

申请上海交通大学硕士学位论文

**骨扫描图像异常检测和分析方法研究**

论文作者 \_\_\_\_\_ 耿世杰 \_\_\_\_\_

学 号 \_\_\_\_\_ 1130329006 \_\_\_\_\_

导 师 \_\_\_\_\_ 乔宇 \_\_\_\_\_

专 业 \_\_\_\_\_ 控制科学与工程 \_\_\_\_\_

答辩日期 \_\_\_\_\_ 2016 年 6 月 \_\_\_\_\_



Submitted in total fulfillment of the requirements for the degree of Master  
in Control Science and Engineering

# Research on Abnormality Detection and Analysis Methods for Bone Scintigraphy

Shijie Geng

Advisor

Yu Qiao

DEPART OF AUTOMATION, SCHOOL OF ELECTRONIC INFORMATION AND ELECTRICAL  
ENGINEERING  
SHANGHAI JIAO TONG UNIVERSITY  
SHANGHAI, P.R.CHINA

June , 2016



## 上海交通大学 学位论文原创性声明

本人郑重声明：所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的作品成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到本声明的法律结果由本人承担。

学位论文作者签名： 耿世杰

日期： 2016 年 5 月 29 日



# 上海交通大学

## 学位论文版权使用授权书

本学位论文作者完全了解学校有关保留、使用学位论文的规定，同意学校保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅。本人授权上海交通大学可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。

本学位论文属于

保 密 ，在 \_\_\_\_\_ 年解密后适用本授权书。

不保密

(请在以上方框内打√)

学位论文作者签名: 耿世杰

指导教师签名: 王立波

日期: 2016 年 5 月 29 日

日期: 2016 年 5 月 30 日



# 骨扫描图像异常检测和分析方法研究

## 摘要

骨扫描是一种全身性医学影像检查，它可以比其他方法更早的发现肿瘤及肿瘤骨转移，因此获得了广泛的应用。临床医生对骨扫描图像的诊断工作主要包括热点的检测、识别和分析。实际中由于图像成像质量和读片医师的主观差异，导致不同医生对同一病人的诊断结果存在不一致，从而影响到病人的后续治疗。骨扫描图像计算机辅助分析系统以图像处理和机器学习算法为核心，对骨扫描图像进行分析，检测出图像中的热点，将热点精准分割出来并进行量化分析。这可以很大程度上克服读片医师的主观差异，降低人工处理的工作量，提升诊断的灵敏度和准确率。同时，由于骨扫描图像数量的不断增长，医生需要在最快的时间内通过图像检索系统从过往病例中找到类似的部位、类似的热点类型，为医生的诊断提供必要的参考。本文搭建了骨扫描图像计算机辅助诊断系统的整体框架，并从三个方面解决了医生的需求。

本文根据骨扫描图像低信噪比、图像边界模糊等特点，创新地尝试使用卷积稀疏编码提取深度特征描述骨扫描图像的细节。并结合 AdaBoost 分类器，从整体上判断胸廓骨扫描图像中是否含有热点，并实现三个级别的分类。本文进而采用多实例学习算法构建骨扫描图像热点概率图，结合局部符号差异水平集，实现对热点的初步自动分割。本文通过相关实验证明了所提出的诊断算法和分割算法的有效性。

由于自动分割不可能完美地分割出所有热点，所以本文创新性地提出了一种热点交互式分割算法，使得医生仅仅提供一个点的信息，便可完成对热点的分割。本文使用变分推断中的平均场方法不断迭代向邻域传播输入信息，并采用水平集实现热点分割。同时，本文利用图割的方法，使得医生可以对热点分割的结果进行局部修饰。实验表明，本文提出的交互式分割方法，可以简单、快速和精准地分割出骨扫描图像中的热点。

另外，为了方便医生查找以往相似的病例，本文构建了全新的基于内容的骨扫描图像检索系统。由于传统的基于 SIFT 和词袋的图像特征对于骨扫描图像描述效果一般，本文改用深度特征代替。为了同时保证检索的准确性和快速性，本文采用了基于核函数的有监督哈希建立索引。实验结果表明，本文所建立的骨扫描图像检索系统可以帮助医生高效而准确找到数据库中的相似病例。

**关键词：**骨扫描 热点检测 交互式分割 计算机辅助诊断 医学图像检索

# **Research on Abnormality Detection and Analysis Methods for Bone Scintigraphy**

## **ABSTRACT**

Bone scintigraphy is a whole-body medical imaging examination. It has been widely used since it is able to detect tumor and bone metastasis earlier than other methods. The diagnostic work of physicians on bone scintigraphy comprises of detection, recognition and analysis of hotspots. In fact, due to the poor image quality and physicians' experience, different physicians may not reach agreement on the diagnosis of the same patient. This will affect the subsequent treatment of the patient. Computer-aided diagnosis (CAD) system for bone scintigraphy is based on the algorithms of image processing and machine learning. It is able to analyze the bone scan images, detect the hotspots in it, segment out the hotspots accurately and conduct quantitative analysis. The CAD system can not only decrease differences between different physicians and manual workload, but also improve the diagnosis accuracy and sensitivity. At the same time, with the increasing volume of bone scan images, physicians need a image retrieval system to find out corresponding part and similar hotspots from case database and to offer reference for them.

In this paper, in view of low signal noise ratio and weak boundary of bone scan images, convolution sparse coding is applied to extract deep features to represent details of bone scan images. The author trained a classifier with AdaBoost, and used the classifier to diagnose if there are hotspots in a certain bone scintigraphy. Similar with physicians, a 3-point interpretation scale is used to score bone scan images. The probability map of hotspots is then constructed by

multiple instance learning algorithms. Finally, local signed difference level set is performed to segment hotspots from bone scintigraphy automatically. The experimental results demonstrate that the proposed diagnosis and segmentation methods are effective.

Since automatic segmentation is not perfect, the author raised up an innovative interactive segmentation algorithm for hotspot segmentation. Physicians can accurately segment a hotspot with merely offering one point. In this paper, mean field inference is used to propagate information from input points to their neighborhood. At the same time, segmentation is performed with level set methods. The author also proposed an approach utilizing Graph Cut that allows physicians to improve segmentation result. Experiment shows that the proposed interactive approach can segment hotspots from bone scan images easily, fast and accurately.

Besides, in this paper, a new content-based bone scan image retrieval system is built for physicians to search similar cases. The author has showed that traditional SIFT feature and Bag of Words model is not capable to represent bone scintigraphy well. In the paper, the author substitutes it with deep features. To keep the effectiveness and accuracy, the author uses kernel-based supervised hashing to construct index for bone scintigraphy database. As shown in the experiments, with the help of the retrieval system, physicians can find similar cases from the database effectively.

