

数字图像处理

实验报告

学号: 181604

姓名: 陈李易

目录

- 1. 设计思路3
- 2. 实现步骤/细节.....3
 - 2.1 图像预处理3
 - 2.2 白细胞核与白细胞边界的提取4
 - 2.3 轮廓算法的选择5
 - 2.4 轮廓算法的改进6
 - 2.5 白细胞的分割6
 - 2.6 白细胞的计数8
 - 2.7 Reference.....9
- 3. 处理流程图10
- 4. 参数设置11
- 5. 实验结果11
- 6. 结果分析12
 - 6.1 改进12

1. 设计思路

本次实验的目的有两个:1. 白细胞的计数，2. 计算白细胞的核占比。

对应的子任务即:1. 白细胞核的提取，2. 白细胞边界的提取，3. 白细胞的计数(边界的校准)，4. 算法需要具有鲁棒性(抗噪声能力)

由于白细胞核，白细胞边界都有较明显的像素值突变，所以考虑的基本思路是阈值+轮廓得到第一步的效果，之后再第二步的去噪声等干扰，通过先验知识对第一步的结果进行修正。

2. 实现步骤/细节

2.1 图像预处理

为了更好的完成前两个任务，提取白细胞核的边界与白细胞的边界。通过PS工具制定了标准答案，统计了四张图像的(白细胞核，白细胞质，全图)在不同通道(R, G, B, GRAY)的平均像素比例分布。如下图2-1：

