

## 基于CT图像的肺结节检测与识别

唐思源<sup>1</sup>, 刘燕茹<sup>2</sup>, 杨敏<sup>1</sup>, 徐瑞英<sup>1</sup>

1. 内蒙古科技大学包头医学院计算机科学与技术系, 内蒙古 包头 014040; 2. 内蒙古科技大学包头医学院医学技术系, 内蒙古 包头 014040

**【摘要】目的:**将肺结节从含有背景、噪声的胸腔区域里检测并识别出来。**方法:**首先,将DICOM格式的医学图像转换成JPG图像后,应用区域生长法分割出肺实质区域,去掉肺区外的干扰信息。然后,利用多尺度高斯滤波器增强图像后,应用模糊C均值聚类算法提取肺结节感兴趣区域。最后,对肺结节特征进行提取及归一化处理,应用支持向量机分类器识别并标记出肺结节。**结果:**在随机抽取的120例图像中,检测肺结节的准确率达到92.3%,分类识别肺结节的准确率达到95.6%。实验结果表明,本文方法有效地排除了交叉状和条形状血管等干扰,实现了肺结节的精确检测和识别。**结论:**本方法在保证检测和识别出正确结节的前提下,降低了误判率,算法也得到了较好的收敛。

**【关键词】**肺结节;CT图像;区域生长法;多尺度高斯滤波器;模糊C均值聚类算法;支持向量机分类器

**【中图分类号】**R318;TP391.41

**【文献标志码】**A

**【文章编号】**1005-202X(2019)07-0800-08

## Detection and recognition of pulmonary nodules based on CT images

TANG Siyuan<sup>1</sup>, LIU Yanru<sup>2</sup>, YANG Min<sup>1</sup>, XU Ruiying<sup>1</sup>

1. Department of Computer Science and Technology, Baotou Medical College of Inner Mongolia University of Science & Technology, Baotou 014040, China; 2. Department of Medical Technology, Baotou Medical College of Inner Mongolia University of Science & Technology, Baotou 014040, China

**Abstract: Objective** To detect and identify pulmonary nodules from thoracic regions with background and noise. **Methods** After DICOM-format medical images were converted into JPG images, region growing method was applied to segment lung parenchyma and remove interference information outside lung area. Subsequently, multi-scale Gaussian filter was used to enhance images, and fuzzy C-means clustering algorithm was applied to extract regions of interest of pulmonary nodules. Finally, the features of pulmonary nodules were extracted and normalized, and the pulmonary nodules were identified and marked with support vector machine classifier. **Results** For the random sample of 120 images, the detection rate of pulmonary nodules reached 92.3% and the accuracy rate of the classification and recognition of pulmonary nodules was up to 95.6%. The experimental results revealed that using the proposed method could effectively eliminate the disturbances from crossing- and strip-shaped blood vessels and other disturbances, realizing an accurate detection and recognition of pulmonary nodules. **Conclusion** Using the proposed method can not only achieve an accurate detection and recognition of pulmonary nodules, but also reduce misjudgment rate. Moreover, the proposed algorithm has a better convergence.

**Keywords:** pulmonary nodule; CT image; region growing method; multi-scale Gaussian filter; fuzzy C-means clustering; support vector machine classifier

## 前言

**【收稿日期】**2018-12-17

**【基金项目】**内蒙古自治区自然科学基金(2016MS0601);包头医学院科学研究基金(BYJJ-QM 201637);包头医学院大学生创新创业训练计划项目(BYDCXL-201922)

**【作者简介】**唐思源, 硕士, 副教授, 研究方向: 医学图像处理, E-mail: 617682453@qq.com

随着我国空气污染的加重及雾霾天气的频发,以及吸烟人群的逐年增加,肺癌的发病率及死亡率不断上升。有关早期肺癌的治疗得到研究者的高度重视,肺癌早期病灶是以肺结节(Lung Nodules)的形式表现,因此,精确检测出肺结节,准确定位出病灶的位置,并对肺结节进行特征提取及分析,是发现早期肺癌的关键。目前,CT影像扫描技术是检测肺结节最有效的方法,放射科医生需要从成百上千张胸部CT影像序列中找出可疑病灶区域,在长时间重复

阅片的繁重工作下,容易产生视觉疲劳,对病灶区域的判断可能会造成误诊或者漏诊等情况。

从20世纪90年代开始,计算机辅助检测系统(Computer-Aided Detection, CADe)开始应用到医学影像方面的研究<sup>[1]</sup>,在早期肺癌筛查方面,该系统可以定量检测出肺部区域特征并定位出可疑结节区域,为放射科医生提供可靠的参考信息,同时还可以减轻医生阅片的工作量,辅助放射科医生提高检测肺结节的准确率、降低假阳性率。

