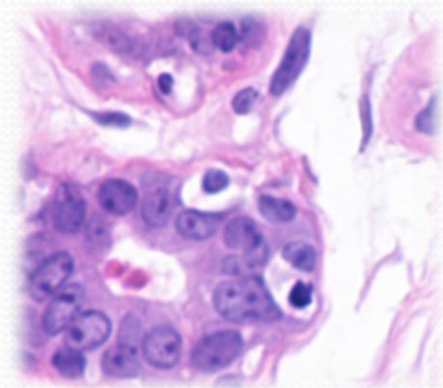


基于活动轮廓模型的病理显微图像 分割若干关键技术研究

刘子纬
数理提高班
华中科技大学

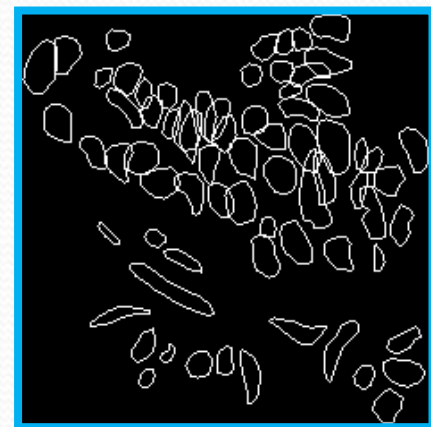
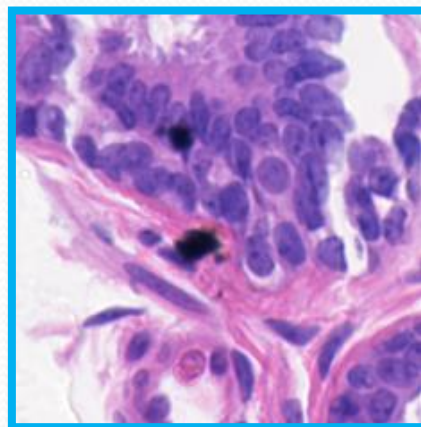
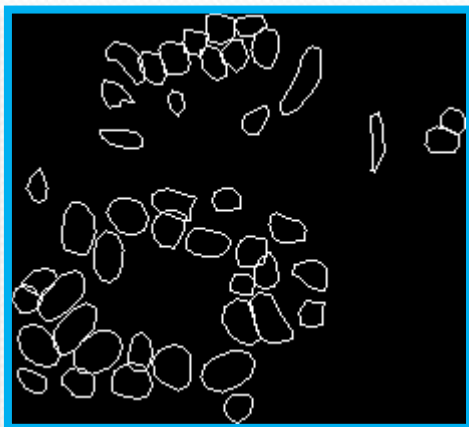
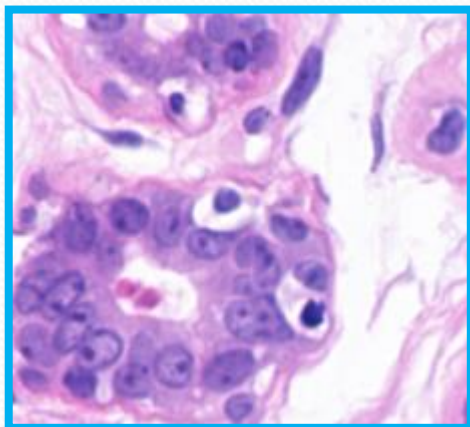
大纲

- 课题研究的背景与意义
- 活动轮廓模型
- 基于多曲率边缘信息的细胞核识别
- 基于稀疏编码的细胞边缘检测
- 未来的工作



课题研究的背景与意义

- 自动化诊断比人工诊断更准确可靠
- 促进统计学习、模式识别的自身发展与完善
- 细胞形态与分布千差万别，有许多尚待深入的研究



课题研究的背景与意义

- 图像分割指的是将数字图像细分为多个图像子区域的过程，使得图像更容易理解和分析。
- 图像分割算法举例：聚类法、直方图法、边缘检测、区域生长、水平集方法、分水岭方法
- 图像分割依据：图像局部特征的相似性和互斥性
- 研究对象：

活动轮廓模型

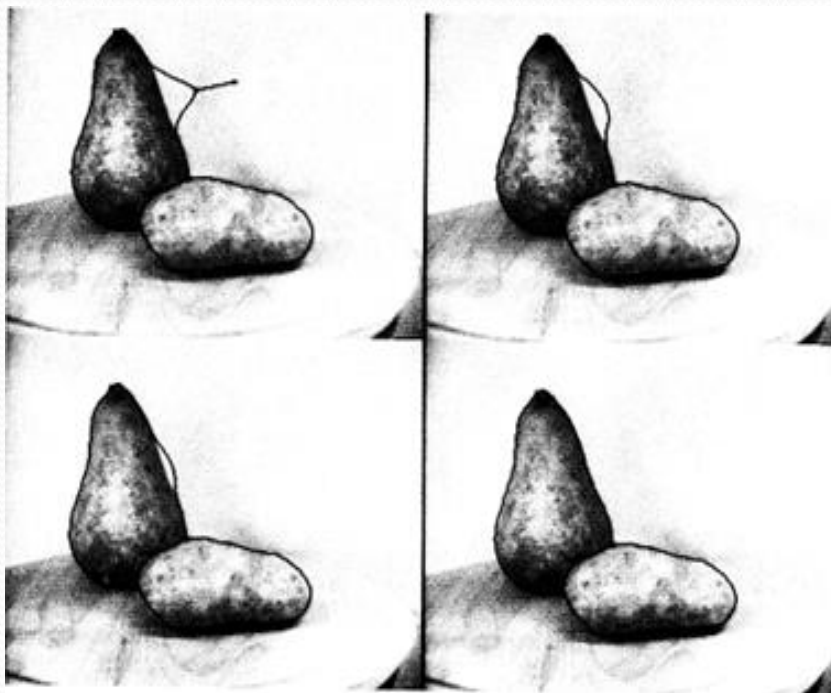
面临的挑战及解决方法：

- 一般方法提取的图像轮廓的精度不高——
活动轮廓模型中能量泛函是在连续状态下实现的
- 一般方法的鲁棒性不强——
活动轮廓模型约束目标轮廓，并融入先验信息
- 一般方法需要繁杂的预/后处理过程——
活动轮廓模型使用光滑闭合曲线表示物体的轮廓