**KEY WORDS:** Bone Scan, Hotspot Detection, Interactive Segmentation, Computer-aided Diagnosis(CAD), Medical Image Retrieval

# 目 录

摘要	i
<b>ABSTRACT</b>	iii
目录	v
<b>第一章 绪论</b>	1
1.1 研究背景 . . . . .	1
1.2 国内外研究现状 . . . . .	2
1.2.1 骨扫描辅助诊断 . . . . .	2
1.2.2 医学图像检索 . . . . .	5
1.3 本文组织结构与创新点 . . . . .	6
1.4 本章小结 . . . . .	7
<b>第二章 基于卷积稀疏编码和多示例学习的热点检测</b>	9
2.1 问题概述与框架介绍 . . . . .	9
2.2 基于卷积稀疏编码的图像诊断 . . . . .	10
2.2.1 卷积稀疏编码 . . . . .	10
2.2.2 AdaBoost 分类器 . . . . .	13
2.2.3 图像级别的诊断 . . . . .	14
2.3 基于多示例学习的区域诊断 . . . . .	17
2.3.1 多示例学习 . . . . .	17
2.3.2 骨扫描图像特征提取 . . . . .	20

2.3.3 区域级别的诊断 . . . . .	22
2.4 局部符号差异水平集分割 . . . . .	25
2.4.1 水平集方法 . . . . .	25
2.4.2 局部符号差异水平集 . . . . .	26
2.4.3 热点自动分割 . . . . .	28
2.5 实验结果 . . . . .	30
2.6 本章小结 . . . . .	33
<b>第三章 基于平均场和水平集的热点交互式分割</b>	<b>35</b>
3.1 问题概述 . . . . .	35
3.2 平均场推断 . . . . .	36
3.3 基于平均场和水平集的交互式分割 . . . . .	38
3.4 热点轮廓局部修饰 . . . . .	42
3.4.1 图割方法简介 . . . . .	43
3.4.2 局部修饰的实现 . . . . .	45
3.5 实验结果 . . . . .	48
3.6 本章小结 . . . . .	51
<b>第四章 基于深度特征的有监督哈希骨扫描图像检索</b>	<b>53</b>
4.1 哈希索引技术 . . . . .	53
4.2 核函数有监督哈希算法 . . . . .	55
4.3 骨扫描图像检索系统 . . . . .	57
4.3.1 SIFT 和词袋特征提取 . . . . .	57
4.3.2 深度特征提取和骨扫描图像检索的实现 . . . . .	58
4.4 实验结果 . . . . .	60
4.5 本章小结 . . . . .	63

---

<b>第五章 总结与展望</b>	<b>65</b>
5.1 主要工作 . . . . .	65
5.2 后续研究 . . . . .	66
<b>参考文献</b>	<b>69</b>
<b>致 谢</b>	<b>77</b>
<b>攻读学位期间发表的学术论文</b>	<b>79</b>



# 第一章 绪论

## 1.1 研究背景

随着医学成像技术的发展，医学影像已经成为临幊上诊断疾病的主要手段。核磁共振（MRI）、计算机断层扫描（CT）、超声成像、单光子发射计算机断层显像术（SPECT）等成像技术，都被广泛应用于临幊疾病的分析和诊断<sup>[1]</sup>。与此同时，以医学图像处理和机器学习为核心的计算机辅助诊断技术（CAD）正发挥着越来越重要的作用。利用计算机技术来进行病灶的检测，可以大大地减轻人工处理的工作量、克服由于读片人员经验而产生的主观差异，同时提高诊断的灵敏度、准确率以及速度，为医生做出正确的决策提供帮助。计算机辅助诊断系统研究的课题主要包括检测医学影像中的病灶、帮助医生方便地分割出病灶以及对病灶进行后续的量化分析。同时，医学图像数量的不断增长催生了对图像检索系统的需求，医生需要在最快的时间从过往病例找到类似的部位、类似的病灶类型，从而为医生的诊断提供必要的参考和帮助。

骨扫描成像是一种全身性的核医学影像检查，它可以比其他方法更快且敏感地发现肿瘤骨转移，因此是诊断肿瘤和相关骨骼疾病的重要临幊手段。典型的骨扫描图像如图1-1所示。检查前病人先要通过静脉注射放射性药物，然后在数个小时后检测骨骼对放射性药物的吸收是否存在异常。在医学术语中，肿瘤及肿瘤骨转移被称为“热点”，并表现出比周围区域更亮的特性。前列腺癌和乳腺癌的晚期患者一般会发生肿瘤骨转移，而肿瘤骨转移的程度直接影响到病人的存活率以及后续治疗方案的选择<sup>[2]</sup>。因此，对骨扫描成像的结果做出正确的诊断至关重要。然而，由于骨扫描成像质量、病人之间的差异、医生本身的经验等原因，不同的医生对同一病人的诊断结果平均一致率仅为64%<sup>[2]</sup>。同时，医生的诊断需要对骨扫描图像进行量化分析，尤其是对热点数量、大小、亮度的测量。若仅靠人工完成，所需工作量非常大，这可能导致医生的诊断速度下降，甚至因

为疲倦而做出不够准确的判断。

建立骨扫描图像计算机辅助诊断系统可以解决以上的问题。它能够完成热点的检测、分类、分割和定量分析，为医生的诊断提供准确的参考信息。但是，要建立符合临床应用要求的骨扫描计算机辅助诊断系统却不容易，热点检测一直是该领域的一大难点。本文根据骨扫描图像的特点，搭建了骨扫描图像计算机辅助诊断系统的整体框架。分别实现了整图级别的诊断、热点的初步自动检测及分割，提出了热点交互式精准分割算法，构建了基于内容的骨扫描图像检索系统。有效的提高了热点识别和分割的准确率，为医生提供了热点定量分析的工具，帮助医生做出正确的诊断。

本文中使用的骨扫描图像来自于上海交通大学仁济医院核医学科。他们对于每一位病人提前注释 Tc-99m 显影液，并于 3 至 4 小时之后采集前身和后身两张骨扫描图像。每幅图像均为 16 位 Dicom 格式，大小为  $1024 \times 512$ ，其中同时包含病人的 ID、姓名、身高、性别等信息。我们请核医学科的医生们手动对采集的图像进行标注，并将医生的标注作为金标准。最后一共标注了 780 例正常病人（无热点）和 583 例异常病人（存在热点）。由于热点主要集中在胸廓区域，同时胸廓区域的热点情况对于肿瘤骨转移的诊断最为关键。因此本文将研究重点放在胸廓骨扫描图像上，并通过特征点分区算法完成胸廓区域的提取。