目前,国内外专家学者提出很多有效的肺结节检测算法,陈树越等<sup>[2]</sup>提出使用K-means聚类算法对肺结节的细长度、类圆形度等几何特征进行多次的聚类,对识别孤立性肺结节达到一个比较理想的效果,降低误判率。Alilou等<sup>[3]</sup>和Demir等<sup>[4]</sup>国外专家学者利用肺结节的3D特征,在分割肺部区域及疑似肺结节区域的基础上,计算疑似肺结节区域的质量中心点,并评价其偏移程度,最终确定疑似区域是否为阳性肺结节。Stember等<sup>[5]</sup>提出用伪分子运动模型(NMA-SD)模拟肺结节形状,实验表明对圆形、椭圆形及不规则形的肺结节识别准确率达到97%。尽管已有很多研究者提出更多的检测方法,但是还是没有达到应用到临床的标准,主要原因在于检测的结节受到周围血管等组织的干扰,检测出的结节含有非结节组织。为了更精确地检测出肺结节,本文提出一种基于增强滤波器和三维特征的肺结节检测与识别方法,滤波器用于增强结节与血管之间的密度、灰度对比度值。构建的三维图像能全面展示肺结构组织的细节信息,可明显区分与血管交叉部分结节的检测困难等问题,提高检测准确率。

## 1 算法原理

### 1.1 建立三维形状模型

利用高斯分布模型建立分别代表肺结节(球体)、血管(圆柱体)、面的3种模型结构,表达式如公式(1)所示,其中 $\sigma$ 代表不同形状的尺寸大小。

$$\begin{aligned} \text{球状: } s(x, y, z) &= \exp\left\{-\frac{x^2 + y^2 + z^2}{2\sigma^2}\right\} \\ \text{圆柱体: } c(x, y, z) &= \exp\left\{-\frac{x^2 + y^2}{2\sigma^2}\right\} \\ \text{面状: } p(x, y, z) &= \exp\left\{-\frac{x^2}{2\sigma^2}\right\} \end{aligned} \quad (1)$$

球和圆柱体模型示意图如图1所示。

### 1.2 构造增强滤波器

高婷等<sup>[6]</sup>提出由海森(Hessian)矩阵构造滤波器,Hessian矩阵 $H$ 由任一点的9个三阶可导函数构成,用矩阵表示为:

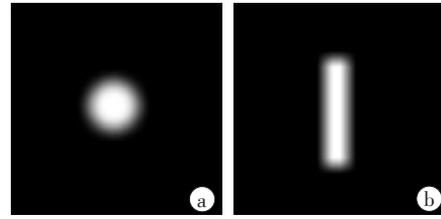


图1 球(a)和圆柱体(b)模型图

Fig.1 Spherical (a) and cylindrical (b) models

$$H = \nabla^2 f = \begin{Bmatrix} f_{xx} & f_{xy} & f_{xz} \\ f_{yx} & f_{yy} & f_{yz} \\ f_{zx} & f_{zy} & f_{zz} \end{Bmatrix} \quad (2)$$

设 $\lambda_1$ 、 $\lambda_2$ 、 $\lambda_3$ 为 $H$ 的特征值,其中要求 $|\lambda_1| > |\lambda_2| > |\lambda_3|$ ,对应的特征向量分别为 $e_1$ 、 $e_2$ 、 $e_3$ ,对应的关系用公式表示为:

$$H = [e_1 \ e_2 \ e_3] \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} \begin{bmatrix} e_1^T \\ e_2^T \\ e_3^T \end{bmatrix} = \lambda_1 e_1 e_1^T + \lambda_2 e_2 e_2^T + \lambda_3 e_3 e_3^T \quad (3)$$

球状模型、管状模型和面状模型特征值满足如下关系:

$$\text{球状: } \lambda_1 \approx \lambda_2 \approx \lambda_3 \approx -\frac{1}{\sigma^2} < 0 \quad (4)$$

$$\text{管状: } \lambda_1 \approx \lambda_2 \approx -\frac{1}{\sigma^2} < 0, \lambda_3 \approx 0 \quad (5)$$

$$\text{面状: } \lambda_1 \approx -\frac{1}{\sigma^2} < 0, \lambda_2 \approx \lambda_3 \approx 0 \quad (6)$$

为了提高区分结节和血管的能力,De Carvalho等<sup>[7]</sup>和Iqbal等<sup>[8]</sup>设计了增强结节和血管图像的两种函数,经过一系列的转换及简化得出结节滤波器( $E_{\text{ball}}$ )和血管滤波器( $E_{\text{line}}$ )增强函数的表达式为:

$$\begin{aligned} E_{\text{ball}} &= \begin{cases} \frac{4\lambda_3/\lambda_1}{1 + \left(\frac{\lambda_2}{\lambda_1}\right)^2 + \left(\frac{\lambda_3}{\lambda_2}\right)^2 + \left(\frac{\lambda_3}{\lambda_1}\right)^2}, & (|\lambda_1| \geq |\lambda_2| \geq |\lambda_3|, \lambda_1 < 0, \lambda_2 < 0, \lambda_3 < 0) \\ 0, & \text{otherwise} \end{cases} \\ E_{\text{line}} &= \begin{cases} \exp(\lambda_3) \times \frac{2\lambda_2/\lambda_1}{1 + \left(\frac{\lambda_2}{\lambda_1}\right)^2}, & (|\lambda_1| \geq |\lambda_2| \geq |\lambda_3|, \lambda_1 < 0, \lambda_2 < 0, \lambda_3 \rightarrow 0) \\ 0, & \text{otherwise} \end{cases} \end{aligned} \quad (7)$$