活动轮廓模型

基本思想：

用显式参数化的形式表达曲线，通过极小化能量函数，使曲线逐渐变形，直到抵达分割区域的边界。



活动轮廓模型

数学模型：

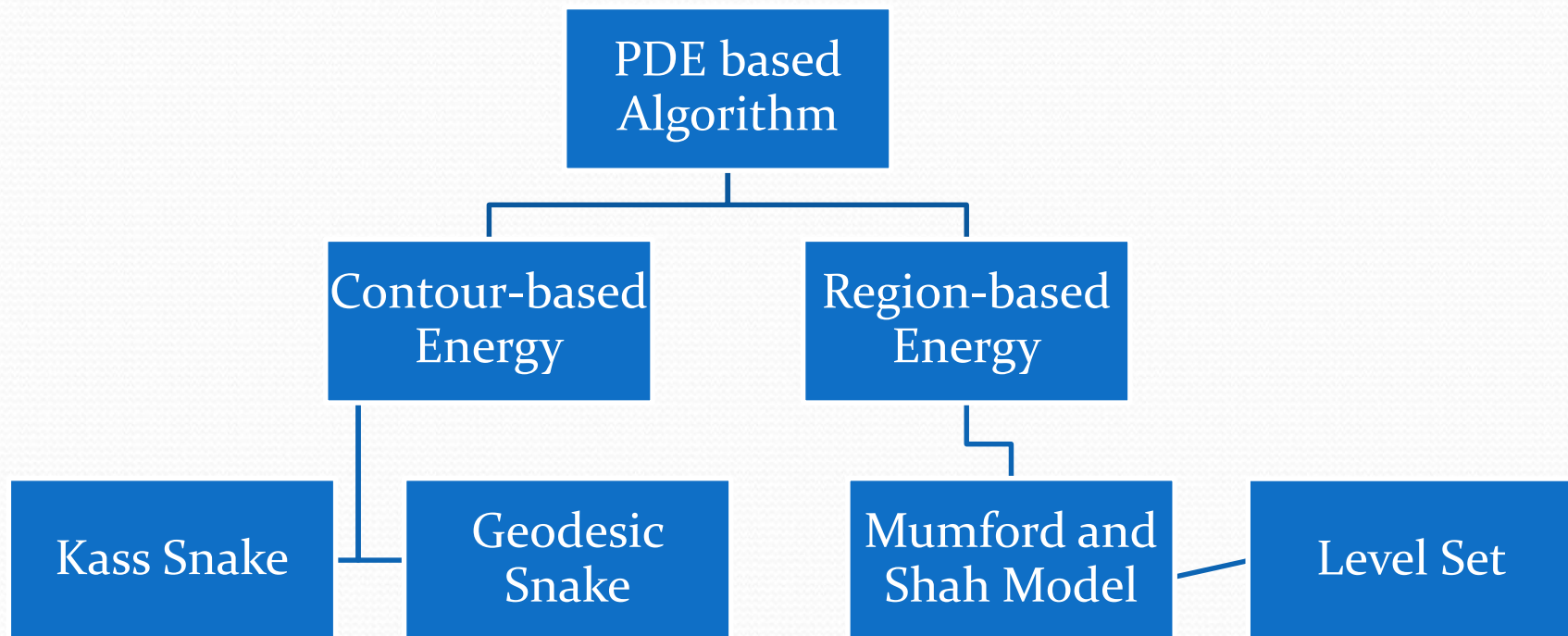
$$E = \int_0^1 \{E_{\text{int}}(X(s)) + E_{\text{ext}}(X(s))\} \cdot ds$$

其中, s 表示弧长参数,活动轮廓在图像中的位置可通过参数向量 $X(s) = [x(s), y(s)]$ 表示

- 内部能量：曲线演化自身的限制条件，同时保持曲线的可变性与光滑性
- 外部能量：图像信息对于曲线演化的限制条件，包括图像的灰度或灰度梯度

活动轮廓模型

分类：依据能量项的定义不同



活动轮廓模型

Kass Snake方法的具体形式:

- 内部能量:

$$E_{\text{int}} = (\alpha(s)|\mathbf{v}_s(s)|^2 + \beta(s)|\mathbf{v}_{ss}(s)|^2)/2 \quad (2)$$

一阶微分 $\mathbf{X}'(s)$ 表达的是活动曲线长度的变化率.

二阶微分 $\mathbf{X}''(s)$ 表达活动曲线曲率的变化率.