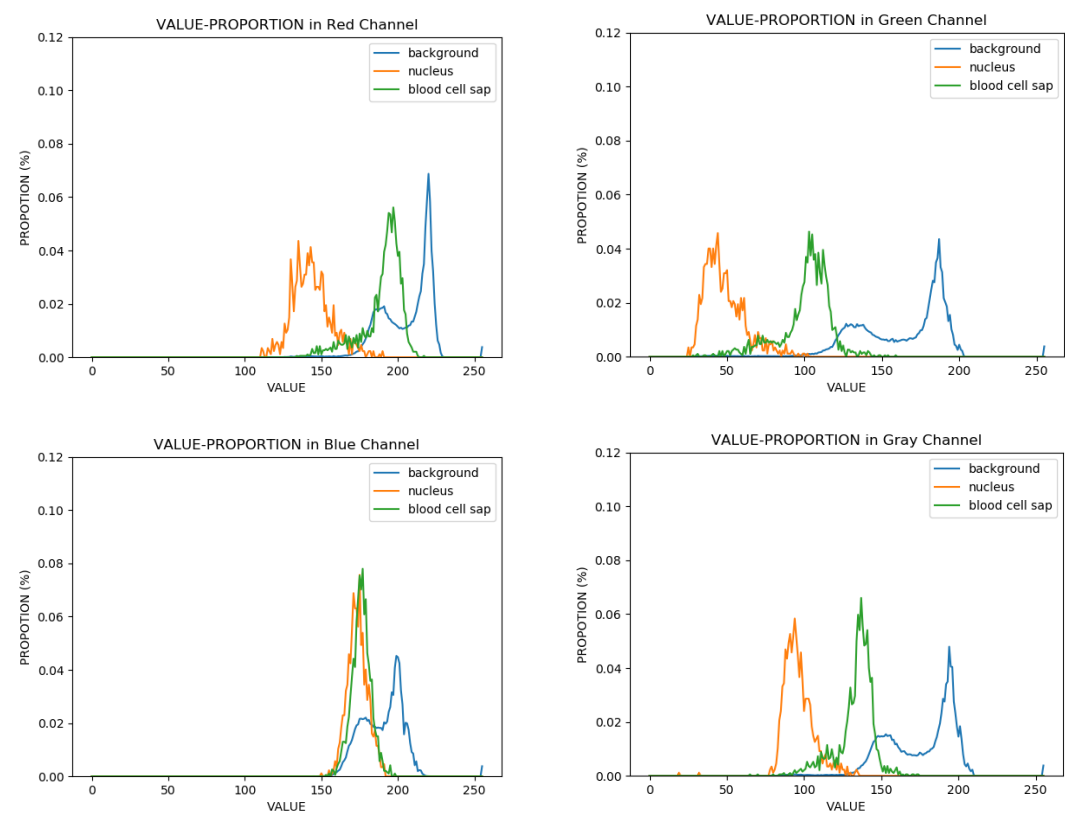


图2-1 像素值分布

可以明显的发现在G通道上，(核，质，背景)的像素分布相差最大。所以在之后的操作中对图像提取G通道作为预处理。(四张图像通道的分布相差不大，即对于这一数据源，抽取任一通道都能较好的反映原始图像的分布特点)。为了不破坏原始图像信息，不进行滤波等处理。

2.2 白细胞核与白细胞边界的提取

白细胞核的像素值相对于细胞质/背景的像素值都有明显的不同，所以通过阈值就能提取到白细胞核的大致区域。而白细胞的细胞质与背景的分度度较低，所以阈值过滤这种全局性的方法不适用，所以考虑使用活动轮廓等更注重局部梯度的算法。

所以基本思路是:1. 通过阈值提取到白细胞核的大致区域(也包含噪点)，2. 基于细胞核区域作为初始边界，使用如活动轮廓向外扩展，以期望找到白细胞的边界。这样就得到了非常粗糙的白细胞核和白细胞边界，如下图2-2:

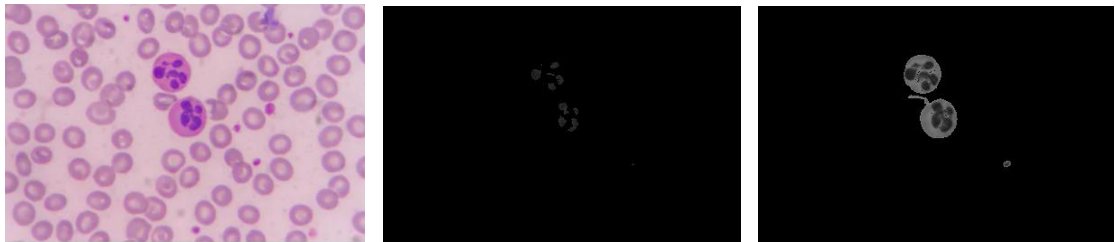


图2-2 初始化白细胞核与白细胞边界

那么如何进一步优化这个结果呢？通过观察最右图的粗略细胞边界，可以发现在这一粗略边界图中，白细胞核和白细胞质的区分度很高，即在剔除了大量背景后，白细胞核和白细胞质的区分度进一步显著了。所以容易想到可以基于粗略边界图再次通过snake等主动轮廓等算法得到更精细的细胞核边界图。

在进一步思考，有了更精细的细胞核边界，我们就能基于此，基于更显著的白细胞质与背景的分度度得到更精确的白细胞边界。依次反复迭代，得到更精确精细的白细胞核和白细胞边界。以下是迭代过程图2-3:

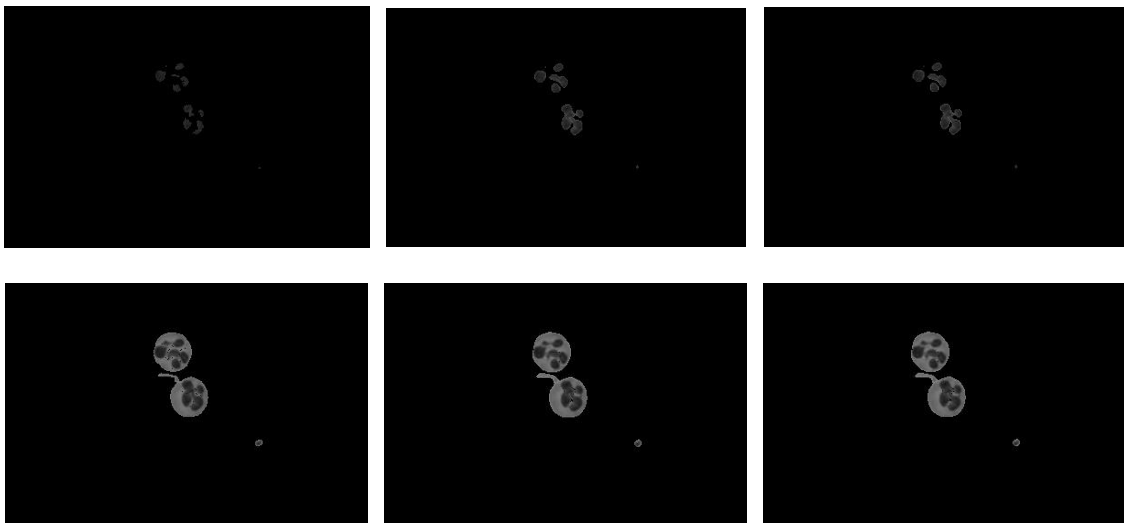


图2-3 迭代更新白细胞核与白细胞边界 (iteration=1, 2, 3)

2.3 轮廓算法的选择

早期的snake算法，Kass Snake的计算公式如下：

$$\begin{aligned} \text{parametric curve: } \mathbf{X}(s) &= (x(s), y(s))', \quad (s = [0,1]) \\ \text{active contour: } E_{snake} &= \int_0^1 E_{snake}(\mathbf{X}(s)) ds \Rightarrow \min \\ \text{energy: } E_{snake}(\mathbf{X}(s)) &= E_{int}(\mathbf{X}(s)) + E_{image}(\mathbf{X}(s)) + E_{con}(\mathbf{X}(s)) \end{aligned}$$

之后的GVF snake，在Kass snake的基础上对外部能量梯度部分做出了改进。但是这些算法实现时需要计算一阶和二阶导数，计算偏微分代价较大，同时局部噪声可能会导致数值计算结果的偏差等。

所以倾向于选择动态轮廓中更具有鲁棒性，计算更快速的算法。

测地线动态轮廓（GAC）将snake公式中的曲线的二阶导删去，因为其对分割结果影响较小。同时引入水平集的概念，使得动态扩展轮廓的过程中，能综合考虑整张图像中位于不同位置的具有相近特征的同一种类物体的边缘特征。对应细胞图像中不同位置的白细胞特征相近（每个白细胞的边缘像素的分布都相近）。相较于传统snake算法，计算量更小，参数数量减少，同时更贴合提取白细胞边界这一问题。

Active Contours like Geodesic Active Contours:

$$\frac{\partial u}{\partial t} = g(I) |\nabla u| \operatorname{div} \left(\frac{\nabla u}{|\nabla u|} \right) + g(I) |\nabla u| \nu + \nabla g(I) \nabla u.$$

对于轮廓边缘是否扩展取决于水平集梯度以及图像梯度，注重局部细节。

Active Contours without Edges:

$$\frac{\partial u}{\partial t} = |\nabla u| \left(\mu \operatorname{div} \left(\frac{\nabla u}{|\nabla u|} \right) - \nu - \lambda_1 (I - c_1)^2 + \lambda_2 (I - c_2)^2 \right).$$