### 1.3 区域生长算法原理

区域生长算法基本思想是:从选择好的一组种子点开始把属性值(灰度级、颜色、纹理等)相似的像素或子区域按照预定义好的生长准则合成较大区域的过程<sup>[9-10]</sup>。区域生长的关键问题是相似性准则的选择和区域生长中停止规则的设定。准则的选取可以通过图像的亮度值、纹理、色彩等相似性设定,当所选区域的像

素不满足相似性准则的时候,停止区域的生长,分割出目标区域。基于区域生长方法计算简单,可实现肺部区域的初始分割,获得区域的初始轮廓信息。

### 1.4 模糊C均值聚类算法原理

模糊C均值聚类(Fuzzy C-Means, FCM)算法是由Bezdek等提出的。该方法对图像进行分割的思想是:不断寻找聚类的中心  $v_i$  和隶属度的函数  $u_{ik}$ ,当  $J_{FCM}$  值及误差达到一个临界最小值,就实现了图像分割<sup>[11-14]</sup>。计算目标函数  $J_{FCM}$  函数的表达式为:

$$J_{FCM} = \sum_{i=1}^c \sum_{k=1}^n u_{ik}^p \|y_k - v_i\|^2, \quad (9)$$

$(k = 1, 2, \dots, N; i = 1, 2, \dots, c; p \in [1, \infty))$

其中,  $c$  表示聚类个数,  $n$  表示所有像素值,  $u_{ik}$  表示第  $i$  个聚类中第  $k$  个样本的隶属度函数矩阵,  $p$  表示控制模糊度的常数,通常情况下  $p=2$ 。  $y_k$  表示像素灰度值的集合,  $v_i$  表示聚类中心的点集合。

在满足  $i = \sum_{i=1}^c u_{ik} = 1$  的条件下,不断迭代使  $J_{FCM}$  取得最小的必要条件是  $v_i$ 、 $u_{ik}$  要满足如下公式:

$$v_i = \frac{\sum_{k=1}^n u_{ik}^p y_k}{\sum_{k=1}^n u_{ik}^p} \quad (10)$$

$$u_{ik} = \frac{(1/\|y_k - v_i\|^{2/(p-1)})}{\sum_{j=1}^c (1/\|y_k - v_j\|^{2/(p-1)})} \quad (11)$$

在迭代过程中,当  $u_{ik}$  的值较高的时候,像素点就找到匹配的聚类中心,最终对每个像素点寻找最大  $u_{ik}$  值,若满足  $u_{ik} > u_{jk}$ ,则将  $y_k$  划到  $i$  类中,不断迭代将目标提取出来,完成图像的分割。

### 1.5 支持向量机(Support Vector Machine, SVM)分类方法原理

SVM分类方法原理是:把不同类别的数据从不可分类的低维空间映射到高维空间,在特征空间中求得最优解,使得不同类别的数据在超平面上的距离达到最大的过程<sup>[15]</sup>。设有一组训练样本集  $T$ :

$$T = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}, x_i \in X = R^n, 1 \leq i \leq N \quad (12)$$

应用训练样本集构造超平面,构造超平面的函数为:

$$w^* \cdot x + b^* = 0 \quad (13)$$

其中,  $w^*$  表示超平面的法向量,  $b^*$  表示超平面的截距。

在构造的多个超平面中求解最优分类平面,函数公式为:

$$Q(a) = \min_a \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n y_i y_j a_i a_j \exp(-\gamma \|x_i - x_j\|) - \sum_{i=1}^n a_i, i = 1, 2, \dots, n \quad (14)$$

$$\sum_{i=1}^n y_i a_i = 0, 0 \leq a_i \leq c$$

然后利用分类函数:

$$f(x) = \text{sign} \left[ \sum_{i=1}^n y_i a_i \exp(-\gamma \|x_i - x_j\|) + b^* \right] \quad (15)$$

求解得到最佳的数据分类效果。

## 2 肺结节自动检测与识别

在得到DICOM格式的CT图像后,首先把DICOM格式的图像转换成JPG格式;然后,把完整的肺实质区域精确分割出来;最后,检测并识别出肺结节,图2给出肺结节自动检测与识别过程的流程图。

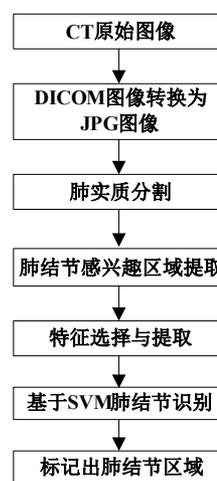


图2 肺结节检测与识别流程图  
Fig.2 Flow chart of pulmonary nodules detection and identification

### 2.1 图像格式转换

CT设备扫描的DICOM格式图像不能直接在普通电脑屏幕上显示,要转换成JPG格式,本文应用垂直镜像变换进行图像格式的转换。垂直镜像公式表示为:

$$G_m(n_0, k_0) = G_0(-n_0, k_0) \quad (16)$$

其中,  $G_0(-n_0, k_0)$  表示DICOM格式原始图像  $I_0$  中像素点  $(-n_0, k_0)$  的像素值,  $G_m(n_0, k_0)$  表示垂直镜像后JPG格式图像  $I_m$  中像素点  $(n_0, k_0)$  的像素值。

### 2.2 基于区域生长的肺实质分割

肺实质分割是肺结节检测的前提条件,转换后的CT影像,在肺区周围还存在骨骼、衣物、设备等干扰,笔者主要提取肺区内部的肺结节,因此要去掉肺区外的噪声信息。肺实质的分割流程如图3所示。

