**病理切片大作业阶段三报告**

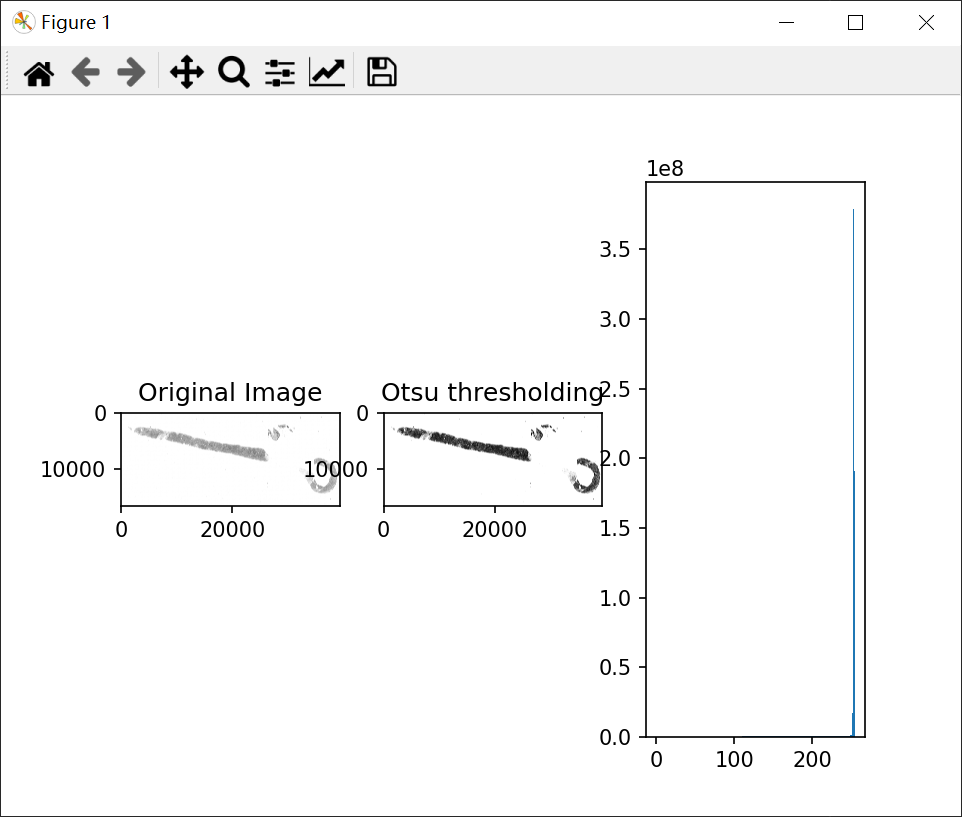
1. **相关思路**

**（1）分块**

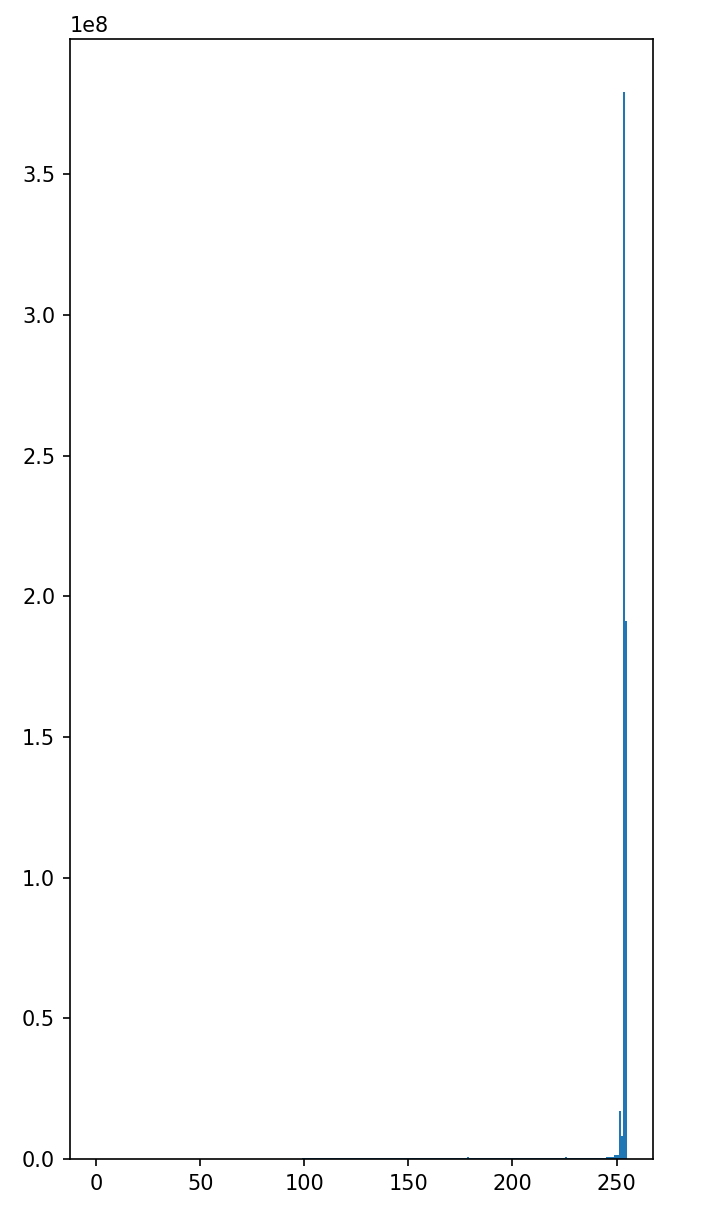
分块的目的是提升图像处理的并行度，从而在实际应用时提高性能。分块时，为了正确处理边缘信息，块与块之间还需要保留一定的overlap区域。考虑到样例图的尺寸在10k~40k \* 