

档案软件 - 批量 OCR 处理功能【云丰档案软件功能介绍】

档案管理软件 · 功能介绍

存量档案成千上万页，一页页点识别不现实，必须能批量自动跑。批量 OCR 处理把成批扫描件一次性提交，系统自动排队逐页识别，无需人工逐份操作，大批量数字化加工效率倍增。

一、为什么需要批量 OCR 处理

数字化加工往往是几十万页的工程。若每识别一页都要人工点一下，加工队再多也忙不过来。

逐份操作慢：一页页手动触发识别，人工盯在电脑前效率低；

排队不可控：大批量任务何时跑完没人知道，只能干等；

失败难定位：几百页里某几页识别失败，难以单独挑出重跑；

资源浪费：不能利用夜间闲时，白天电脑被占着干别的。

批量 OCR 把识别变成后台流水线：一次提交、自动排队、进度可见，大批量加工不再靠人肉逐页点。

二、怎么用：整批提交，后台自动跑

选定一个案卷或一批扫描件，提交批量识别任务，系统后台排队处理，实时显示进度。

表1 批量 OCR 处理流程

步骤	操作要点
选择批次	选定一个案卷或一批扫描件；
提交任务	发起批量 OCR 任务；
后台排队	系统自动逐页识别，无需人工盯守；
查看进度	实时查看完成页数与成功率；
失败重跑	识别失败页单独挑出重处理。

批量不等于放任。批量跑完后仍需按质量抽检：识别率低的批次要抽查校对。系统记录每页识别状态，失败页与低置信度页可单独拎出返工。

失败页与低置信度页自动隔离，不影响整批入库，可单独重跑。任务可排队闲时执行，利用夜间算力。

任务队列支持优先级，急件优先处理。处理进度实时可见，大批量夜间自动跑完，次日直接验收。

批量 OCR 让大规模数字化工程成为现实。系统把待识别文件排成队列，支持优先级设置，急件优先、闲时与夜间自动跑完，白天人力可投入校对与著录。失败页与低置信度页自动隔离，不影响整批入库，可单独重跑。处理进度实时可见，几十万页的存量工程后台流水线自动完成。这种批处理能力把逐页手工识别升级为工业化流水线，既大幅缩短加工周期，又通过错误隔离保证质量，是存量档案数字化的效率底座。

三、核心能力

表2 核心能力一览

能力	一句话价值	关键细节
整批提交	一次不用管	成百上千页一次提交，后台自动逐页识别。
进度可视	跑了多少知道	实时显示完成页数、成功率与预计时间。
失败隔离	坏页不拖累	识别失败页单独标记，不影响整批入库。
闲时利用	夜里也在跑	任务可排队在闲时自动执行，利用算力。
状态留痕	加工可追溯	每页识别时间与结果留记录，便于返工与审计。

几十万页的存量工程，靠后台流水线夜间跑完，白天人力可投入校对与著录。

四、典型使用场景

存量数字化：数十万页老档案批量提交 OCR，夜间自动跑完。

新卷加工：一新卷扫描完成后整卷提交识别。

失败返工：对低识别率批次单独重跑问题页。

五、价值与产品定位

批量 OCR 的价值在于把数字化加工从“逐页手工”升级为“批量流水线”：一次提交、自动跑完，大规模存量数字化才现实可行。把加工从逐页手工升级为批量流水线，大规模数字化才现实。

它是云丰档案数字化加工能力的效率引擎，与任务分配、进度跟踪、质量抽检配合，支撑大批量加工工程。