对于轮廓边缘是否扩展取决于水平集梯度以及边缘与轮廓内部/外部均值的接近程度。更适用于颜色分布集中的情况。

以下图2-4是两种算法在图像1上的结果，显然第二种方法效果更佳。

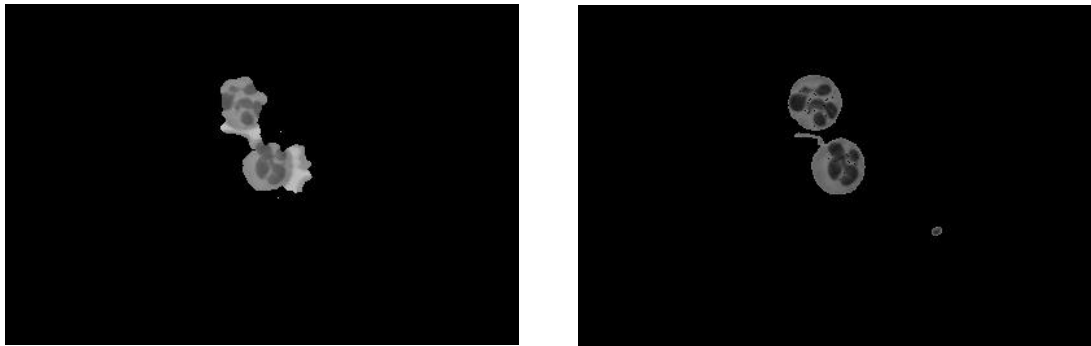


图2-4 snakeGAC与snakeACWE效果对比

所以选择使用snakeACWE作为轮廓算法。

2.4 轮廓算法的改进

虽然snakeACWE基本提取出了白细胞的边界轮廓，但是如果细胞出现粘连的情况，算法的提取会出现偏差，如上图中白细胞边界判定将周围的正常细胞部分包括进来。这一问题的产生是由于边界处的像素梯度较小，所以考虑引入先验知识解决(白细胞的形状为近似圆形)，所以可以在snakeACWE的基础上增加“圆形近似度”项。设计如下：

$$a \sum_{i=0}^I \int_0^1 |C'_i(q) - \overline{C'_i}|^2 dq$$

对于第*i*条轮廓线，其平均曲率为 $\overline{C'_i}$ ，计算轮廓线上曲率的方差。累计*I*条轮廓的曲率方差。设计思想是圆形的曲率处处相等，方差为0，而随着圆形的形变，形状轮廓的曲率方差逐渐增大。但是考虑到实际情况中可能有两个或多个白细胞相连，需要允许非圆形的轮廓，所以没有引入此附加项，而等待之后的后场处理此问题。

2.5 白细胞的分割

通过以上的思路可以得到白细胞的边界。但是存在两个问题：

- 1. 将与白细胞相接的其他区域也纳入范围
- 2. 如果有两个或多个白细胞相接，如何将其分割开以便之后的计数

考虑到每个细胞的形状特点，我们可以近似认为每个细胞的边界轮廓为凸包，而非凸包的轮廓必然包含了两个或多个细胞。

所以设计思路是对于每个轮廓，计算凸包，计算拟合程度如果较差，进行分割。此操作也进行迭代，直到每个轮廓都近似凸包。

记*C*为轮廓的点集合，*C'* 为凸包的点集合。

实现伪代码如下：

```
while ALL( $Area(C'_i) - Area(C_i) > T$ )  
  
    for  $(C_i, C'_i)$  in  $\{(C, C')\}$   
  
        if  $Area(C'_i) - Area(C_i) < T$  continue  
  
         $diffs = Area(C'_i) - Area(C_i)$   
  
        按照面积对  $diffs$  数组进行排序  
  
        取最大和次大区域的轮廓记做  $C_{max1}$  和  $C_{max2}$   
  
        计算  $C_{max1}$  与  $C_{max2}$  的最近点对，连线  
  
        对图像进行分割
```