**2.2.1 OSTU法二值化图像** OSTU方法是根据阈值把图像区域分为两组,当被分成的两个区域方差达到最大时,用得到最佳分割阈值来对图像进行分割<sup>[16]</sup>。应用MATLAB软件中的imhist函数绘制直方图,公式为:

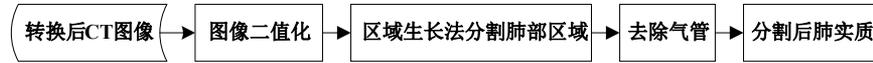


图3 肺实质分割流程图

Fig.3 Flow chart of pulmonary parenchyma segmentation

$$M = \text{imhist}(I, n) \quad (17)$$

其中,  $I$  为原图像,  $n$  为灰度级, 直方图中重叠区域最小的位置的值为最佳分割阈值, 选定阈值后, 把分

割出的肺部区域作为目标区域, 其他区域作为背景区域, 肺实质分割过程如图4所示。

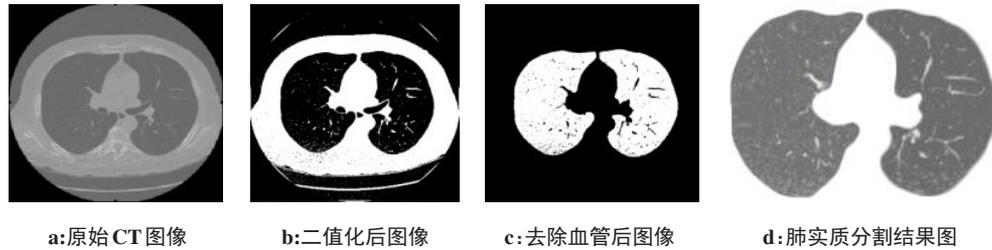


图4 肺实质分割过程示意图

Fig.4 Images in the process of pulmonary parenchyma segmentation

**2.2.2 区域生长法分割肺部区域** 二值化后的肺部CT图像, 没有把肺实质与图像背景精确分离开, 肺实质外的信息还混淆在图像区域内, 同时, 肺实质区域边缘还包括一些气管、空洞等噪声, 本文应用区域生长法定位肺实质边界并剔除噪声等干扰信息。

(1) 区域生长法初步分割肺实质。首先, 设定一个四邻域模板, 从肺边界最左侧出发, 比较生长点与待生长点的相似性, 找到满足条件的种子点; 其次, 设定种子点的生长准则; 最后, 确定种子点在生长过程中的终止条件。经过分割后, 肺实质周围大块的噪声被去除, 接下来去除肺实质周围的气管等干扰信息。

(2) 去除气管。在经过区域生长分割后, 肺实质周围还有一些血管噪声, 本文应用区域轮廓面积筛选法去除血管。先用边界跟踪算法记录肺实质完整轮廓, 接着根据计算得到的轮廓面积判断边缘是否有气管, 如果有应用数学形态学运算的腐蚀和膨胀操作去除噪声, 去噪后肺轮廓区域显示结果如图4c所示。把去除噪声后得到的蒙板图像与原图像进行掩膜运算, 得到最终分割的肺实质区域, 如图4d所示。

**2.3 肺结节感兴趣区域(Region Of Interest, ROI)的提取**

常用的获取肺结节ROI方法有阈值法、聚类法<sup>[17]</sup>等, 本文应用FCM算法提取ROI区域。首先, 应用多尺度高斯滤波器增强结节和血管的形状, 可比较清晰地区分血管端点及血管交叉区, 降低检测肺结节假阳率。对图像进行增强的算法步骤如下:

(1) 设滤波器的尺度个数  $N=5$ , 肺结节直径范

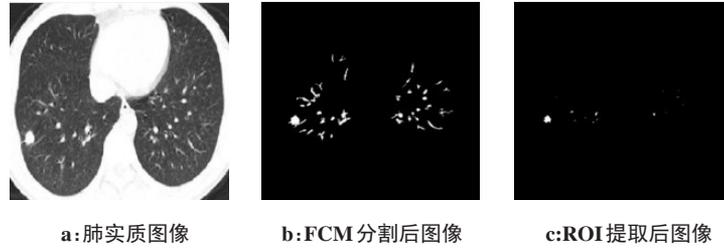
围  $d[d_0, d_1]$ ,  $d \in [4 \sim 30 \text{ mm}]$ , 尺度大小由  $\sigma$  表决定,  $\sigma \in [d_0/4, d_1/4]$ ,  $\sigma$  表的计算公式为:

$$\begin{cases} \sigma_1 = \frac{d_0}{4}, \sigma_2 = r\sigma_1, \dots, \sigma_N = r^{N-1}\sigma_1 \\ r = \left(\frac{d_1}{d_0}\right)^{\frac{1}{N-1}} \end{cases} \quad (18)$$

(2) 计算Hessian矩阵中  $\lambda_1, \lambda_2, \lambda_3$  的特征值向量, 并利用式(7)和式(8)计算结节 ( $E_{\text{ball}}$ ) 和血管 ( $E_{\text{line}}$ ) 增强函数; (3) 用增强函数得到的滤波器对图像的体素进行增强; (4) 每个体素值达到最大后, 循环到  $N=5$ , 结束算法; (5) 得到增强的结节和血管图像。