## 1.2 国内外研究现状

### 1.2.1 骨扫描辅助诊断

目前，国内外的一些团队致力于骨扫描计算机辅助诊断的研究。下面针对其中有代表性的一部分做简单的介绍。

Tang-Kai Yin 等<sup>[3]</sup>提出了一种基于特征点的模糊推断系统进行骨扫描图像的热点检测和识别。它利用局部最大值分割出潜在的热点。并模仿医生的诊断方法，对这些候选点提取分别提取亮度和非对称性两个特征，其中“亮度”定义为候选区域与其邻域的灰度差异，“对称性”定义为候选区域与人体中心线对称区域的灰度差异。此方法将骨扫

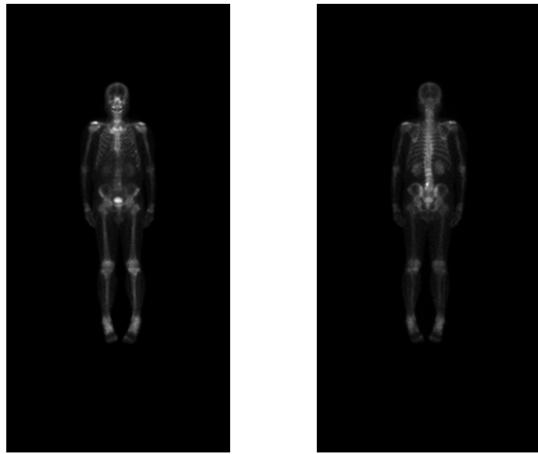


图 1-1 典型的骨扫描图像

Fig. 1-1 Classical Bone Scan Image.

描图像分成六个区域，对每个区域，都以提取的两个特征作为输入，训练了相应的模糊推理规则。此方法可以达到 91.5% 的灵敏度，但是却存在假阳性率偏高的问题，会给医生带来额外的工作量。

Jia-Yann Huang 等<sup>[4]</sup> 在 2007 年开发了一套骨扫描辅助诊断系统，该系统采用人体解剖学的作为先验知识，通过寻找人体特征点，将骨扫描图像分割成 23 个人体区域。对于每个分割出来的区域，该系统通过此区域的灰度均值和标准差建立线性回归分析模型，再计算得到区域阈值用于热点的检测和分割。该算法的热点检测敏感率达到 92.1%，但如果前期分割的精准度不高，就会导致阈值选择的偏差，从而影响到系统的效果。

瑞典的哥德堡大学研究团队对骨扫描图像计算机辅助诊断系统的研究最为深入和成熟。May Sadik 团队在 2006 年到 2010 年期间，提出了一套完整的骨扫描计算机辅助诊断系统，该系统实现了图像分割、热点检测、特征提取和病灶识别等功能。May Sadik 等<sup>[2]</sup> 于 2006 年提出的骨扫描辅助诊断系统利用人体解剖先验知识完成图像分区，根据分区结果选择区域阈值分割热点。该系统对分割出的热点抽取 14 维的特征向量，利用人工神经网络（ANN）训练了图像级别的分类器，此算法的敏感性为 90%，特异性为

74%。由于该系统特异性较低，因此 May Sadik 团队于 2008 年对原系统进行了改进<sup>[5]</sup>。改进主要集中在 3 个方面：一是采用主动形状模型（ASM）进行图像分区，使分割的结果更为准确；二是阈值大小通过特定区域的均值和方差信息自适应获得；三是利用人工神经网络训练全新的热点分类器，其中对每个热点提取 45 个特征，而原有针对病人的分类器提取的特征提高到了 26 维。改进后的系统特异性达到了 89%，取得了不错的效果。

在 May Sadik 所做工作的基础上，Mattias Ohlsson 等<sup>[6]</sup> 提出了骨扫描全自动诊断系统。它首先建立人体标准骨骼模板，接着对于每个病人采用 Morphon 算法进行配准，将图像分割成 12 个区域。该系统不断迭代，完成区域灰度归一化和热点分割，并将收敛后的热点分割结果作为诊断系统的输入。Mattias Ohlsson 同样训练了针对热点和病人的两个神经网络分类器，使系统的敏感性和特异性达到 95% 和 64%。该系统虽提高了热点检测的敏感性，特异性却有严重的问题。

Reza Kaboteh<sup>[7]</sup> 于 2012 年对原有骨扫描辅助诊断系统进行了新一轮的改进。该系统通过计算骨扫描指数（Bone Scan Index, BSI），为医生的诊断提供参考。骨扫描指数是指热点面积与区域面积之比，骨扫描指数值越高，病人的存活率也就越低，医生可以根据该指数判断病人的病情进展。该系统采用改进的 Morphon 算法完成图像配准，并结合带通滤波器和阈值进行热点分割。该系统对每个热点提取骨扫描指数以及几何、位置和灰度直方图等特征，并训练了 12 个神经网络分类器，分别对应每个图像区域。该系统的敏感性和特异性都很高，分别达到了 95% 和 87%。根据该系统已经开发了第一款商用骨扫描诊断软件 EXINI-bone。

在国内，上海交通大学图像处理与模式识别研究所首先致力于骨扫描辅助诊断系统的研究，并在脊柱分割和脊柱热点检测上取得了一定的进展<sup>[8,9]</sup>。在脊柱分割方面，该团队提出的算法<sup>[9]</sup> 利用解剖学知识和统计学方法定位特征点完成人体的分区。对于胸廓区域，利用 OTSU 阈值和形态学方法实现拓扑学操作和宽度限制，从而分割出脊柱区域。对于热点检测，该团队先对脊柱图像进行灰度标准化，然后对灰度概率分布进行最

优高斯拟合确定区域阈值<sup>[9]</sup>。对于检测出的热点，该算法通过自适应生长算法<sup>[10]</sup>勾勒出热点的边界。该团队提出的热点检测算法获得了更好的敏感性和特异性。但目前研究工作只局限于在脊柱的热点检测，对于其他骨骼组织的热点诊断还有待研究。

### 1.2.2 医学图像检索

医学图像数量的增长，催生了对基于内容的医学图像检索的需求。利用基于内容的图像检索（CBIR），医生可以从医学图像数据库中找到与查询图像内容上最相关和一致的图像，从而对医生的诊断起到帮助和参考。最近几年，针对不同成像设备、不同疾病的医学图像检索系统都被广泛地研究。下面列举一些比较典型的医学图像检索系统。