10k~40k像素的尺度左右，单个汇管区的尺寸基本在1k \* 1k像素以内，因此选择了10k作为分块大小，1024作为overlap区域长度。经过测试，该参数在测试样例上的表现较为理想。

**（2）提取前景**

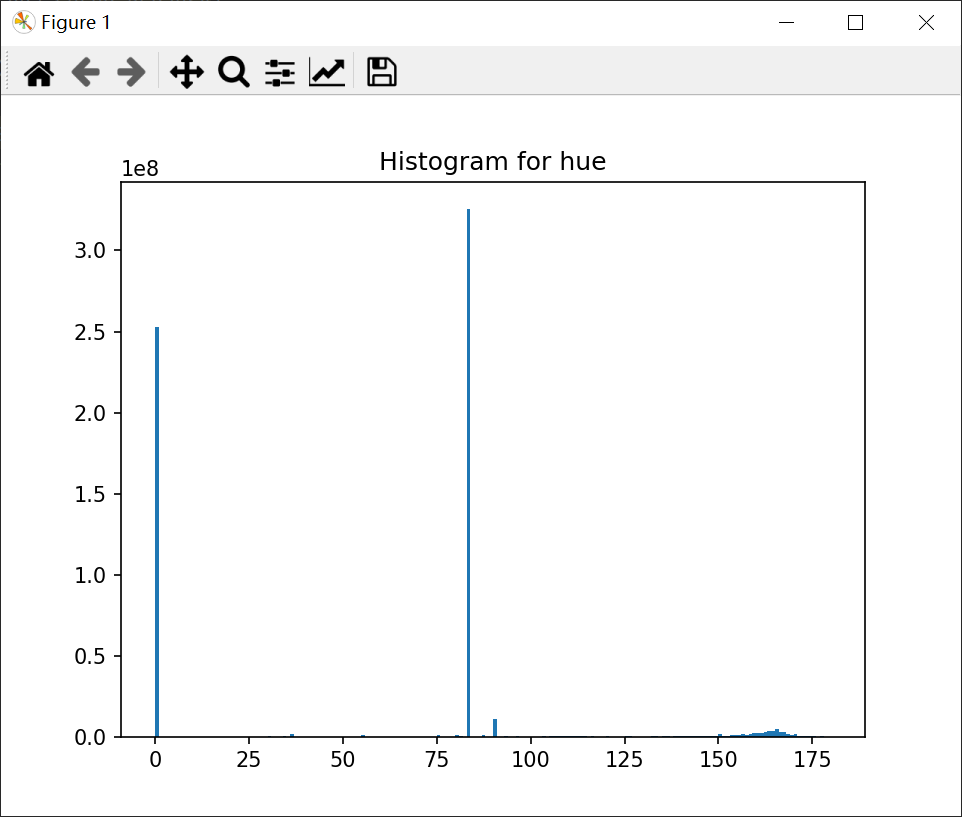
从样例图中可以看出，纤维化区域被事先染成蓝色，肝脏其他区域呈现红色，而背景为白色。为了提取汇管区，尝试对原图进行灰度分析，并使用Otsu算法对其进行阈值分割。如下图所示：