接下来应用聚类方法中的FCM算法提取肺结节的感兴趣区域。算法的步骤如下: (1) 设定聚类个数  $c=3$ , 常数  $p=2$ ; (2) 确定初始聚类中心值  $v$ ; (3) 应用式(11)计算隶属度函数  $u$ ; (4) 应用式(10)计算下一个聚类中心值  $v$ ; (5) 重复步骤(3)、(4)直到找到符合条件的  $v, u$  值, 完成模糊聚类的划分, 肺结节感兴趣区域提取过程如图5所示。

聚类后的分割结果里除了含有结节, 还包含气管及其他组织器官等噪声, 本文应用相关匹配法再次筛选疑似结节, 减少非结节区域, 获得包含结节的ROI。判断疑似结节区域的算法步骤如下: (1) 可检测最小结节直径为  $4 \text{ mm}$ , 所以设定阈值的范围为  $T \in [50, 150]$ ; (2) 定义分割候选区的面积为  $R$ , 对应的特征值为  $C_R$ , 若有  $C_R > T$ , 则  $R$  为目标区域, 否则为非结节区域; (3) 根据肺结节图像, 设符合高斯分布的模板为  $w(x, y)$ , 通过模板扫描候选区图像  $f(x, y)$ , 寻



a:肺实质图像 b:FCM分割后图像 c:ROI提取后图像

图5 肺结节感兴趣区域提取结果图

Fig.5 Extraction of regions of interest of pulmonary nodules

找到相似区域后,记录模板左对角线下方的位置,即为疑似结节的点的坐标;(4)重复步骤(2)、(3)扫描整个候选区域,标记疑似结节位置的像素点,去除未标记的位置,并提取疑似结节区域,结果如图5c所示。

### 2.4 基于SVM肺结节识别

肺结节分类识别分为3个步骤:特征选择与提取、训练分类器和测试分类器。即先把特征提取的数据生成训练样本集,然后对SVM模型进行分类训练及参数的优化,最后对测试样本数据进行分类识别。

**2.4.1 肺结节特征选择与提取** 在提取后的ROI区域里,依然含有血管、气管等假阳性结节,为了精确检测并识别出肺结节,需要提取肺结节特征,去掉假阳性结节。有关肺结节的特征有很多种,但提取过多特征会使算法复杂化,分类效果也不理想。根据肺结节与血管在形状及三维空间上的差异性,本文主要以形状特征为主对肺结节特征进行提取。为了更全面地分析肺结节形态结构、局部特性,在建立三维形状模型的基础上,提取如下(4)、(5)、(6)新的三维特征。提取的特征如下所示:

(1)圆形度  $K$ 。数值越接近1,区域形状越接近似圆形。 $P$ 表示周长, $A$ 表示面积,公式为:

$$K = \frac{P^2}{4\pi A} \tag{19}$$

(2)细长度。候选区域长轴与短轴的比值。值越大越狭长,与血管的形状越接近。

(3)紧凑度(comp)。表示所选区域内切圆轮廓与外切圆轮廓的比值,比值越大,越接近血管的形

状,反之,近似圆形。其中  $R_{in}$  为内切圆半径,  $R_{out}$  为外切圆半径,公式表示如下:

$$comp = \frac{R_{in}}{R_{out}} \tag{20}$$

(4)球心距  $E_2$ 。表示所选区域中的像素点与球心距离的标准差,  $E_2$  值越接近0,形状越接近球体,  $C_{cen}(A_i)$  表示区域的中心点,  $std$  表示计算标准偏差的函数,  $mean$  表示求数值的平均数。公式表示为:

$$E_2 = std \left( \frac{\|C(S_i) - C_{cen}(A_i)\|}{\text{mean}(\|C(S_i) - C_{cen}(A_i)\|)} \right) \tag{21}$$

(5)球形体积比  $E_1(A_i)$ 。表示所选区域与外接球体的体积  $VS(A_i)$  比,  $E_1(A_i)$  越小,所选区域越接近球体形状,公式表示为:

$$VS(A_i) = \frac{4}{3} \times \pi \times \left( \frac{dim}{2} \right)^3 \tag{22}$$

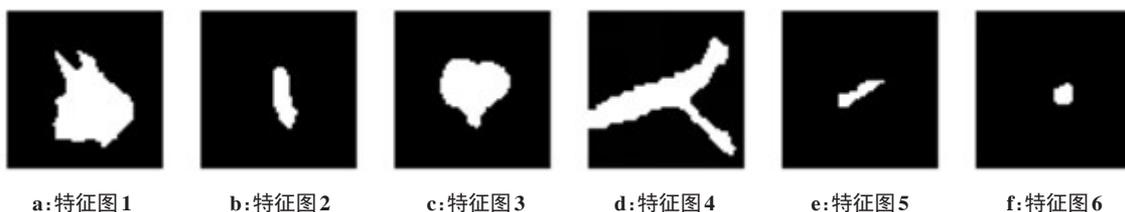
$$E_1(A_i) = \frac{A_i}{VS(A_i)} \tag{23}$$

