

本课题针对 OCR 技术的发展,对经典文字检测模型 **advancedEAST** 和文字 识别模型 **CRNN** 进行研究, 并且构建有针对的数据集进行重新训练,将检测与 识别二者结合,开发一款基于 OCR 技术的图像信息转可编辑文字信息的文章 分析平台.主要解决如今用户在工作和学习中遇到的一些无法编辑的文章如图片、 pdf 时,却想要去二次编辑文章的问题。针对市面上的类似的网站只有单调 的实现图像转文字信息,本课题结合了自然语言处理 (**NLP**) 技术对识别后的文 字信息作为处理。将多项技术集成于一个网站上。同时,也加入了文章管理 的系统,给用户提供更多操作的空间和便利,方便了用户学习和工作需要, 节省了大量的时间。