色和淡蓝色的结合，使得整体颜色观感和谐，便于用户长时间的观看而不至于眼睛疲劳。同时结构主题清晰，响应时间短，能够较好的实现人机交互的目的。

两个地址选择栏，通过直观的图像化的界面进行可视化，实现图像的批量上传与结果下载。在不选择保存文件夹的时候，结果保存路径和图像文件夹同一路径；如果路径不完全的情况下，系统会警告用户需要确认路径是否合法。一个运行按钮和服务器建立连接，一个退出按钮防止遇到未预料到的错误强制退出，共两个按键，清晰简洁的提示用户前端窗口的使用方法。



图 4-1(a)(b) Tkinter 前端窗口

### H. socket 通讯

Socket 是应用层与 TCP/IP 协议族通信的中间软件抽象层，它是一组接口，在本地测试环境中使用 Socket，可以减少网络通信的复杂程度。Socket 通讯的流程由服务器端开始，服务器端先初始化 Socket，然后与端口绑定(bind)，对端口进行监听(listen)，调用 accept 阻塞，等待客户端连接。客户端发送数据请求，服务器端接收请求并处理请求，然后把回应数据发送给客户端，客户端读取数据，最后关闭连接，一次交互结束。对于我们的程序而言，需要服务端与客户端互相发送指令来协同完成工作。同时使用 TCP 连接，可以有效保证通讯内容的安全稳定，个人信息不会轻易的被窃取。



## I. Linux 云服务器

论文所述系统采用的是 python 中的 socket 包。因此对于云服务器的环境配置要求不高。在安装好通用中文 OCR 模型后，只需根据需要调整 python 及其模块的版本即可。

## V. 总结

本文详细描述了一款基于在线 OCR 技术的核酸报告信息录入系统的构建过程及所用到的技术。完成了由核酸报告截图的图像输入，到信息汇总表的文件输出。并且采用云计算和打包前端的方式，避免了环境和项目的重复部署，提升了应用程序的可用性。本系统在实际检测核酸报告的时候，不仅可以有效提升检测的效率，减少误筛漏筛的可能性；而且在保留检测记录的服务器上，可以有效实现对于阳性病例的流调追踪，也可以提供途径地区的密接人群的相近时刻的查询记录，对于闭环管理提供了一种新的可能性。

本项目包含三种版本，分别是前期是测试使用的，包含被放弃的 CRNN 的文本识别的 DeepLearning 版本；服务器提供者需要进行部署的 Server 版本；以及用户使用的，只包含可执行文件的 Client 版本。

## 参考文献

- [1] 百度新闻. 常态化核酸检测堵在哪? 专家提出多环节改进方案[EB/OL].(2022-6-2)[2022-6-2]. <https://baijiahao.baidu.com/s?id=1734508672153079984&wfr=spider&for=pc>.
- [2] 辜双佳, 栗智. 基于 CRNN 模型的中文场景文字识别 [J]. 科技 风,2021(17):108-110. DOI:10.19392/j.cnki.1671-7341.202117047.
- [3] LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition. Proceeding of the IEEE, 1998, 86(11):2278-2324.
- [4] MIKE SCHUSTER, KULDIP K. PALIWAL. Bidirectional Recurrent Neural Networks[J]. IEEE Transactions on Signal Processing: A publication of the IEEE Signal Processing Society,1997,45(11):2673- 2681.
- [5] Hinton G E, Srivastava N, Krizhevsky A, et al.Improving neural networks by preventing co-adaptation of feature detectors [J].Computer enee,2012,3(4):212—223.
- [6] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, Hartwig Adam: Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. CoRR abs/1802.02611 (2018)