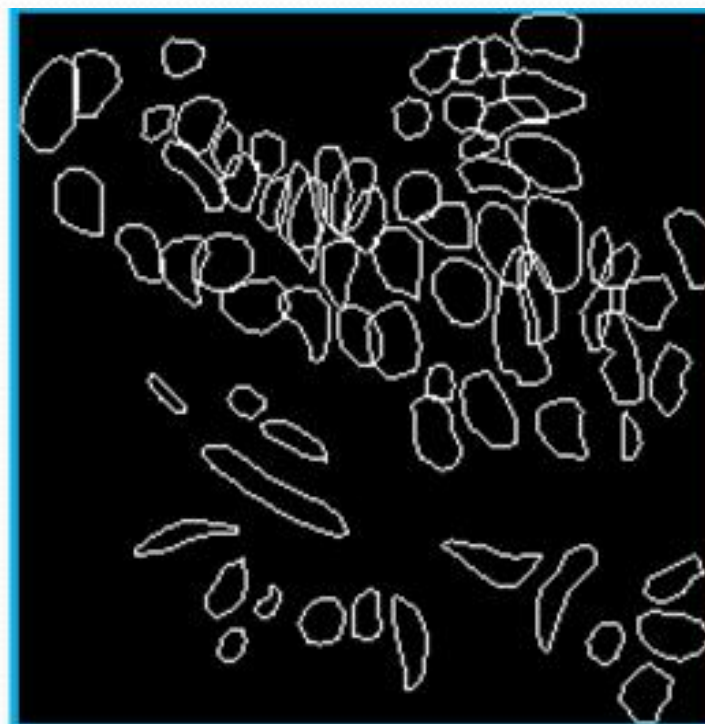
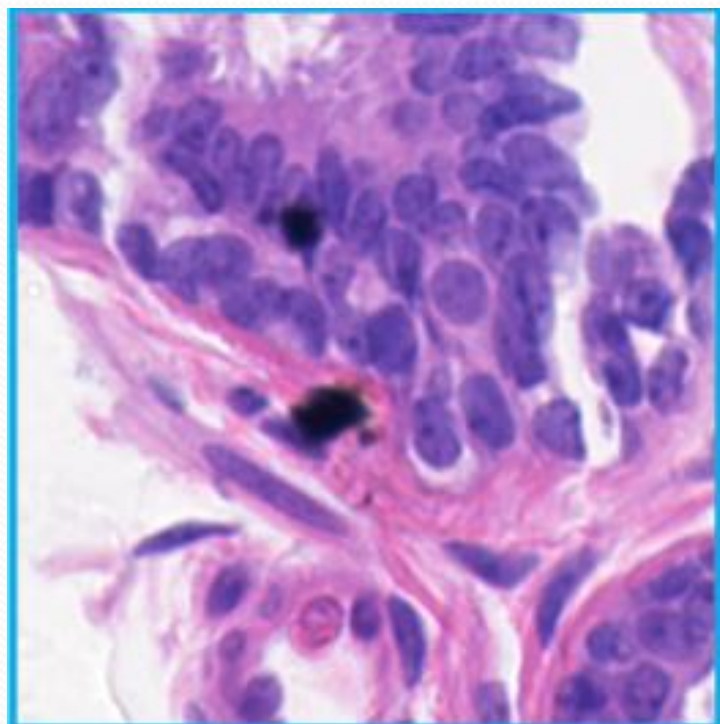
- 外部能量:

$$E_{\text{line}} = -(G_{\sigma} * \nabla^2 I)^2 \quad (5)$$

采用标准偏差为 σ 的Gauss函数来进行卷积低通滤波
图像的灰度梯度

活动轮廓模型

Kass Snake方法的实验结果:



活动轮廓模型

Kass Snake方法的缺点:

- 对初始位置敏感,必须使初始轮廓接近真实的边界
- 模型的非凸性将导致收敛到局部极值点,甚至发散
- 不能处理曲线的拓扑变化
- 解的最优性、计算的稳定性以及无法加入外部强约束

活动轮廓模型

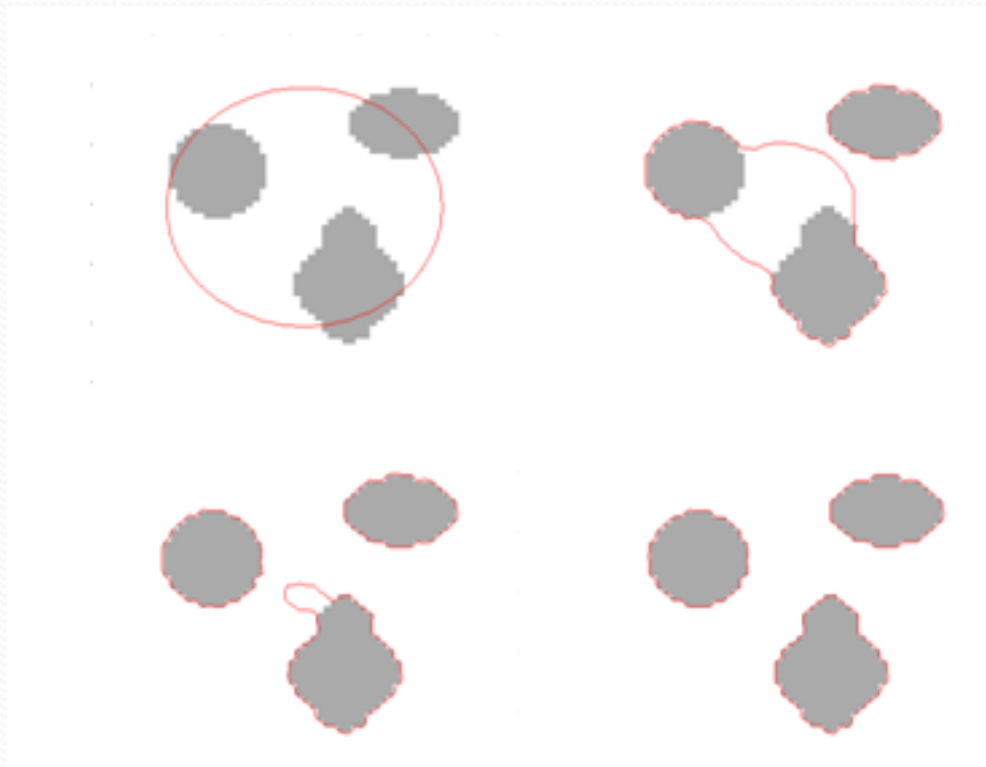
Level Set方法的具体形式:

- 其基本思想是将随时间演化的平面闭合曲线 $C(p, t)$ 隐含地表达为同样随时间演化的三维连续函数曲面 $D(x, y, t)$ 的一个具有相同函数值的同值曲线，通常是 $\{D = 0\}$ ，称为零水平集，而 $D(x, y)$ 称为水平集函数。
- 给定平面上的一条封闭曲线，以曲线为边界，把整个平面划分为两个区域：曲线的外部 and 内部区域。用符号函数定义这条曲线：

$$C(p, t) = \{ (x, y) | D(x, y, t) = 0 \}$$

活动轮廓模型

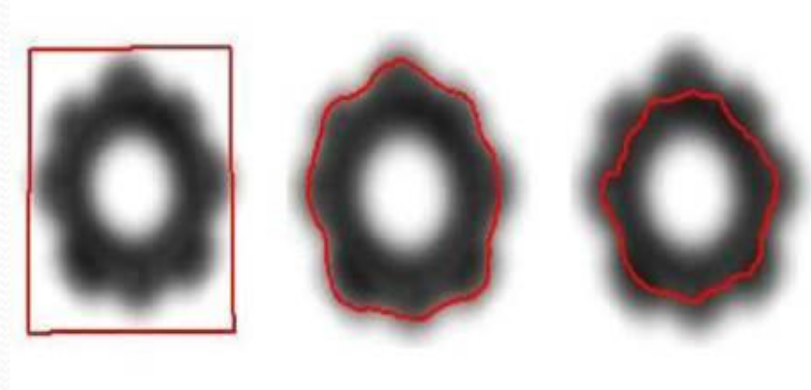
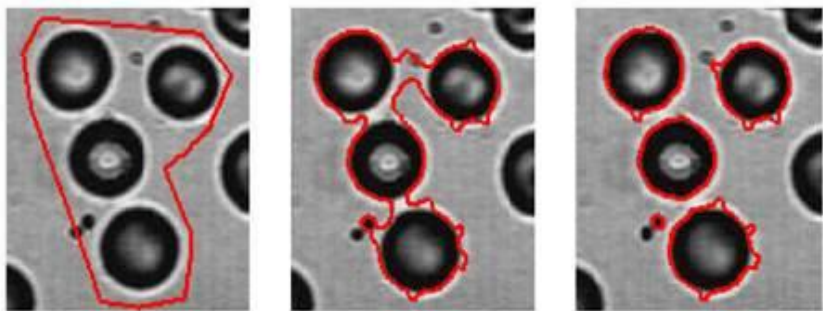
Level Set方法的实验结果:



活动轮廓模型

Level Set方法的缺点：

- 只有速度函数光滑，水平集函数才能保持一个函数形式，因此需要重新初始化；
- 分割结果产生边界泄漏现象；
- 函数复杂，计算费时，分割速度较慢。
- 对强边缘分割结果较好，但对弱边缘分割结果较差



基于多曲率边缘信息的核识别

研究动机:

- 细胞核识别是自动病理诊断中的一个重要步骤;
- 通过视觉检验及人工分割非常耗时且不准确;
- 大多数文献都只是解决了对于普通细胞学图像的分割; 对于病理组织学图像的分割还没有完全令人满意的方

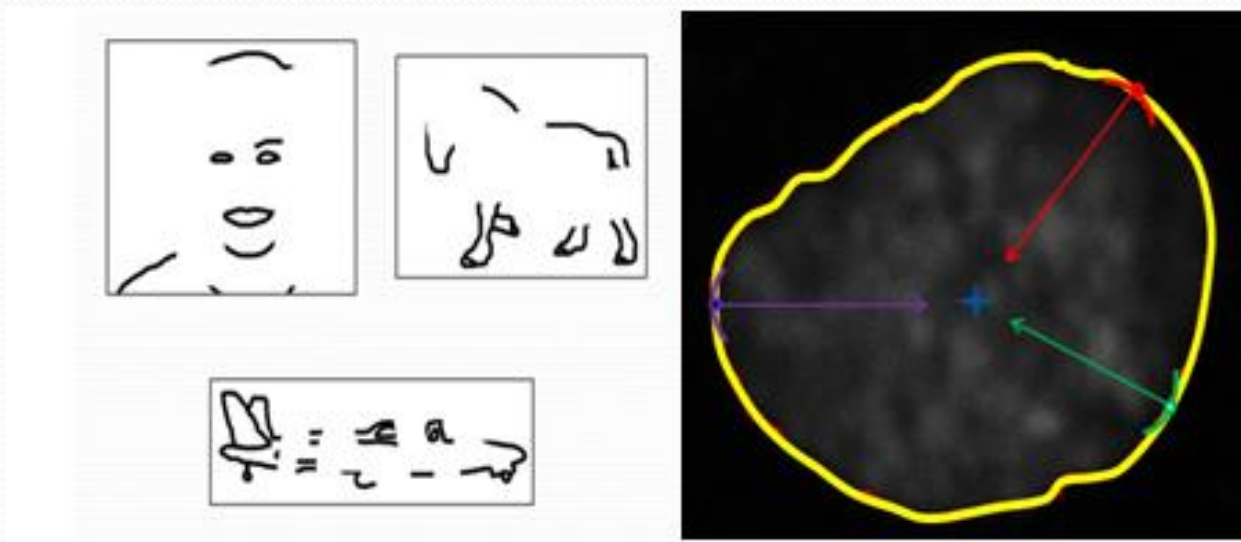
面临的挑战:

- 生物标本内部的多样性;
- 图像背景的杂乱性;
- 细胞之间存在着接触, 甚至重叠。

基于多曲率边缘信息的核识别

基本思想：

心理学研究表明，有时物体边缘（即轮廓片段）是识别物体的最重要信息。且对于细胞核检测这个特定任务来说，轮廓片段还具有作为支撑细胞核存在的证据的作用



基于多曲率边缘信息的核识别

方向能量滤波器：

利用局部的梯度大小信息，同时在 $[0, 2\pi)$ 的方向内计算轮廓片段的曲率。该局部描述子具有一定的对于细胞重叠干扰的鲁棒性。（缺点：缺少旋转不变性）

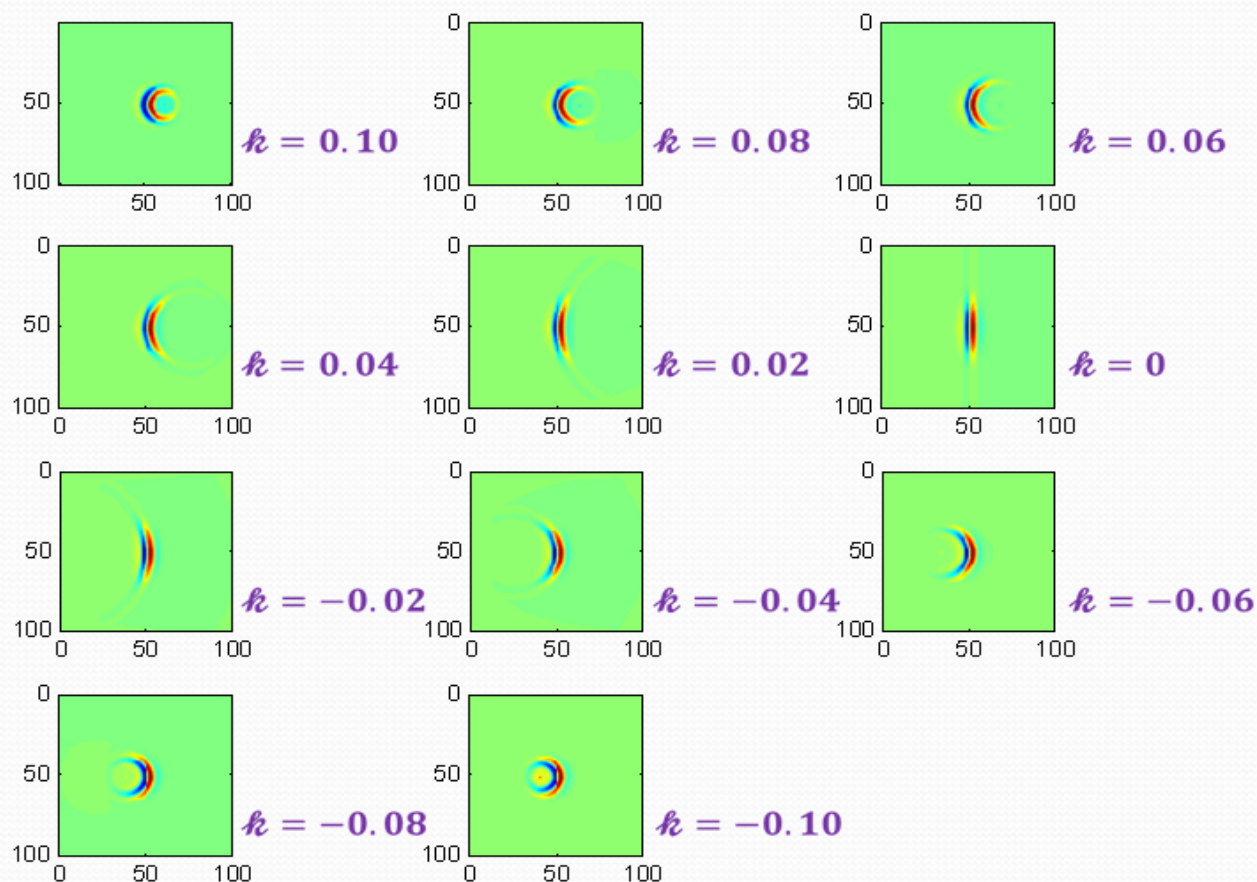
$$f_1(x, y) = \frac{d^2}{dy^2} \left(\frac{1}{C} \exp\left(\frac{y^2}{\sigma^2}\right) \exp\left(\frac{x^2}{l^2 \sigma^2}\right) \right)$$

$$f_2(x, y) = \text{Hilbert}(f_1(x, y))$$



基于多曲率边缘信息的核识别

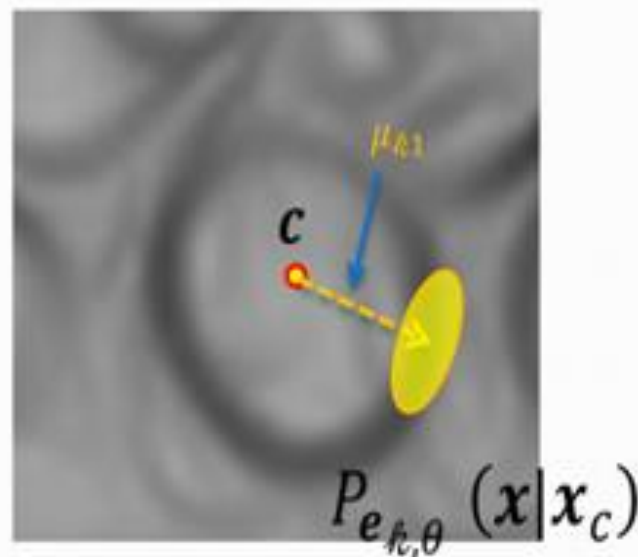
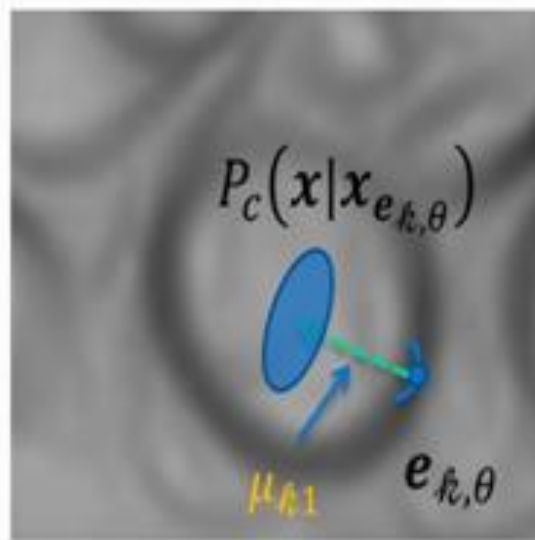
方向能量滤波器：