Simonyan 等<sup>[11]</sup>于 2011 年提出了一种交互式的 X-射线影像搜索引擎。不同于其它检索系统搜索全局上相似的图像，这种算法允许用户选择感兴趣区域（ROI）进行检索。作为预处理的步骤，这种方法使用薄板样条对全图进行配准。并根据感兴趣区域的内容进行检索和排序，比全局检索方法更加精准和高效。Zhang 等<sup>[12-14]</sup>利用哈希方法实现了组织病理学图像的大规模检索系统。这种方法可以将 10000 多维的图像特征向量压缩成几十位的哈希码。使得即使对于较大的图像数据库，也可以实现实时的检索。该方法可以达到 87.3% 的准确率，且对于 50 万级别的数据库，查找时间仅需 1.68 秒。Jiang 等<sup>[15,16]</sup>提出了一套基于词汇树的乳房肿瘤 X 射线图像检索系统。该系统具有良好的可扩展性，图像数量的增长不会对检索准确率产生影响，同时检索的效率也很高。

另外，学者们也提出了许多表达医学图像特征的方法，提升了对医学图像的描述能力和检索准确率。Pauly 等<sup>[17]</sup>通过多重随机分配法完成对特征空间的全局量化，建立视觉词汇字典，实现对医学图像成像设备的识别。由于采用了随机方法，该方法比传统的 K-Means 聚类要快很多，并且在准确率上也有提升。传统的词袋方法（Bag of Visual Words）使用高斯差分（DoG）或者密集网格来提取兴趣点。Haas 等<sup>[18]</sup>通过超像素方法产生图像的超分割，并将超像素区域的中心点作为兴趣点。实验证明，利用这种方法提取兴趣点，生成的图像特征区分能力更强，建立的医学图像检索系统检索准确率也更高。传统的医学图像检索系统一般采用低级视觉特征描述图像的细节。但低级视觉特征

和高级语义特征之间的“语义鸿沟”会影响检索的准确率。Kurtz 等<sup>[19,20]</sup>定义了全新的相似性度量方法——层级语义距离 (HSBD)，该方法将图像特征和本体语义特征相结合来检索相似的图像。由于该方法有效利用了图像的标注信息，使得检索的准确率显著提升。

### 1.3 本文组织结构与创新点

图1-2给出了本文研究内容的整体框架，本文的组织结构如下：

第一章首先介绍了本文的研究背景与国内外相关研究的现状，并给出了全文的大体框架与创新点。

第二章提出了基于稀疏编码的图像级别诊断、基于多示例学习的区域级别诊断以及一套热点自动分割的框架。

第三章介绍了平均场推断方法，并提出了一套基于平均场推断的热点交互式分割算法和后续的局部修饰方法。

第四章实现了一套基于有监督哈希的骨扫描图像检索系统，并对不同的特征提取方法进行了比较。

第五章总结了本文的研究工作，并指出了未来有价值的研究路线和方向。

针对骨扫描图像热点检测、热点分割和图像检索等任务，本文的主要创新点如下：

1. 针对骨扫描图像，本文创新地尝试使用卷积稀疏编码提取图像的特征，并通过 AdaBoost 训练图像级别的分类器，实现了对胸廓骨扫描图像的三级诊断。
2. 本文对  $4 \times 4$  大小的骨扫描图像区域分别提取直方图、邻域对比度、对称对比度以及局部二进制模式特征，并创新性地通过多示例学习算法训练了区域级别的热点分类器。本文进而利用训练好的分类器构建热点概率图，结合水平集分割方法，实现对热点的自动分割。
3. 为方便医生对热点定量信息的获取，本文首次提出了一种基于平均场推断和水平集的热点交互式分割算法。使得医生可以通过最少的输入，快速而准确地完成对热点

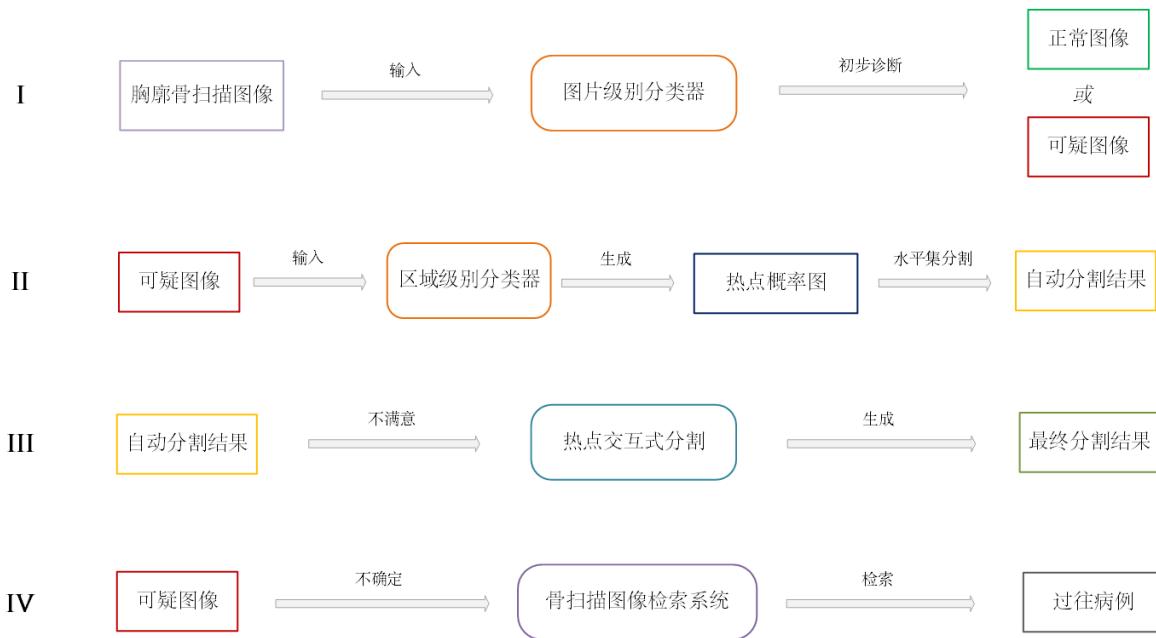


图 1-2 本文整体框架

Fig. 1-2 The overall framework of the thesis.

的分割。

4. 为了给医生的诊断提供帮助，本文首次构建了一套基于内容的骨扫描图像检索系统。本文采用深度特征描述图像的细节，并通过哈希算法完成图像的检索。

## 1.4 本章小结

本章首先介绍了本文研究的原因和背景，接着介绍了该领域国内外相关研究的已有成果，最后介绍了本文研究的整体框架并列举了本文的创新成果，具体的研究成果将在接下来三章进行详述。



## 第二章 基于卷积稀疏编码和多示例学习的热点检测

传统的骨扫描辅助诊断系统一般采用区域生长法<sup>[10]</sup> 或自适应阈值法<sup>[21]</sup> 完成热点的检测和分割。然而，由于骨扫描图像的低信噪比以及模糊边界，以上两种方法都不能取得满意的分割结果。本章提出了一种骨扫描图像热点的自动检测方法，和以往的方法不同，它尝试去理解骨扫描图像的细节，并通过机器学习算法，训练出图像级别和区域级别的分类器。利用这些分类器，可以生成热点的概率图，并通过水平集完成热点的精确分割。