从灰度直方图中可以看出，染色区域与白色背景的灰度存在明显差异，故使用Otsu算法可以很好地分离背景。同时，我们也可以观察到，原图的灰度图整体呈现单峰状，很难仅根据灰度就把红色和蓝色区域分离开来，所以我们还要对图像的色彩进行分析。



为了根据颜色特征对图像进行分割，从而获得蓝色前景，我们需要把图像转化到HSV空间，并且对其进行色彩分析。样例图的色相值分布如下：



从上图中可以看出，第一个波峰(0附近，红色)对应非纤维化区域，而第二、第三个波峰（75-100之间，青蓝色），第四个波峰（150-175，紫色）则对应纤维化区域。整体来看，纤维化和非纤维化区域的色相偏离较为明显，故可以尝试使用红色和蓝色两个遮罩掩码进行图像分割。此外，为了消除掩码带来的噪声，使用了形态学中的开操作进行降噪。降噪后的掩码如下：

可以看到，遮罩的噪声被基本去除，同时，图像的特征也被保留。

基于上述思路，我们取蓝色遮罩与非红色遮罩的与，再与原图进行求交即可得到前景。

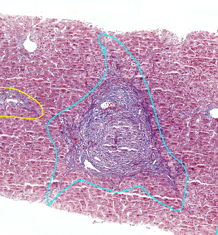
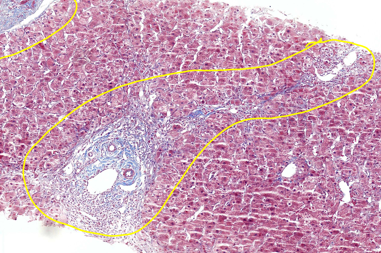


**（3）图像分类**

作业中，要求对正常汇管区，汇管区纤维化和纤维间隔等进行区分。考虑到纤维化区域具有特定的形状特点，并且细胞核较为密集，因此最终采用形状分析和纹理分析的方法进行分类。

**（a）形状分析**

对于正常的汇管区而言，由于其纤维化程度较低，因此蓝色区域的外轮廓往往表现为椭圆形，且区域面积较小，空腔占比较大；而对一级纤维化区域而言，由于纤维化的影响，蓝色区域的面积相对更大，然而外轮廓依然类似椭圆，无明显的凸包；对二级纤维化而言，由于纤维化已经向四周发展，因此轮廓往往表现出更大的复杂性，具备若干个凸包或拐点；对三级纤维化（纤维间隔）而言，由于纤维化区域将两个独立的汇管区连接在了一起，因此外轮廓在连接的中心往往表现出内凹的形态，纤维化的面积也更大，并且整个区域呈现出“长条形”。



基于上述的分析，我们可以从区域面积和外轮廓几何特性出发，对区域进行分类。经过实际测试和筛选，最终选用了area（区域面积），solidity degree（区域面积/凸包面积）、hu矩的某些项和aspect ratio（最小包围盒长宽比）作为分类的形状依据。具体的阈值由实际的人工测试得出。

**（b）纹理分析**

尽管上述的形状分析在实践中已表现出不错的效果，但对于正常汇管区和一级纤维化区域，只靠形状分析并不能予以很好的区分，存在误判的情况。为此，进一步地，我们尝试对汇管区纤维化进行纹理分析。

首先，轻度纤维化的汇管区往往出现不同程度的纤维化组织塌陷，并且空腔附近细胞核较为密集，而细胞核在切片中又表现出深红色（也就是较重的灰度）。基于上述观察，我们认为纤维化区域的频率较高，灰度的变化更剧烈，纹理模式更加复杂。为此，笔者尝试提取不同区域的GLCM矩阵，并比较各个参数（contrast，homogeneity，energy，entropy）。