基于多曲率边缘信息的核识别

细胞核的概率模型:

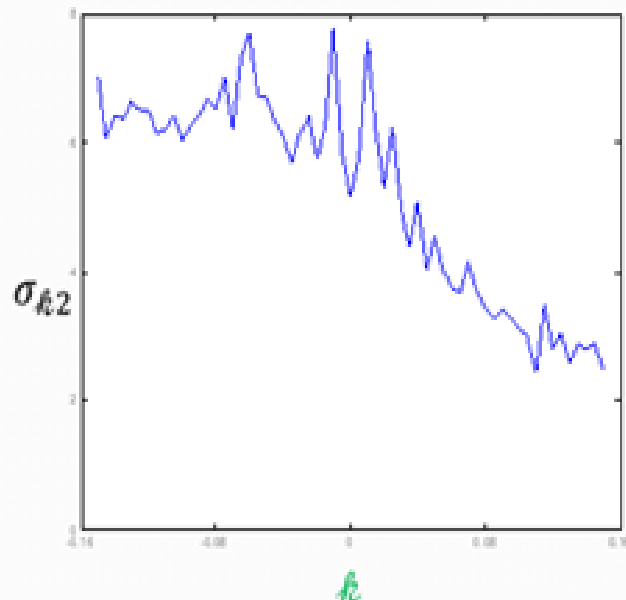
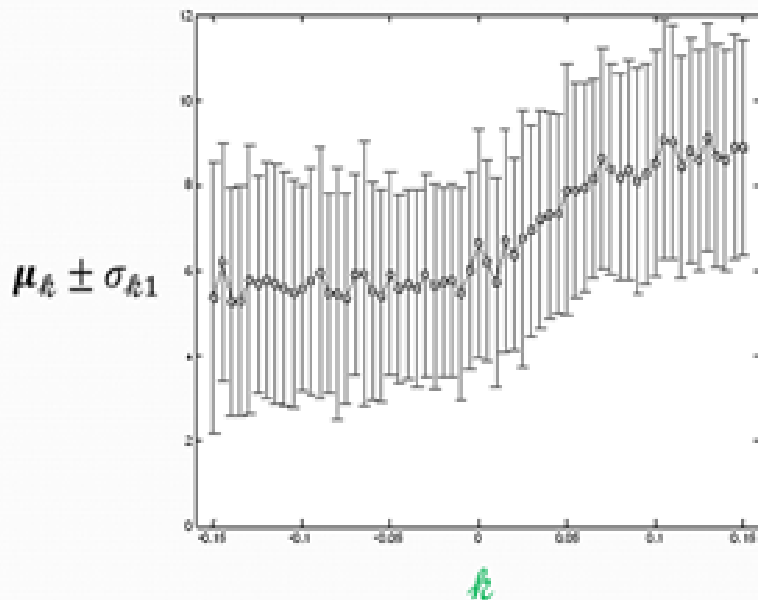
细胞核相对于边缘片段的位置是利用极坐标(θ , r)来表示的, 对于每个曲率半径为 r 的边缘片段, 我们利用高斯分布对细胞核的位置进行拟合。



基于多曲率边缘信息的核识别

实验结果——参数估计：

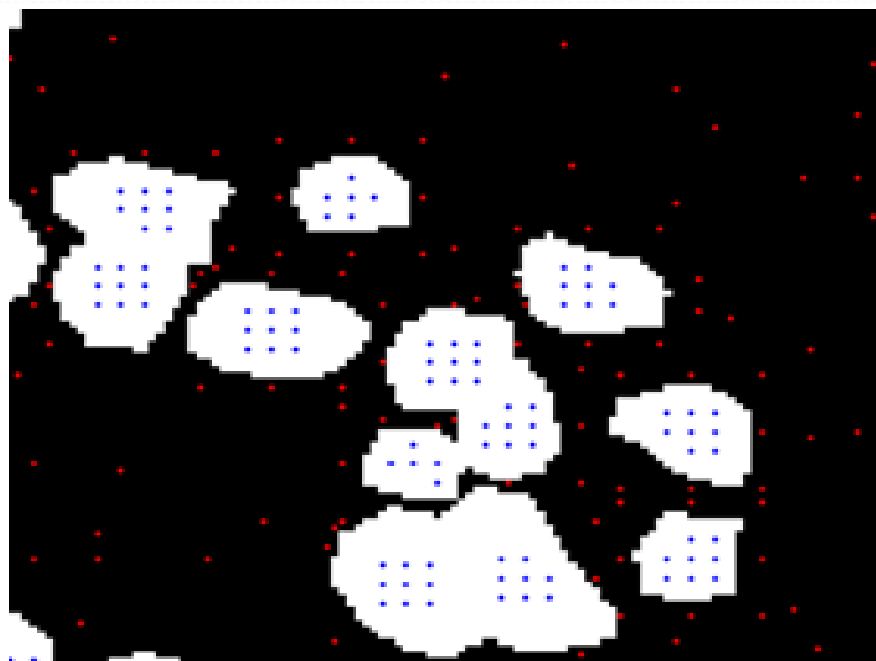
左边的图像显示了细胞核中心沿着边缘片段方向的分布
右边的图像显示了细胞核中心沿着垂直于边缘片段方向的分布