### 2.1 问题概述与框架介绍

本章将实现热点的自动精确分割。由于水平集方法在拓扑变化和像素精确度上的优势，因此本章采用局部符号水平集来完成热点的分割。为了保证水平集分割的精度，设置的初始轮廓要尽量的接近要分割的目标，而热点的概率图可以对水平集的初始化起到很大的帮助。本章中，我们将结合卷积稀疏编码、AdaBoost 算法和多实例学习来完成热点概率图的生成。

本章的整体框架如图2-1所示。首先利用卷积稀疏编码对输入图像抽取深度特征，并通过 AdaBoost 算法训练出图像级别的分类器。这是为了提前对图像进行判断，如果图像被诊断为正常图像，就不用再进行后续的分割步骤。为了找出可疑图像中的热点区域，本文通过多实例学习算法训练了区域级别的热点分类器，并利用它对可疑图像进行扫描，从而生成输入图像的热点概率图。最后，将得到的热点概率图进行二值化，作为水平集的初始轮廓，并通过水平集的不断进化将热点自动分割出来。

接下来，将分别详细介绍提出的热点自动精确分割算法的每个部分，并通过具体的例子和相关的实验证明所提出方法的有效性和准确性。

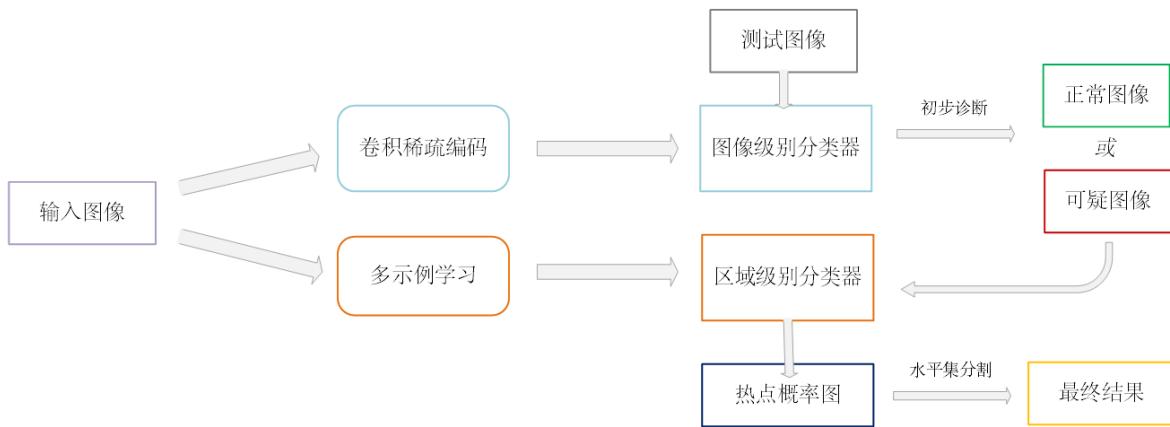


图 2-1 热点自动分割整体框架

Fig. 2-1 The overall framework of automatic hotspot segmentation.

## 2.2 基于卷积稀疏编码的图像诊断

### 2.2.1 卷积稀疏编码

良好的特征表达对机器学习的效果起着十分关键的作用。传统的机器学习算法需要靠人工来提取特征，虽然有很多区分性和不变性强的特征被提出，但这些特征却并不是对所有图像都有效。对每一类图像设计和提取不一样的特征虽然可行，却也耗费大量的人力和时间。随着稀疏编码和卷积神经网络的发展，深度学习技术已经在计算机视觉领域得到广泛的应用。与传统的机器学习方法不同，深度学习可以通过无监督的方式自动地学习图像的特征，它通过模拟人脑的视觉机理，对图像的特征进行逐层抽象。从最底层的像素，到更高层的边缘、目标，实现对特征的表达。

无监督特征学习需要大量的图像进行训练，互联网上大量的图片数据恰恰可以满足这一要求。大量的训练需要非常大的计算量并耗费很多的时间，并行计算和图像处理器（GPU）加速技术的发展解决了这一问题。这些因素都直接促进了深度学习的发展。在过去的几年，一系列深度学习的方法和架构被提出。Raina 等<sup>[22]</sup>于 2007 年提出了自我学习框架（Self-taught learning），它利用稀疏编码对大量的未标注数据构建高层次特征，并利用这些特征完成有监督分类问题。由于学到的特征具有很好的表达能力，因此显著提升了分类的效果。而直接采用未标注数据进行训练也使得该方法的应用范围十分广泛。

Lee 等<sup>[23]</sup>于 2009 年提出了卷积深度信念网络 (Convolutional deep belief networks)，它是一种层级生成模型，可以实现有效地层间上下的概率推断。该网络被用于训练高维的图片，它采用卷积和概率最大池化完成对高层的特征表达，能够学到非常有用的高层视觉特征，比如目标片段 (Object parts)。Coates 等<sup>[24]</sup>于 2011 年提出了一种深度学习网络的参数调整方法，他们指出隐含层的节点数目比感受野大小、步长等参数对深度学习网络性能的影响更大。他们利用较大的隐含层节点数和密集的特征提取实现了在 CIFAR-10 和 NORB 公共数据集上的最好效果。Krizhevsky 等<sup>[25]</sup>于 2012 提出了一种深度卷积神经网络的结构，它包含 5 个卷积层、3 个最大池化层、3 个全连接层，共 6000 万个参数和 65 万个神经元，结合 GPU 卷积运算加速和 dropout 正则化方法，实现了 Imagenet LSVRC-2010 数据集上的最好效果。

在医学领域，深度学习和医学图像处理的结合也起到了很好的效果。Kim 等<sup>[26]</sup>利用两层的无监督深度卷积网络完成了 7.0 Tesla 核磁共振图像的海马体分割，与人工提取的特征相比，深度学习提取的特征更能描述 7.0 Tesla 核磁共振图像的丰富细节。他们利用学习到的特征结合多重图谱实现分割，取得了很好的效果。Xu 等<sup>[27]</sup>将深度学习和多示例学习相结合，利用深度神经网络提取组织病理学图像的特征表达，再通过多示例学习训练区域分类器，实现对组织病理学图片的区域诊断。既节省了大量标注所需的人力，又取得了很好的诊断效果。Chen 等<sup>[28]</sup>通过深度神经网络学习三维脊髓 CT 图像的细节，完成对图像中脊柱的自动精识别和定位。他们利用一种联合学习模型，同时考虑当前维度脊柱的外观和其与相邻维度的关系，有效地降低了定位失误率。Roth 等<sup>[29]</sup>利用多层深度神经网络实现了对腹部 CT 图像中胰腺的分割，他们同时训练了关于小块区域和大块区域的两个网络，并结合两个分类结果完成胰腺的精确分割。

