表格识别

1. 以灰度格式读取图片，gray =0.3\*R+0.59\*G+0.11\*B，白色为255，再将所有255的值转换为1，非255的值转换成0.
2. 识别表格交点，即如图一所示，识别方式，遍历每个像素点，找出符合交点格式的所有点，这个点的特点是，它的上下左右最少有两根黑线如L型、T型。将识别后的点存在一个三维数组中，三维数组形式：由n行组成，每一行内存放的是在表格内同一行的交点，每个交点有[x,y]组成。一般在这一步耗时最大。



图1 表格交点

1. 识别出交点后，有三个问题：1.有些文字太过方方正正也会导致出现交点的形式；2.线条太厚会出现连续的交点。3.渐灭层是独立的一行，与它原本所在行独立出来，如图2所示，有些交点可能会丢失（描述表部分）;

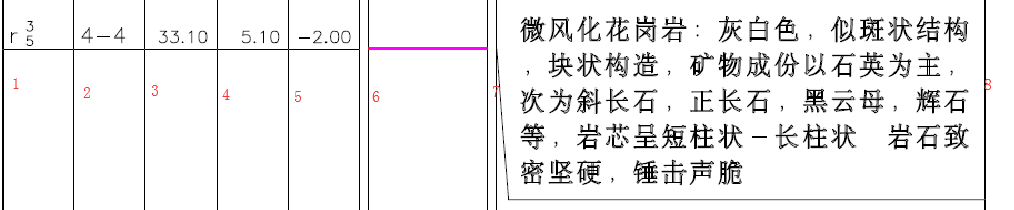


图2 渐灭层

1. 第一个问题解决方案：找出表格左上角的坐标，即横向min\_x，文字的x肯定大于min\_x的，并且独占一行，而表格框架的起始x必定是min\_x，所以去除row[i][0][0]<min\_x的行（row[i][0][0]为每一行的第一个点的x数值），由于渐灭层也是单独一行，切x大于min\_x，所以需要单独找出歼灭层的x坐标将其忽略掉。
2. 第二个问题解决方案：这个问题好解决，加个判断去除即可。
3. 第三个问题解决方案：将描述表部分的列从1-8标号，判断每行前7个点是否具有相同的x坐标，如果没有则补齐。由于第8列上的交点个数与前7列每列上的交点个数和相等，所以将第8列上的点提取出来，依次添加到描述表的每一行的末尾。
4. 生成单元格，交点已经全部找出，遍历所有交点，将改点当做单元格左上角，找出位于改点右下方的点，进行组合，生成[[a,b],[c,d]]矩阵即单元格，使用数组按行存储
5. 切割图片，文字识别，根据生成的单元格，依次将图片进行切割，再调用BaiduAI的借口进行文字识别。
6. 将获取到的结果存储下来，待全部识别后，将内容依次写入excel表格中