对应处理过程如下：



图2-5 左:白细胞 右:白细胞(凸包-轮廓)结果

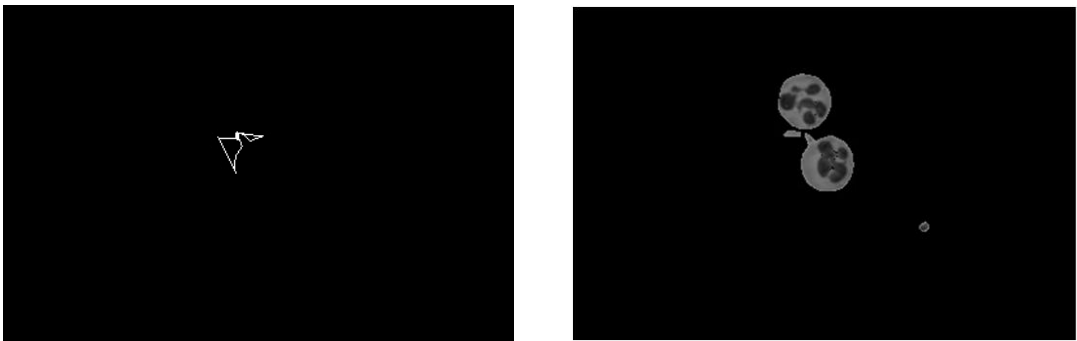


图2-6 左:面积最大区域与次大区域间最短连线 右:白细胞在分割

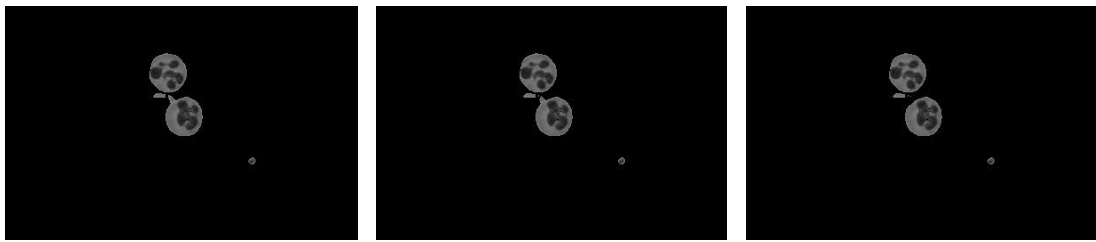


图2-7 迭代分割 左:(iteration=1) 中:(iteration=2) 右:迭代收敛结束

2.6 白细胞的计数

得到了完整的多个轮廓，此时就可以进行白细胞的核占比计算以及计数了。
判断选中区域是否为白细胞基于区域的面积和区域内的核占比，逻辑如下：

```

for each 选中区域 $S_i$ 

    if  $Area(S_i) > T_{Amax}$  or  $Area(S_i) < T_{Amin}$ 

        remove  $S_i$ 

     $P_i = Area(S_{nucleus}) / Area(S_i)$     # $P_i$ 为 $S_i$ 的核占比

    if  $P_i > T_{Pmax}$  or  $P_i < T_{Pmin}$ 

        remove  $S_i$ 
  
```

得到的结果如下图2-8：



图2-8 左:过滤前的白细胞图像 右:过滤后的白细胞图像

过滤之后图像的最外层轮廓个数即为白细胞个数。

2.7 Reference

- [1]*Geodesic Active Contours*, Vicent Caselles, Ron Kimmel and Guillermo Sapiro. In International Journal of Computer Vision (IJCV), 1997, DOI:10.1023/A:1007979827043
- [2]*Active Contours without Edges*, Tony Chan and Luminita Vese. In IEEE Transactions on Image Processing, 2001, DOI:10.1109/83.902291