本章将采用卷积稀疏编码完成对骨扫描图像特征的提取。稀疏编码是一种无监督学习的算法，它通过寻找输入图像的基本元素，即基底，来学习这组图像的高级视觉表达。对于输入图像，它首先从中随机地提取大量的小块区域，假设这些小区域的大小为  $d \times d$ 。接着这些小区域被转换为未标记的向量  $\{x_1, \dots, x_m\}$ ，其中  $x_i \in \mathbb{R}^s, s = d \times d$ 。

对于输入向量  $x_i$  和基底  $\phi_j (j \in 1, \dots, k)$ , 定义  $x_i$  对  $\phi_j$  的激励系数为  $a_{i,j}$ 。最后, 稀疏编码通过求解如下的优化问题得到这组输入图像的基底:

$$\min_{a, \phi} \sum_{i=1}^m \left\| x_i - \sum_{j=1}^k a_{i,j} \phi_j \right\|_2^2 + \lambda \sum_{i=1}^m \sum_{j=1}^k \|a_{i,j}\|_1 \quad (2-1)$$

其中的基底  $\phi_j$  满足  $\|\phi_j\|_2 \leq 1, \forall j \in 1, \dots, k$ 。因为要同时求解对于  $a$  和  $\phi$  最小的目标函数, 稀疏编码采用类似 EM 算法的交替迭代的方法, 将迭代分成两部分交替进行:

- 1) 首先固定住基底  $\phi$ , 求关于  $a$  的目标函数最小值, 由于  $a$  为 L-1 范数, 相当于求解一个 Lasso 问题。
- 2) 接着固定住激活系数  $a$ , 求关于  $\phi$  的目标函数最小值, 由于  $\phi$  为 L-2 范数, 相当于求解一个凸二次规划问题。

不断重复上面的两个步骤, 直到收敛, 就可以得到一组关于输入图像的基底  $\Phi = \{\phi_1, \dots, \phi_k\}$ , 其中  $\phi_j \in \mathbb{R}^s$ , 这组基底也称为输入图像的字典。得到字典之后, 就可以利用如下的方法对新的输入图像抽取特征向量, 也称为对输入图像的编码:

$$\min_a \sum_{i=1}^m \left\| x_i - \sum_{j=1}^k a_{i,j} \phi_j \right\|_2^2 + \lambda \sum_{i=1}^m \sum_{j=1}^k \|a_{i,j}\|_1 \quad (2-2)$$

对于较大的图像, 如果将所有的像素直接以全连接的方式输入神经网络, 将会有巨大量的参数需要学习, 一方面会导致训练非常耗时, 另一方面较大维数的特征也会导致过拟合的产生。对于大型图像, 一般采用卷积和池化进行处理, 从而减少提取特征的维数。如图2-2所示, 上面步骤得到的每个基底可以看成是一个滤波器, 利用每一个  $d \times d$  的小滤波器去和  $n \times n$  的大图像做卷积运算, 这样就可以得到大图像任意位置关于此基底的激活值, 即  $k$  个  $(n - d + 1) \times (n - d + 1)$  的卷积图像特征。但是这样虽然得到了关于不同基底的激活图像, 其维数却依然很大。由于图像具有位置静态属性, 因此可以尝试对不同位置的特征进行聚合统计, 比如计算一个区域激活值的平均值或最大值, 这种操作就是池化。池化可以有效的降低特征维数, 同时又能保持原图像的位置特性。对于平均池化, 我们将每幅卷积过的图像分成  $l \times l$  个均等大小的部分, 并计算每个部

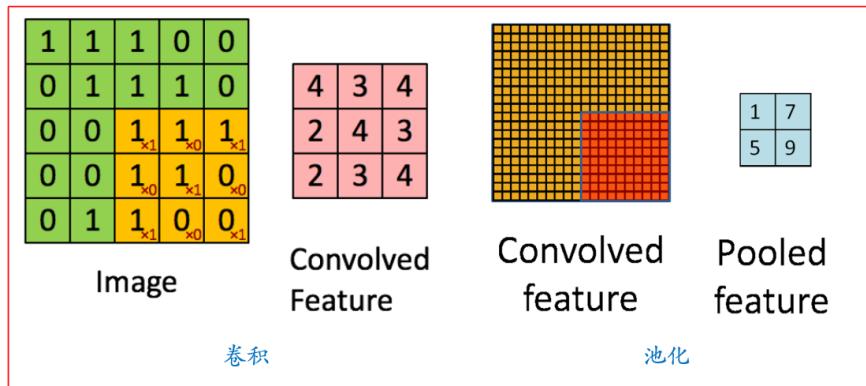


图 2-2 卷积和池化示意图

Fig. 2-2 Illustration of convolution and pooling.

分的平均值，这样就可以得到  $k$  个  $l \times l$  的池化特征。

## 2.2.2 AdaBoost 分类器

AdaBoost 算法最早由 Freund 和 Schapire<sup>[30]</sup> 提出，由于该算法具有较高的分类精度，且无需进行特征筛选、能够防止过拟合等优点而被广泛使用。AdaBoost 算法利用许多个加权的最优弱分类器组合成强分类器，它可以适应性地调整弱分类器的权重。如图2-3所示，AdaBoost 算法根据当前样本分布下的弱分类器错分率大小来适应地更新弱分类器权重和样本权重。如果弱分类器的错分率低，就调高该弱分类器的权重。如果样本被正确分类，就降低该样本的权重。这样在接下来的迭代中，弱分类器会更加关注具有较大权重的样本是否被正确分类，最终的强分类器由所有的最优弱分类器加权组合而成。AdaBoost 算法有两个常用的改进版本。GentleBoost<sup>[31]</sup> 由 Friedman 等提出，它比 AdaBoost 更稳定且鲁棒，作为一种高效的提升算法而被广泛使用。例如 Viola 等<sup>[32]</sup> 提出的人脸识别算法中就采用了 GentleBoost。Modest AdaBoost<sup>[33]</sup> 由 Vezhnevets 等于 2005 年提出，它具有比 AdaBoost 更好的可扩展性和抑制过拟合的能力。本章将同时采用 GentleBoost 和 Modest AdaBoost 两种方法训练图像级别的分类器。