(6)长方体交线平均距离  $E_2(A)$ :表示所选区域中心点  $C_{cen}(A_i)$  与外接长方体12条线中点  $dim(L_i)$  距离的平均值,距离分布越均匀,越接近球体,公式表示为:

$$E_2(A) = \sum_{i=1}^{12} \frac{C_{cen}(A_i) - C_{cen}(dim(L_i))}{C_{cen}(dim(L_i))} \tag{24}$$

本文提取肺结节图像的部分特征,特征图如图6所示。肺结节特征提取部分数据如表1所示。

**2.4.2 特征数据归一化处理** 提取样本数据的取值范围各不相同,给分类器训练数据带来了不便,为了使分类器在训练过程中不受数据的影响,根据表1数据值的特点,大部分数值范围在[0, 1]之间,本文应用线性函数归一化方法把特征数据取值按照相同的权



a:特征图1 b:特征图2 c:特征图3 d:特征图4 e:特征图5 f:特征图6

图6 肺结节特征图

Fig.6 Features of pulmonary nodules

表1 样本特征提取数据

Tab.1 Sample feature extraction data

特征	特征图1	特征图2	特征图3	特征图4	特征图5	特征图6
圆形成度	0.487	0.546	0.776	0.389	0.501	0.998
细长度	1.052	3.231	1.243	1.765	2.908	1.014
紧凑度	0.333	0.164	0.597	0.212	0.196	0.806
球心距	0.677	6.445	0.549	4.675	6.125	0.001
球形体积比	0.321	7.954	0.298	5.665	7.182	0.014
长方体交线距离	0.487	0.976	0.394	0.798	0.945	0.121

重映射到 $[0, 1]$ 之间。归一化方法定义如下:

$$x^* = \frac{x - x_{\min}}{x_{\max} - x_{\min}}, x \in D, x^* \in D^* \quad (25)$$

其中,  $x$  是归一化前数据集合  $D$  中任意元素,  $x^*$  是  $x$  归一化后数据集合  $D^*$  中元素,  $x_{\min}$ 、 $x_{\max}$  分别表示归一化前数据集合  $D$  中最小及最大元素值。归一化后, 特征数据的取值范围都在 $[0, 1]$ 之间, 大大提高了SVM分类器收敛速度及分类识别精度。

**2.4.3 SVM 参数优化** 在进行分类训练之前, 对式(14)和式(15)中  $(c, \gamma)$  参数进行优化, 可以得到比较理想的分类结果。本文应用网络搜索方法遍历所有网格点来寻找最优参数。设定两个参数的取值范围, 通过该范围内的数值创建合理间隔的网格模型, 并采用10折交叉验证法得到  $c = 2.0, \gamma = 8.0$  时, 分类准确率最好, 确定了参数的最优值。

**2.4.4 分类器训练与测试** 本文应用LIBSVM工具库对SVM分类器进行训练, 该工具库提供的分类器模型支持MATLAB、C语言等多种语言。把特征提取后的320个样本数据集放入MATLAB环境中进行试验, 选择其中190个作为训练样本数据, 在超平面上应用分类函数对数据集进行分类, 剩下的130个数据作为测试样本, 对分类器进行测试, 通过评价参数评判分类结果。

### 3 实验结果及分析

#### 3.1 数据来源

为了验证本方法的可行性和检测的有效性, 从肺部影像数据库联盟(Lung Image Database Consortium, LIDC)中随机抽取120例图像, 经专家标注有267个肺结节, 其中假阳性结节88个, 肺结节大小为4~30 mm, 图像的大小为512\*512, 扫描层大约有30层。

#### 3.2 实验环境

本实验在Windows10操作系统, CPU:i7处理器, RAM:8 G, 编程软件为MATLAB R2015a的环境中完成仿真实验。

#### 3.3 实验结果与分析

通过上述一系列算法过程, 肺结节被检测出来, 并在原图像中用红色矩形区域标注, 本文算法与文献[18-19]检测对比的3组结果如图7所示。

从检测结果可以看出, 应用FCM聚类算法及图像增强技术能准确全面地提取肺结节ROI区域, 再根据建立的三维形状模型, 提取肺结节的三维特征, 降低肺结节检测的假阳性率, 检测的准确率高于其它两种方法, 有效地排除了交叉血管、条形血管等干扰。定义准确率、误判率两个参数评价检测结节效果, 公式表示如下所示:

$$\begin{aligned} \text{准确率} &= \text{检出正确的结节数} / \text{实际总结节数} \\ \text{误判率} &= \text{检出动错误的结节数} / \text{检出结节总数} \end{aligned} \quad (26)$$

使用本文方法与文献[18-19]方法分别对随机抽取的120副图像中267个肺结节进行检测, 具体检测结果如表2所示。

由表2数据对比可知, 本文提出的方法在准确率上高于其它两种方法, 而且通过增强肺结节及血管图像效果和提取三维肺结节特征, 误判率也低于其它两种方法, 实验结果证明, 本文的方法是有效可行的。

为了评价SVM分类器识别性能, 本文使用敏感度、特异度和准确率评价分类器, 定义如下:

$$\begin{aligned} \text{敏感性} &= \text{真阳性} / (\text{真阳性} + \text{假阳性}) \\ \text{特异性} &= \text{真阴性} / (\text{真阴性} + \text{假阳性}) \end{aligned} \quad (27)$$

$$\text{准确率} = (\text{真阳性} + \text{真阴性}) / (\text{真} / \text{假阳性} + \text{真} / \text{假阴性})$$