经过测试，笔者发现homogeneity，即同质化系数对正常区域和非正常区域的划分较为明显。对于正常的汇管区区域，同质化系数的均值一般大于0.9980，而纤维化区域的同质化系数则基本在0.93~0.98区间内。因此，最后选择同质化系数作为分类的纹理依据。

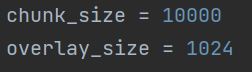
1. **方法设计**

选择python作为编程语言，使用opengCV库进行图像处理，使用tkinter库制作简单的GUI。系统设计如下：

**（1）分块：**点击**Split**按钮，选择一张图片进行分块，并且将分块结果输出到指定文件夹。



分块参数在main.py的中指定：



**（2）提取前景：**点击**Upload**按钮上传分割后的图像。根据预先调试好的参数分别生成红色和蓝色掩码，以及蓝色与非红色的位与掩码。分别对掩码进行开操作和闭操作进行降噪。最后根据掩码得到分割后图像。调节参数直至分割效果良好。点击**Save**按钮保存分割结果。





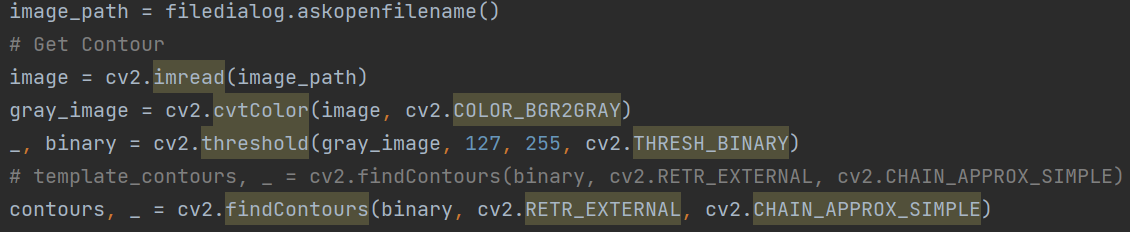