基于多曲率边缘信息的核识别

实验结果——正负样本标记：

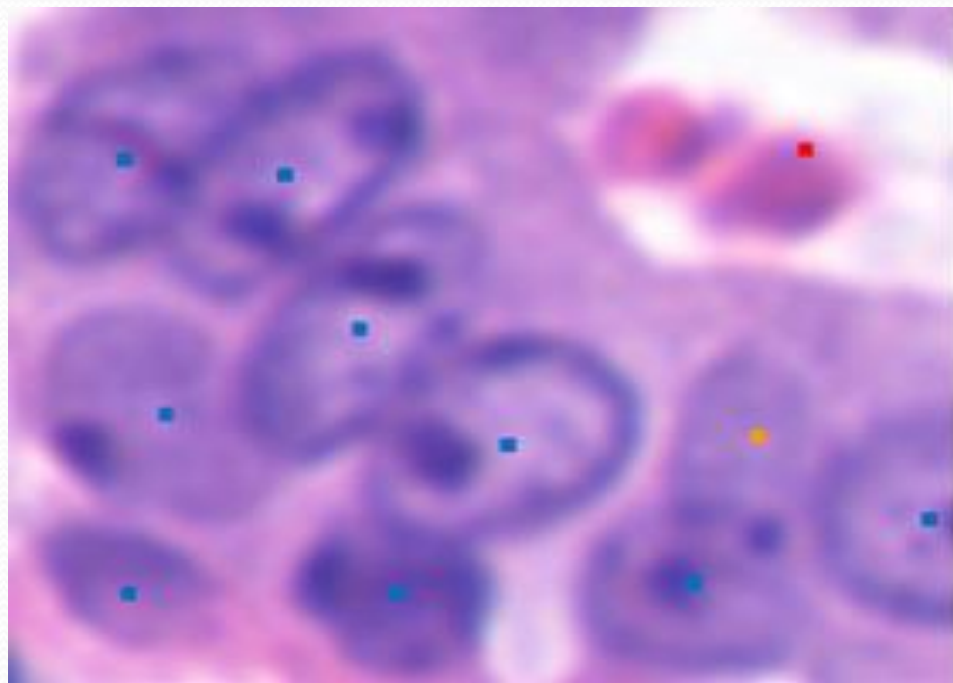
所有的前景像素（细胞核中心）被抽样来作为正样本，而所有的背景像素被抽样作为负样本。此步骤为分类器 Adaboost 的输入



基于多曲率边缘信息的核识别

实验结果——细胞核识别结果：

如图所示，为细胞核检测的结果。蓝色的点代表正确的检测结果，红色的点代表不正确的检测结果。由图可知本方法具有较高的正确率



基于稀疏编码的细胞边缘提取

基本思想：

- 图像信息的“有效”表示是图像处理任务的基础。“有效”表示是指能够用较少的系数捕获感兴趣目标重要信息的能力，即稀疏表示能力。
- 超完备字典中的冗余基取代了传统方法中的正交基，而字典的选择应尽可能的包含被表达信号所含有的信息结构。信号的稀疏分解即从超完备字典中选择具有最佳线性组合的若干原子来表示信号，实际上它是一种逼近过程：

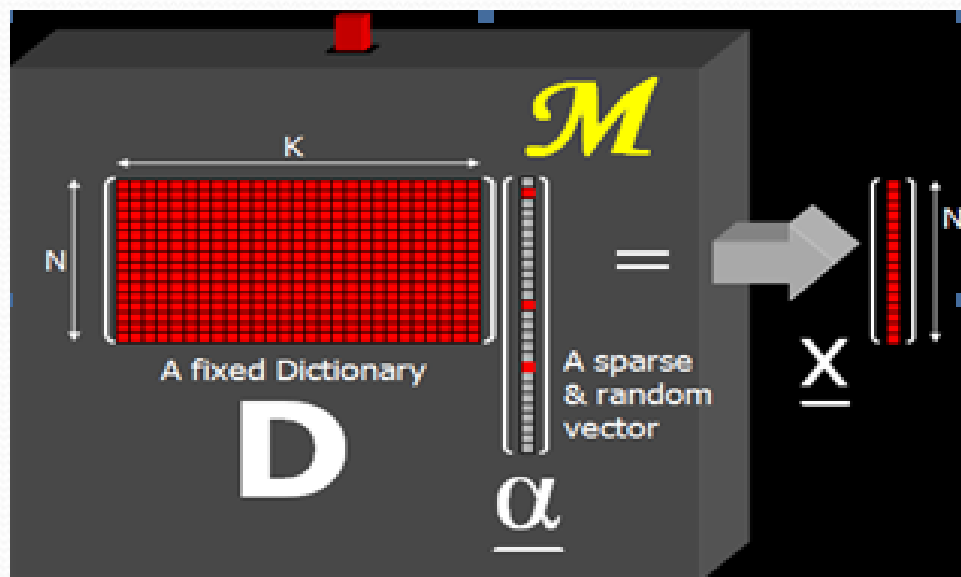
$$f = f_m + f_r = \sum_{k \in \Omega_m} \alpha_k d_k + f_r$$

基于稀疏编码的细胞边缘提取

基本思想：

从稀疏角度出发，希望在 f_r 达到最小的情况下得到 α 最稀疏的一个解。可构造优化函数如下：

$$\min \|\alpha\|_0, s.t. \left\| f - \sum_{k \in \Omega_m} \alpha_k d_k \right\| \leq \varepsilon$$



