

国家知识产权局信息显示，国科环宇（南京）信息技术有限公司申请一项名为“一种基于掩码处理的 PDF 文件版面识别的方法”的专利，公开号 CN121354152A，申请日期为 2025 年 12 月。

专利摘要显示，本发明公开了一种基于掩码处理的 PDF 文件版面识别的方法，旨在实现 PDF 文件中标题、正文、页眉、页脚、表格和图片等多种版面元素的精准识别。该方法通过 MuPDF 解析 PDF 文件中的文字及其坐标，采用不同颜色掩码分别替换文本和标点符号，保留非文本区域原始内容，生成掩码处理后的页面图片，并基于此构建训练数据集，利用目标检测模型(如 YOLO)进行训练，得到高精度的版面识别模型。本发明有效提升了 PDF 版面识别的准确性和自动化水平，具有广泛的应用前景。

天眼查资料显示，国科环宇（南京）信息技术有限公司，成立于 2022 年，位于南京市，是一家以从事科技推广和应用服务业为主的企业。企业注册资本 5000 万人民币。通过天眼查大数据分析，国科环宇（南京）信息技术有限公司共对外投资了 1 家企业，参与招投标项目 1 次，专利信息 28 条，此外企业还拥有行政许可 3 个。