通过计算, 应用SVM分类器对130个样本数据做测试, 并与Cascio等<sup>[20]</sup>BP神经网络算法进行对比, 得出评价指标的计算结果如表3所示。

由表3数据可知, 应用SVM模型分离肺结节的特异度和准确率都略高于BP神经网络方法, 敏感度略低的原因是SVM中提取特征少于BP神经网络, 但SVM检测时间及收敛速度都高于BP神经网络, 同时, SVM参数优化及采用交叉验证法保证了分类的准确性。

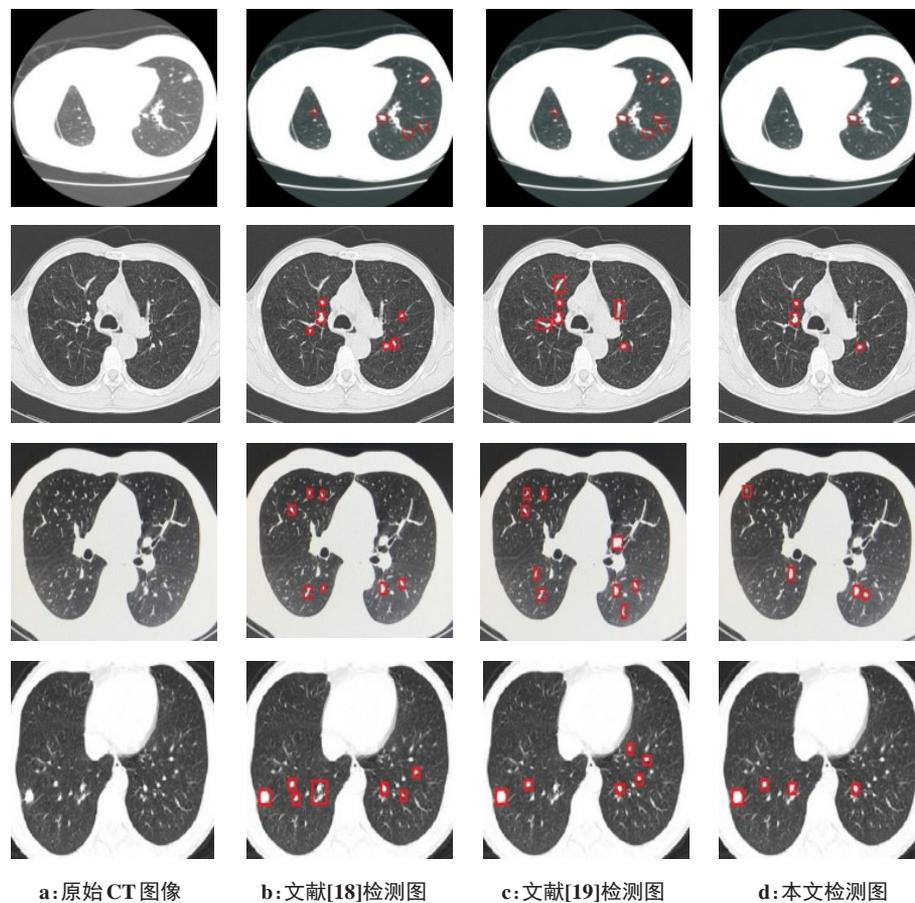


图7 肺结节检测结果图

Fig.7 Flow chart of pulmonary nodules detection

第1~4行图像分别为4个不同患者肺部影像图

表2 本文方法与现有两种方法性能对比

Tab.2 Performance comparison among the proposed method and two existing methods

方法	结节总数	检出结节数	正确结节数	错误结节数	准确率/%	误判率/%
文献[18]	267	456	206	250	88.7	54.8
文献[19]	267	379	228	151	85.6	39.8
本文方法	267	299	299	53	92.3	17.7

表3 评价数据结果对比

Tab.3 Comparative analysis of evaluation indexes

方法	敏感度	特异度	准确率
SVM	0.827	0.929	0.956
BP神经网络	0.967	0.881	0.923

## 4 结论

本文针对二维空间肺结节检测困难及准确率低等问题,提出一种增强滤波器和三维特征肺结节检测和识别方法。经过仿真实验及数据比较,该算法能高效、准确地检测并识别出肺结节。但算法也存检测敏感性

不够高的问题,主要原因是本文只针对肺结节的形状特征进行提取,但没有对纹理等特征进行提取,针对一些对比度低的毛玻璃型或者无规则形状的毛刺型结节的检测并不理想,因此,对肺结节纹理等特征的提取及研究是后续一个发展研究方向。

## 【参考文献】

[1] KUHNIGK J M, DICKEN V, BORNEMANN L, et al. Morphological segmentation and partial volume analysis for volume try of solid pulmonary lesions in thoracic CT scans[J]. IEEE Trans Med Imaging, 2016, 25(4): 417-420.  
 [2] 陈树越,晁亚,邹凌.基于几何特征的孤立性肺结节检测[J].生物医学工程学杂志,2016,33(4): 680-685.  
 CHEN S Y, CHAO Y, ZOU L. Detection of solitary pulmonary