基于稀疏编码的细胞边缘提取

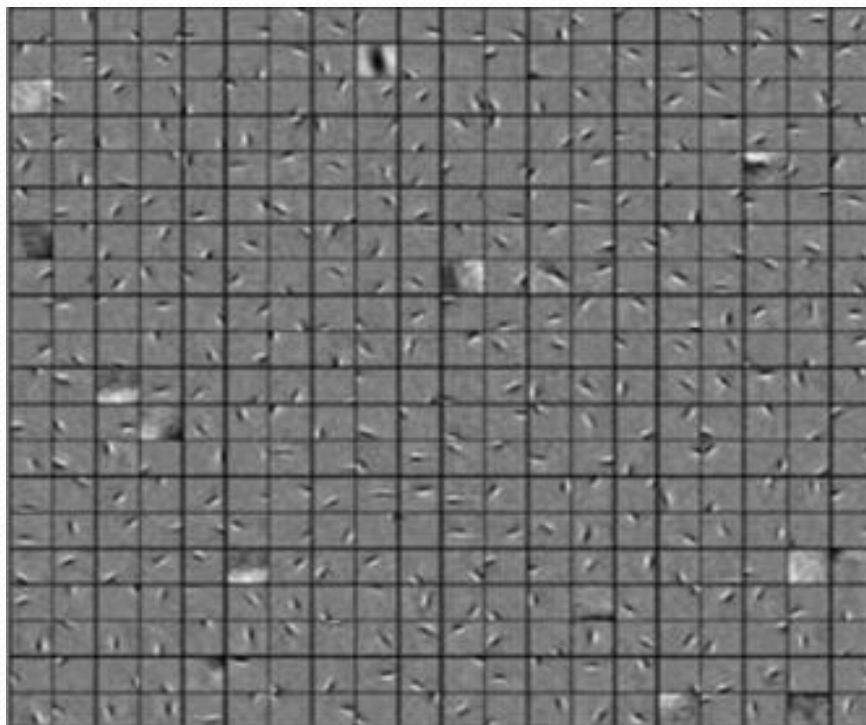
自适应超完备字典构造算法——K-SVD:

- 初始化阶段:
 - 构造初始超完备库
 - 构造示例信号集合
- 训练更新阶段:
 - 选出信号子集
 - 计算残差
 - 构造梯度算子
 - 更新原子

基于稀疏编码的细胞边缘提取

实验结果——字典的学习：

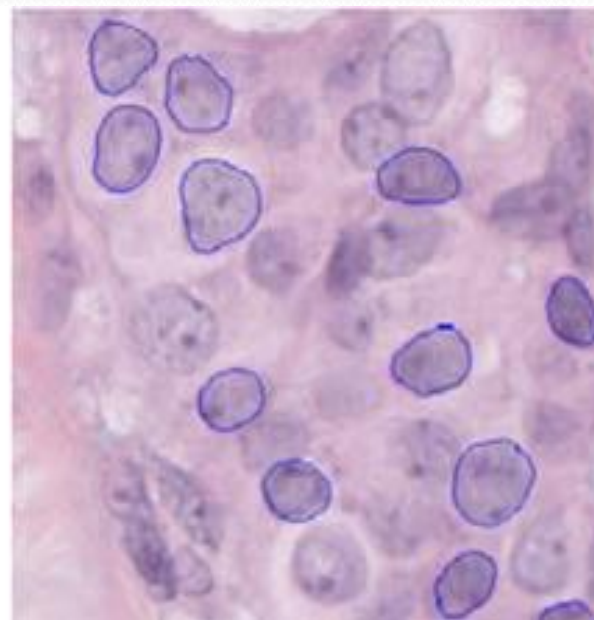
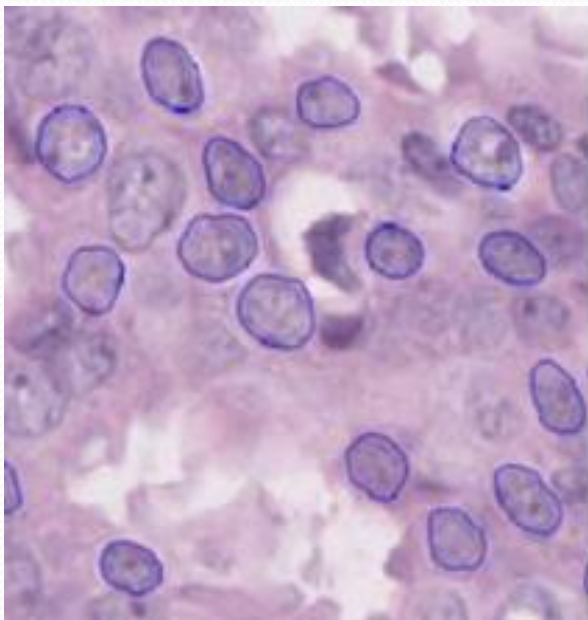
通过将训练集中的图像分成一定大小的**patches**，然后利用**K-SVD**算法，可以在训练集中学出一套关于细胞边缘的模版



基于稀疏编码的细胞边缘提取

实验结果——细胞边缘提取：

利用上一个步骤中学到的字典，对一副测试图像进行稀疏表示，并找出边缘



未来的工作

- 在以下3个数据集上的完整测试：
 - Murphy 3T3 & U20S Nuclei Datasets from Murphy Lab, CMU
 - Breast Cancer & COS1 Kidney Cell Datasets in UCSB Bio-segmentation Benchmark Datasets from UCSB
 - Cell Segmentation Datasets from Lezoray
- 根据细胞学的理论与知识，除所讨论的形态、边界外，还有其它十分重要的特征能够描述细胞，例如纹理、光学特征与颜色特征等。针对这些特征，还须要与染色方法的改进相结合，进行更为细致、全面的分析。

未来的工作

- 有效地利用细胞浆所提供的信息，能够进一步增加**DNA**定量化自动诊断方法的敏感性，并弥补其特异性不足的问题。因此，如何有效地提取细胞浆的位置、轮廓与特征参数，并为诊断提供依据是将来研究的重要与难点
- 对于稀疏编码进行更加深入的研究，或尝试融入小波特征，争取在计算机视觉和机器学习理论方面作出贡献

谢谢！