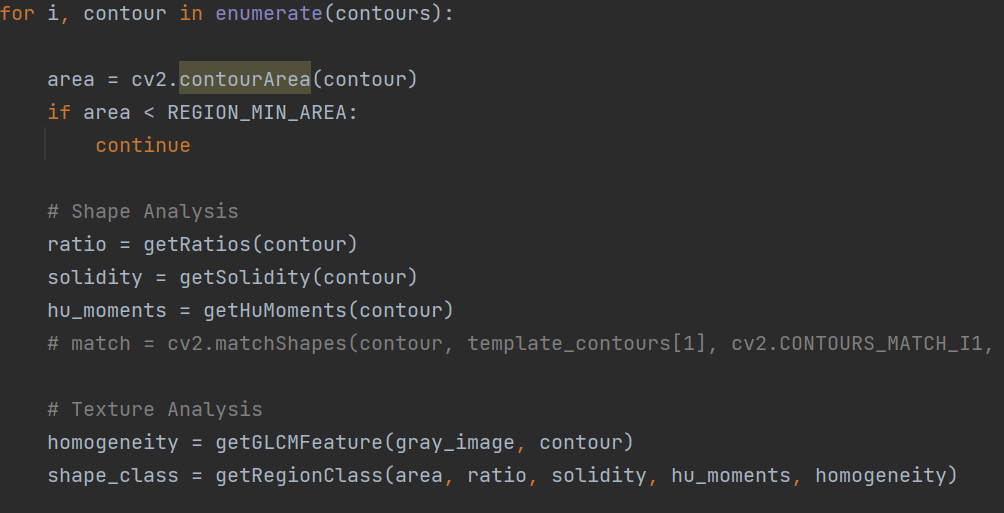
**（3）图像分类：**点击**Classify**按钮上传分割后图片。



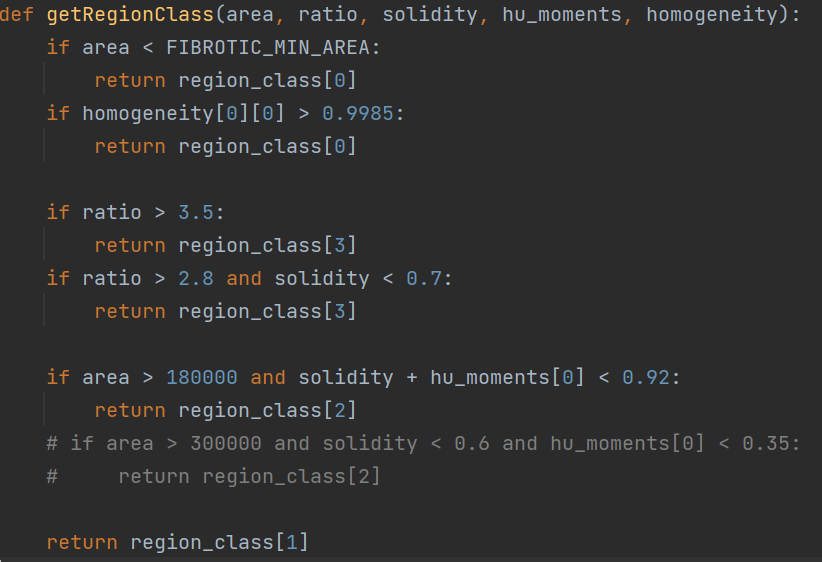
调用process\_image方法。process\_image方法中，将图像转为灰度图，并进行二值化，最后得到各个连通区域的轮廓线。



然后，遍历各个连通区域，对每个区域分别进行形状分析和纹理分析。在分析前，先过滤面积较小的区域，排除脂肪粒和其他因素的影响。



计算出各个指标后，调用getRegionClass获得分类结果。getRegionClass函数中，根据预先调试好的参数进行分类。

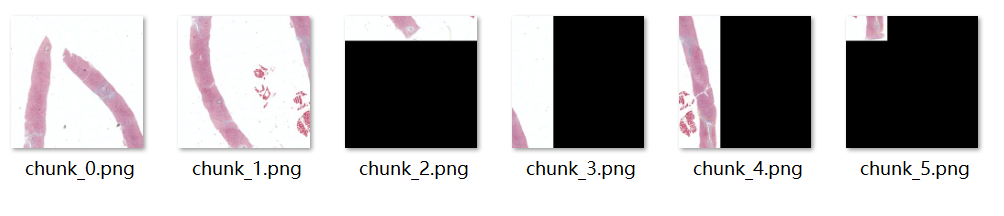


如上图所示，首先根据面积和同质化系数完成对正常汇管区的判断，然后，根据长宽比和solidity度判断三级纤维化（纤维间隔）的可能性，最后，通过面积，solidity和hu钜的首项判定二级纤维化。