- nodules based on geometric features [J]. *Journal of Biomedical Engineering*, 2016, 33(4): 680-685.
- [3] ALILOU M, KOVALEV V, SNEZHKO E, et al. A comprehensive framework for automatic detection of pulmonary nodules in lung CT image [J]. *Image Anal Stereol*, 2014, 33(1): 13-27.
- [4] DEMIR Ö, YILMAZ ÇAMURCU A. Computer-aided detection of lung nodules using outer surface features [J]. *Biomed Mater Eng*, 2015, 26(Suppl 1): S1213-S1222.
- [5] STEMBER J N. The normal mode analysis shape detection method for automated shape determination of lung nodules [J]. *J Digit Imaging*, 2015, 28(2): 224-230.
- [6] 高婷, 龚敬, 王远军, 等. 检测肺结节的3维自适应模板匹配[J]. *中国图象图形学报*, 2014, 19(9): 1384-1391.  
GAO T, GONG J, WANG Y J, et al. Three dimensional adaptive template matching algorithm for lung nodule detection [J]. *Journal of Image and Graphics*, 2014, 19(9): 1384-1391.
- [7] DE CARVALHO FILHO A O, DE SAMPAIO W B, SILVA A C, et al. Automatic detection of solitary lung nodules using quality threshold clustering, genetic algorithm and diversity index [J]. *Artif Intell Med*, 2014, 60(3): 165-177.
- [8] IQBAL S, IQBAL K, ARIF F, et al. Potential lung nodules identification for characterization by variable multistep threshold and shape indices from CT images [J]. *Comput Math Methods Med*, 2014: 241647.
- [9] 王秋萍, 冯筠, 强永乾, 等. 空洞检测算法自动化检测肺结节内空泡和空洞的可行性[J]. *中国医学影像技术*, 2014, 30(9): 1309-1313.  
WANG Q P, FENG J, QIANG Y Q, et al. Feasibility of automatic detection of cavity and vacuole in pulmonary nodules using cavity detection algorithm [J]. *Chinese Journal of Medical Imaging Technology*, 2014, 30(9): 1309-1313.
- [10] KRISHNAMURTHY S, NARASIMHAN G, RENGASAMY U. Three dimensional lung nodule segmentation and shape variance analysis to detect lung cancer with reduced false positives [J]. *J Proc Inst Mech Eng H*, 2016, 230(1): 58-70.
- [11] KANAZAWA K, KAWATA Y, NIKI N, et al. Computer-aided diagnosis for pulmonary nodules based on helical CT images [J]. *Comput Med Imaging Graph*, 1998, 22(2): 157-167.
- [12] KUBO M, KUBOTA K, YAMADA N, et al. CAD system for lung cancer based on low-dose single-slice CT image [C]//*Proceedings of the Medical Imaging*. San Diego, USA: SPIE, 2012: 1262-1269.
- [13] YAMADA N, KUBO M, KAWATA Y, et al. ROI extraction of chest CT images using adaptive opening filter [C]//*Proceedings of the Medical Imaging*. San Diego, USA: SPIE, 2013: 869-876.
- [14] ANTONELLI M, FROSINI G, LAZZERINI B, et al. A CAD system for lung nodule detection based on an anatomical model and a fuzzy neural network [C]//*Proceedings of the 2006 Annual Meeting of the North American Fuzzy Information Processing Society*. Montreal, Canada: IEEE, 2016: 448-453.
- [15] 崔伟东, 周志华, 李星. 支持向量机研究 [J]. *计算机工程与应用*, 2011, 27(1): 55-60.  
CUI W D, ZHOU Z H, LI X. Research of support vector machines [J]. *Computer Engineering and Applications*, 2011, 27(1): 55-60.
- [16] 范九伦, 赵凤. 灰度图像的二维OSTU曲线阈值分割法 [J]. *电子学报*, 2017, 35(4): 751-755.  
FAN J L, ZHAO F. Two-dimensional OTSU's curve thresholding segmentation method for gray-level images [J]. *Acta Electronica Sinica*, 2017, 35(4): 751-755.
- [17] 米红, 张希雯, 陈文生. 基于数据挖掘平台的模糊聚类算法及应用研究 [J]. *厦门理工学院学报*, 2016, 14(2): 17-21.  
MI H, ZHANG X W, CHEN W S. Application of Fuzzy clustering algorithm on data mining platform [J]. *Journal of Xiamen University of Technology*, 2016, 14(2): 17-21.
- [18] 钟明霞, 姜柏军. 基于BP神经网络的肺结节自动检测算法研究 [J]. *计算机时代*, 2014(10): 65-67.  
ZHONG M X, JIANG B J. Research of lung nodules detection algorithm based on BPNN [J]. *Computer Era*, 2014(10): 65-67.
- [19] 王立非, 李彬, 田联房, 等. 基于规则及多特征跟踪的肺结节的智能检测方法 [J]. *生物医学工程研究*, 2016, 29(2): 79-83.  
WANG L F, LI B, TIAN L F, et al. Intelligent detection of lung nodules based on rules and multi-feature tracking [J]. *Journal of Biomedical Engineering Research*, 2016, 29(2): 79-83.
- [20] CASCIO D, MAGRO R, FAUCI F, et al. Automatic detection of lung nodules in CT data sets based on stable 3D mass-spring models [J]. *Comput Biol Med*, 2012, 42(11): 1098-1100.

(编辑:薛泽玲)