

如何提高扫描效果？计算机等级考试 PDF 转换可能丢失图片或格式，建议阅读原文

[https://www.100test.com/kao\\_ti2020/644/2021\\_2022\\_\\_E5\\_A6\\_82\\_E4\\_BD\\_95\\_E6\\_8F\\_90\\_E9\\_c98\\_644272.htm](https://www.100test.com/kao_ti2020/644/2021_2022__E5_A6_82_E4_BD_95_E6_8F_90_E9_c98_644272.htm) 一、黑白类型 通常一般的识别软件(比如清华TH-OCR)在对文字识别后会自动生成一个与其图像文件同名的纯文本文件，但其缺点是每一行的末尾全是硬回车(直接复制到记事本中也是如此)，非常不便于编辑，所以我们一般采用导出功能而不采用直接保存为TXT文本文件的方法。除此之外，OCR识别软件还可以将识别结果以RTF格式导出，这种格式不仅能够直接为WORD或WPS识别，而且导出后的版式与原稿一模一样。对于含有图、表的稿件，可以采用RTF格式导出。特别要说明的是识别后的表格一定要使用导出方式，否则表格的格线将成为字符，而不再是表格线。

一般印刷品的扫描：对于黑字原稿的印刷品，采用黑白模式、300dpi进行扫描，这样做不仅扫描时速度快，而且文件小，识别速度快。完毕之后，我们可以点击校对键，对识别结果进行校对。对于个别错误的地方采用直接输入法改正即可。依笔者之见，区域划分也是提高识别效果的关键之一，区域类型有横排正文、竖排正文、表格、图形图像四种，在识别之前，可拖动鼠标划分区域，一般而言，可将字体字号相同的划在一个区域，选定划分的区域后，单击右键即可设置区域的类型。

报纸的扫描：由于报纸不易摆放端正，所以对扫描结果进行识别前。首先要点击“自动纠斜”。若纠斜效果不理想，还可以采用手动纠斜，不同的识别软件可能方法不同，清华TH-OCR的方法是按住Shift键的同时再按住鼠标右键，画出一条与倾斜着的文字平行的

线，然后松开鼠标右键，这样就可以达到理想的纠斜效果。

第二，要点击“自动版面分析”(报纸类)，这里要说明的是报纸的标题字号极大，若按字去识别，效果可能不理想，建议将此区域改成图形或干脆将该区域取消，识别之后再手工输入即可。同时还要取消不是该篇文章内容的其他区域。识别与导出的过程与上述方法相同。但需说明的是，由于报纸的版面特殊，一般不必保留，所以不宜导出成RTF格式，最好导出成Formatted Text Files的格式，这种TXT文件没有多余的硬回车。

二、灰度类型 当扫描有饼图、直方图的资料(比如建筑图纸、施工进度报告等，以及质量较差的报纸)时，那些饼图、直方图有灰度层次之分，我们称这类的稿件为灰度稿，这时，如果用黑白方式扫描的话，扫描出的图像(二像素图像)可能效果极差；对于那种过期老报纸，扫描出的图像可能对比度非常低，不利于识别。对于灰度稿，我们可以用灰度模式扫描，它可以完整保留饼图、直方图原有的灰度层次，对报纸也无需再像以前的OCR那样需要调整阈值了。为了提高识别效果，扫描的分辨率不要低于200dpi。

三、彩色类型 以前当我们遇到宣传彩页与彩色画报等彩色稿，既想识别里面的文字，又要在WORD文件中重新恢复版面并编辑时，往往需要先在OCR中用黑白方式扫描、识别文字，然后再在Photoshop中扫描印刷品中的彩色图像，最后在WORD中打开文字，插入图像，几乎等于要在WORD中重新排版，工作即复杂又不方便，而目前的OCR软件一般都具有彩色稿件识别功能(例如清华TH-OCR)，使用这样的OCR软件可以使彩色稿件的识别、恢复、编辑变得非常简单容易，一页A4幅面的彩页，从扫描到识别再到导出到WORD中，只要电脑速度

快，仅需短短的两分钟左右，要知道这时所处理的是一个20多兆的文件，如果用以前的方式恢复成WORD文件起码要用十多分钟的时间。为了保证印刷品图像的扫描效果，我们在扫描时要选择去网纹功能，这对文字的识别率不会有丝毫的影响。从效率角度考虑，我们要根据不同原稿类型选择不同的扫描模式，因为彩色文件的数据量是同等条件下灰度文件数据量的3倍，是黑白文件数据量的24倍，如果可以用黑白方式扫描的稿件用彩色方式去扫描，不仅数据量变大了，识别时间也会加长，而且扫描时间也要相应加长，毫无必要。从效果上看，黑白扫描方式识别的效果最好，对于只有文字和表格的清晰原稿(甚至包括一般质量的报纸)，如果方法正确的话，识别率可达95%以上；而对于以灰度方式和彩色方式扫描的文件，特别是当背景太花哨时，文字和表格的识别率就相对要低得多，如果原稿中的某些部分(即使是文字和表格)不需要编辑，最好在识别前划分区域时，将区域类型设置为“图形图表”。

编辑特别推荐: 计算机等级考试三级PC技术模拟试题及参考答案一 计算机等级考试三级PC技术模拟试题及参考答案二 计算机等级考试三级PC技术模拟试题及参考答案三 100Test 下载频道开通，各类考试题目直接下载。详细请访问 [www.100test.com](http://www.100test.com)