1. **成果展示**

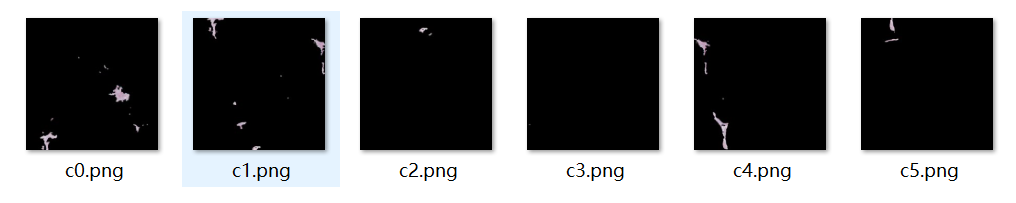
**（1）分块**

分块结果如下：



**（2）提取前景**

结果如下：

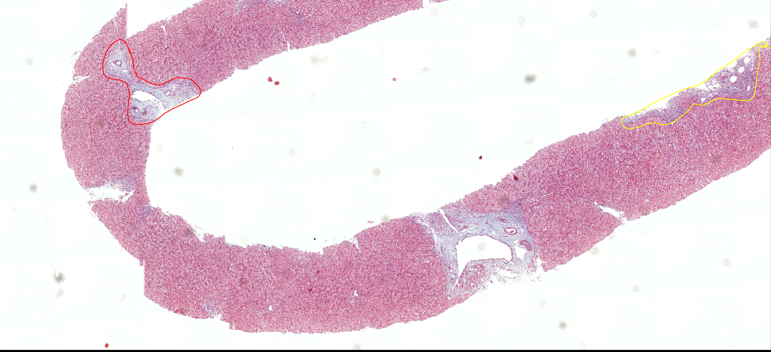


**（3）图像分类**

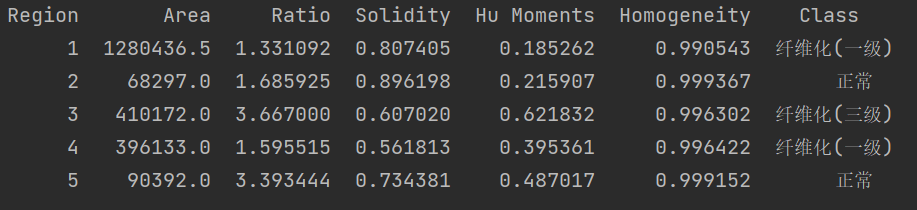
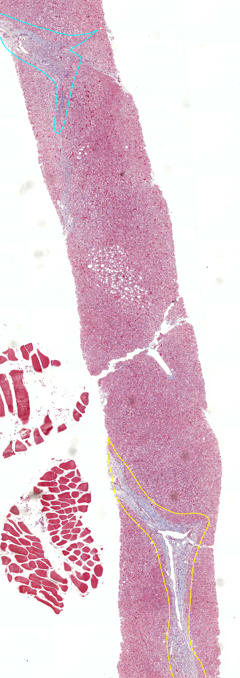
选取4张完整的样例图（01a250m，06a202m，06a283m，06a284m），分块后得到28张图片，分别对这28张图片进行classify。结果统计如下：（正确率与错误率仅针对事先被圈出来的区域，没被圈出但被检测出来的区域不列入下表）

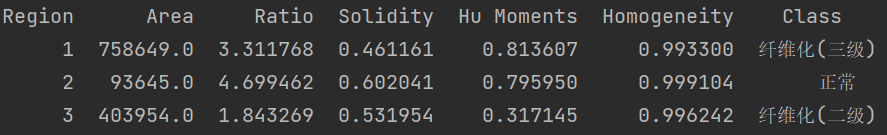
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 实际 | 检测正确 | 检测错误 | 错误率 | 准确率 |
| 正常汇管区 | 7 | 6 | 1 | 0.13 | 0.86 |
| 一级纤维化 | 5 | 4 | 1 | 0.20 | 0.80 |
| 二级纤维化 | 8 | 6 | 2 | 0.25 | 0.75 |
| 三级纤维化 | 16 | 15 | 1 | 0.06 | 0.94 |
| 总和 | 36 | 31 | 5 | 0.14 | 0.86 |

部分结果详情如下：



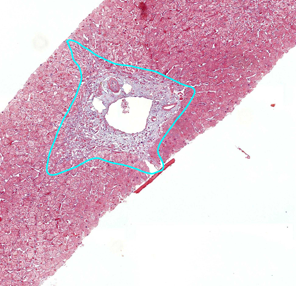


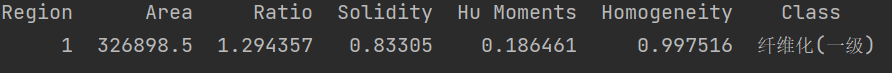
  



1. **结果分析**

从结果来看，分类的准确率达到85%以上，其中对于二级纤维化的检测出现了两例误判，其中一例的结果如下：



从图中可以看出，问题主要出在轮廓的提取上，在进行前景的分割时，上下两个角有断开现象，因此导致solidity值偏高，最终没有被归类到二级纤维化中。

此外，经过验证，其他的误判案例均与轮廓线的划分不准有关。可见，该方案对前景分割和轮廓线的准确性非常敏感，在实际运用中，还需要手动调节降噪半径和结构元半径，以期获得正确的分类结果。

综上所述，本方案主要基于形状分析和纹理分析。形状分析上，考量了包括hu矩，solidity，长宽比等几何特征；纹理分析上，考量了GLCM矩阵的各种参数。方案对于样例图表现出85%左右的准确率。此外，该方案的准确性较为依赖分割结果，在实际运用中，可能需要先调节参数获得良好的分割，才能最终获得较为准确的结果。