对于 AdaBoost 算法中的弱分类器，一般的算法采用单层决策桩 (Stump) 来实现，由

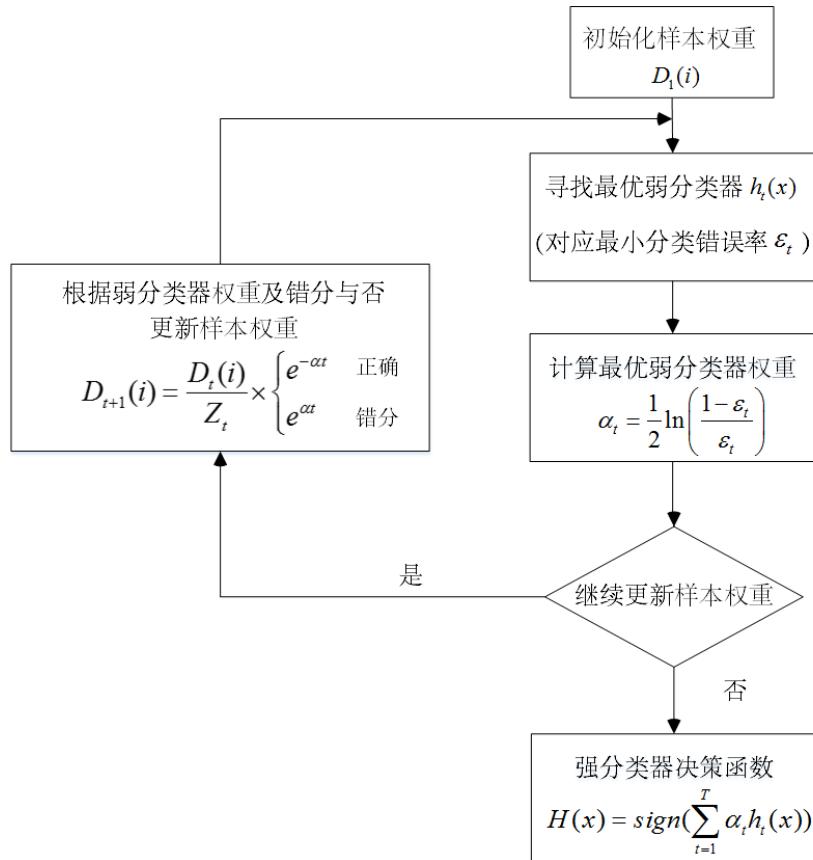


图 2-3 AdaBoost 算法流程图

Fig. 2-3 Flow chart of AdaBoost algorithm.

于它只有简单的一层和一个分叉，参数仅仅只有一个阈值，因此分类准确率不高。本章采用分类回归树（CART）来建立算法中的弱分类器，分类回归树的结构如图2-4所示，它采用二叉树对预测空间自上而下进行划分，如果满足给定的指标就对当前节点进行分裂，并给定该节点的分割阈值。对分裂节点的左右孩子，算法继续递归地进行划分，直到收敛或达到给定的层数，就构建好了一个分类回归树。本章中使用3层的分类回归树作为AdaBooost算法中的弱分类器。

### 2.2.3 图像级别的诊断

这一节将叙述图像级别分类器的训练方法，图像级别诊断的框架如图2-5所示。

为了增加相似灰度值之间的区分程度，本文采用密度分割作为输入骨扫描图像的预

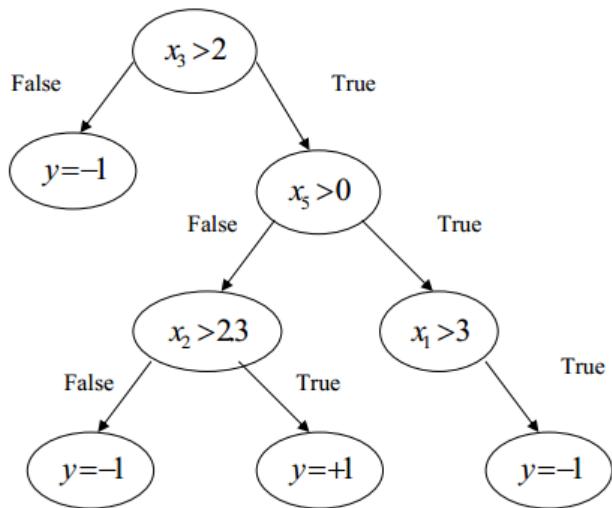


图 2-4 分类回归决策树示意图

Fig. 2-4 A example of CART.

处理步骤，将灰度图像映射到 RGB 彩色图像空间。通过第2.5小节的实验，发现采用密度分割确实可以增加识别的准确率，因此本文在进行卷积稀疏编码时，采用的都是经过密度分割后的骨扫描图像。

我们首先利用卷积稀疏编码对输入图像抽取深度特征：

- 1) 将所有胸廓骨扫描图像缩放到  $100 \times 100$ ，使其具有统一的大小。
- 2) 从骨扫描图像中随机抽取 100000 个  $11 \times 11$  的小块图像，并利用稀疏编码训练出 400 维的字典。
- 3) 对骨扫描图像进行卷积和池化的步骤，对每个基底，抽取 9 维的池化特征。

接着，分别通过 GentleBoost 和 Modest AdaBoost 算法训练图像级别的分类器。两种分类器的输出结果分别表示为 *Result G* 和 *Result M*，其符号代表了具体的分类。如果  $\text{sign}(\text{Result}) > 0$ ，则图像中包含热点；如果  $\text{sign}(\text{Result}) < 0$ ，则图像中无热点。输出结果的绝对值大小，代表了分类器对该结果的置信程度，较小的置信程度就说明分类器对该图像的诊断不是很确定。医生对骨扫描的诊断一般采用多级分类的方法，在有些情况下，医生仅根据图像本身不能做出完全肯定的诊断，因此会使用“可能存在热点”

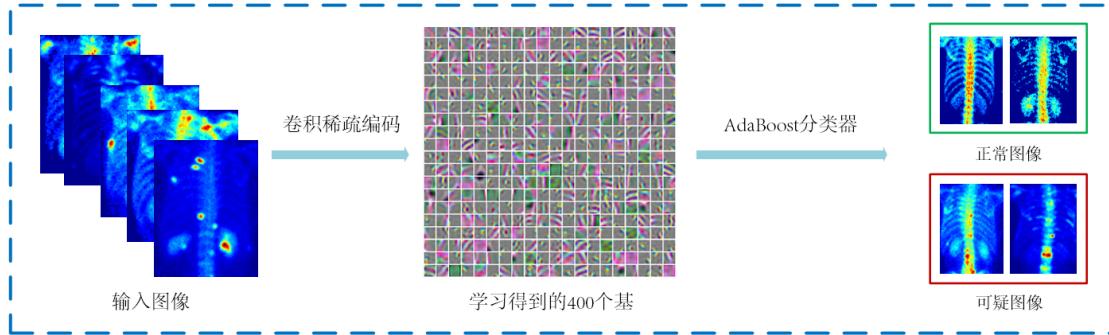


图 2-5 图像级别诊断框架

Fig. 2-5 Framework of Image-level Diagnosis.