3. 处理流程图

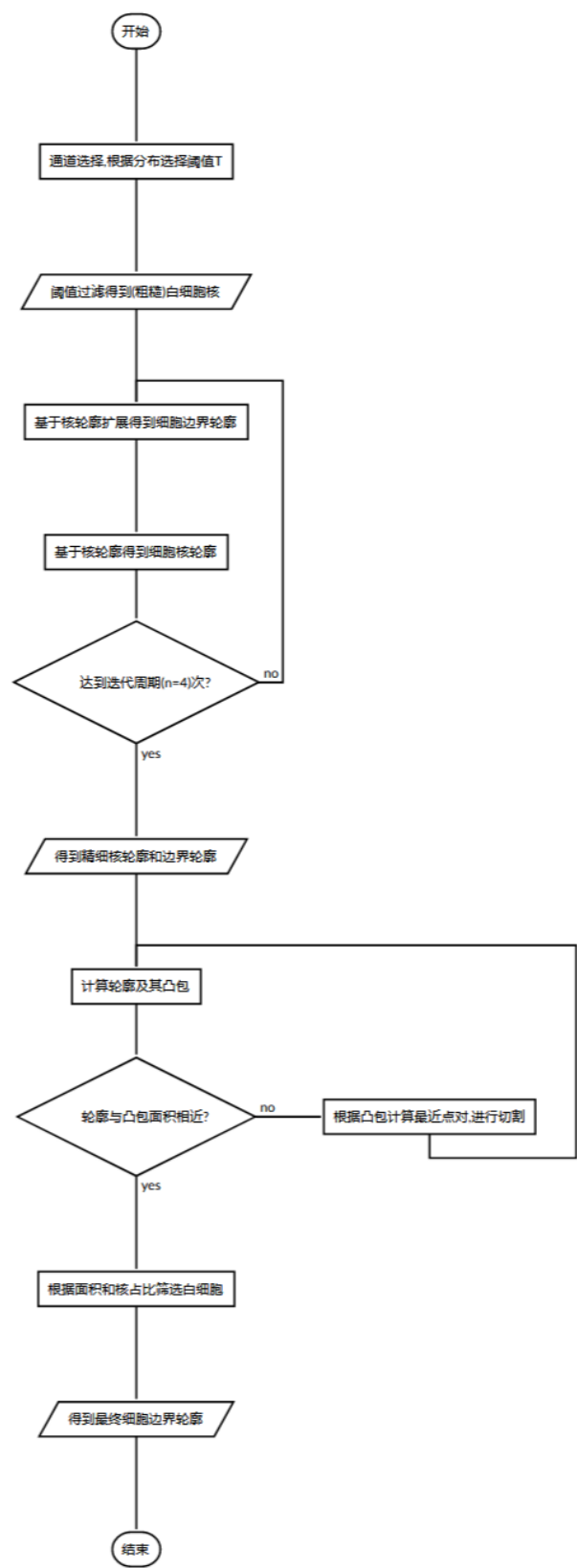


图3-1 算法流程图

4. 参数设置

常量符号	意义	取值	备注
PROPORTION	图像中像素按照分布取第proportion小的值作为阈值，用于提取白细胞核	0.005	[0.001-0.01]都能得到正确的结果,不需要微调
ITERATION	更新白细胞边界和白细胞核的轮廓的迭代次数	4	>2即可,因为一般迭代2轮边界就基本收敛
SUB_ITERATION	在一次轮廓算法中,轮廓变化变形的迭代次数	50	>30即可,因为从核到白细胞边界一般30轮即收敛
END_HULL	区域分割阶段的终止条件,判断该区域与其凸包相似程度	3	如果区域与凸包至少有2处局部面积大于3像素的不同,则需要切割
Ta_max	白细胞筛选阶段,如果区域面积过大,则删除	2100	常规白细胞面积在1400-1800像素之间
Ta_min	白细胞筛选阶段,如果区域面积过小,则删除	1100	常规白细胞面积在1400-1800像素之间
Ta_max	白细胞筛选阶段,如果核占比过大,则删除	0.5	常规白细胞核占比在0.3-0.4之间
Ta_min	白细胞筛选阶段,如果核占比过小,则删除	0.2	常规白细胞核占比在0.3-0.4之间

图4-1 参数设置

5. 实验结果

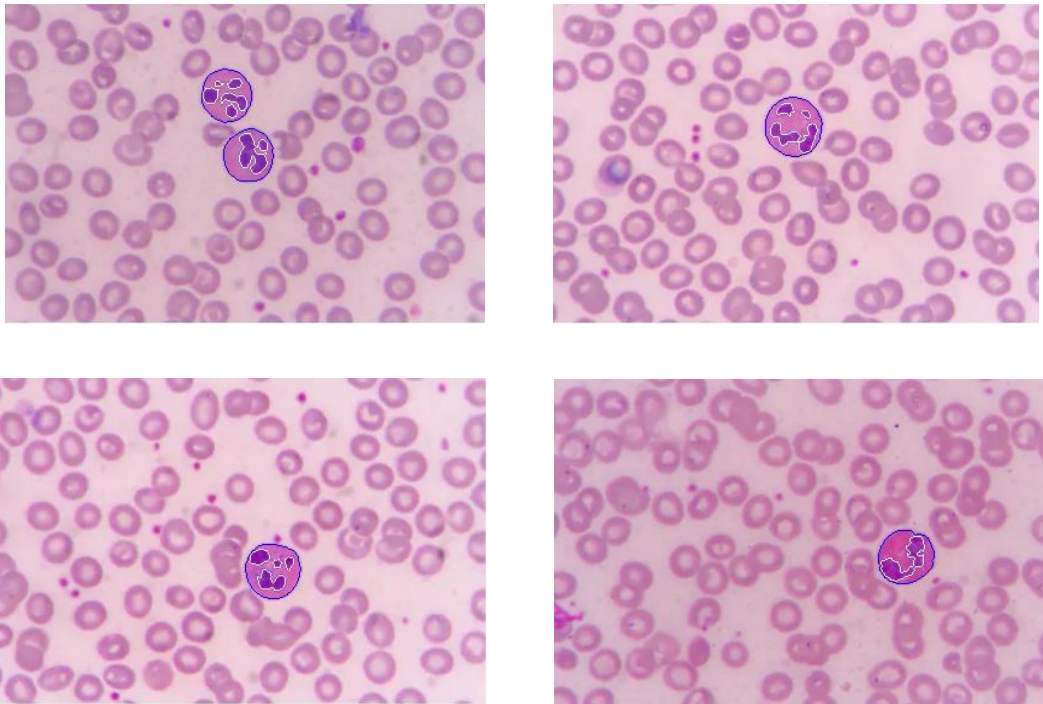


图5-1 4张数据集算法处理结果

图像ID	白细胞核面积/像素个数	白细胞核面积/像素个数	白细胞个数/个	核占比/%
1	2885	1073	2	37.20%
2	1770	609	1	34.41%
3	1704	566	1	33.22%
4	1652	585	1	35.41%
AVG	1602	567	1	35.36%

图5-2 算法结果统计

6. 结果分析

从分割的效果和算法对参数的不敏感性来看，算法是比较好的。

6.1 改进

还有待改进的地方主要是在于边界轮廓的细节处，当白细胞核与边界相距较近时，白细胞的边界判定可能出现轻微偏差。如下图6-1：

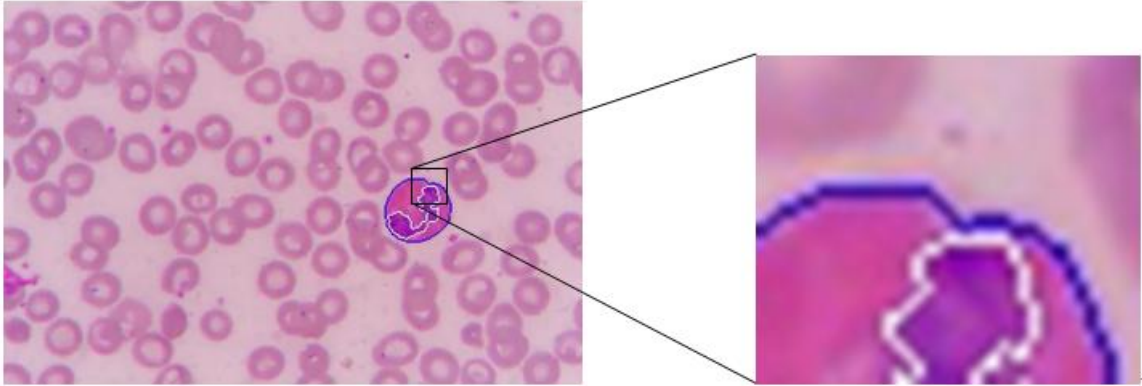


图 6-1 算法处理细节

解决此需要不应该仅仅停留在像素层面的算法处理，应该加入其它先验知识辅助判别更高置信度的曲率。