去描述这些图像。在本章中，我们也按照医生的做法，对骨扫描图像进行三级别分类，三个级别具体描述为：

**级别 3：** 骨扫描图像中一定存在热点。

**级别 2：** 不能确定图像中是否存在热点。

**级别 1：** 骨扫描图像中不存在热点。

本章采用如下的规则融合两个分类器的分类结果，并实现三个级别的分类：

1) 如果  $\text{sign}(\text{Result}(G)) \neq \text{sign}(\text{Result}(M))$ ，则划分为级别 2。

2) 如果不满足 1) 的条件，则继续判断：

a) 如果  $\text{sign}(\text{Result}(M)) = 1$ ，且  $\text{abs}(\text{Result}(M)) < 2$  或  $\text{abs}(\text{Result}(G)) < 20$ ，

则划分为级别 2。

b) 如果  $\text{sign}(\text{Result}(M)) = 1$ ，且不满足 a) 的条件，则划分为级别 3。

c) 如果  $\text{sign}(\text{Result}(M)) = 0$ ，且  $\text{abs}(\text{Result}(M)) < 3$  或  $\text{abs}(\text{Result}(G)) < 30$ ，

则划分为级别 2。

d) 如果  $\text{sign}(\text{Result}(M)) = 0$ ，且不满足 c) 的条件，则划分为级别 1。

在第2.5小节，我们给出了图像级别分类器具体的诊断效果、准确率及分类结果与金标准的对比表。

## 2.3 基于多示例学习的区域诊断

### 2.3.1 多示例学习

传统的机器学习主要分为有监督学习和无监督学习两类。有监督学习算法的训练过程需要大量高质量的标注数据，然而标注数据会消耗大量的人力和时间。同时对于医学图像细节的标注，由于数据本身的模糊性，通过人工方法也很难对数据进行准确的标注，即使对于专业的医师也是一个有挑战性的任务。无监督学习算法虽然不需要任何人工标注，但由于缺少关于数据的知识，训练出的分类器准确率不高。半监督学习可以很好地解决这一问题，它通过较少较粗略的标注提供关于数据的知识，再通过算法完成对精细知识的自动探索和学习。在半监督学习中，多示例学习是应用很广泛的一类。它的训练集中包括一系列包，每个包又包含很多示例，训练集提供了包的标注信息，却没有提供每个具体示例的标注信息。多示例算法利用包的标注信息，通过学习去理解训练数据，并完成包和示例两个级别的分类。

多示例学习的研究最早来源于对药物活性进行预测的问题。在 1997 年，Dietterich 等<sup>[34]</sup> 提出了第一个多示例学习算法，即轴平行矩形（APR）学习算法，它通过在特征空间中寻找合适的轴平行矩形，实现对示例的分类，并在麝香数据集上取得了不错的效果。之后，很多经典的多示例学习算法被不断提出。

Zhang 等<sup>[35]</sup> 于 2001 年提出了 EM-DD 算法，它将 EM 算法与多样性密度算法（Diversity Density，简称 DD）相结合，通过 EM 算法不断迭代并找出一个包中最能代表这个包的示例，并选取这个示例来进行训练。通过这样一个示例筛选机制，EM-DD 可以把和包的标注不吻合的示例剔除，从而提升了算法的分类准确率，并且对于不一致的包内示例数目也十分鲁棒。Zhou 等<sup>[36]</sup> 于 2009 年提出了 mi-Graph 算法，和以往的其它方法假设包内示例为独立同分布不同，该算法将包内示例看成是非独立同分布的。它试着去学习包内示例间的关系，通过建立图模型和反映团簇信息的核函数导出包内示例的邻接矩阵，从而有效地提升了多示例学习算法的效果。Viola 等<sup>[37]</sup> 提出的 MILBoost 算法将多示例学习思想与 AnyBoost 框架<sup>[38]</sup> 相结合，通过提升算法的思想来解决多示例学习问

题。它采用 Noisy OR 函数定义包和示例间分类结果的关系，再通过梯度下降法不断训练最优弱分类器，并将所有弱分类器组合起来成为最终的强分类器。

由于多示例学习只需要很少很粗略的标注，且具有自动学习精细知识的特点，因此它被广泛地应用于计算机视觉和医学图像处理领域。

Babenko 等<sup>[39]</sup>于 2008 年对 Viola 等的工作进行了拓展，并将其应用到人脸识别与跟踪上。他们提出了多姿势学习 (Multiple Pose Learning)，它可以自动对数据进行聚类，并对每一类数据分别训练一个分类器。对于一副人脸图片，采用多姿势学习可以对图片中人脸的姿势进行准确的判断。Wu 等<sup>[40]</sup>于 2014 年提出了名为 MILCUT 的交互式分割方法，这种方法首先利用 SLIC 超像素<sup>[41]</sup>对图像进行划分，分割出的每个超像素被当成一个示例。接着该方法利用扫描线法在输入框内外构建正包和负包，并利用 MILBoost 算法来学习构建好的数据集。最后通过学习到的前景背景模型，该方法完成对所有示例的分类，所有被分为前景的超像素就是图像的分割结果。Kandemir 等<sup>[42]</sup>采用多示例学习完成巴雷特食管癌的检测，他们分别对组织图像提取细胞级别和小块级别的特征，并通过 mi-Graph 算法获得了最好的诊断效果。由于 MILBoost 算法只能完成二分类问题，Xu 等<sup>[43]</sup>提出了它的改进版本——MCILBoost 算法。该算法可以在样本正负分类的同时，完成正样本内的聚类，将正样本分成几个类别。对于 softmax 函数，他们改用效果更好的广义平均 (Generalized Mean) 来定义包和示例间分类结果的关系。在文章中，作者利用 MCILBoost 同时完成了组织病理学肿瘤图像级别的分类、像素级别的分割与区域级别的聚类，取得了很好的效果。

本文采用 MILBoost 来训练区域级别的热点分类器，下面将详细地介绍 MILBoost 算法。

在多示例学习中，训练集中包含正包和负包，每个包都包含很多未标注的示例，在正包中至少有一个示例是正样本，而在负包中所有的示例都是负样本。一个包可以表示为  $X_i = \{x_{i1}, \dots, x_{im}\}$ ，其中  $i$  是包的索引，而  $j$  是示例的索引。包的标签可以表示为  $y_i \in \{-1, 1\}$ ，由于示例的标签没有被明确标出，因此表示为隐含的标签  $y_{ij} \in \{-1, 1\}$ 。

根据多示例学习的设定，如果一个包中的  $m$  个示例里有一个为正样本，则该包是正包；如果一个正样本都没有，则该包是负包。因此，包的标签  $y_i$  和该包中示例的标签  $y_{ij}$  具有如下的关系：

$$y_i = \max_j(y_{ij}) \quad (2-3)$$

多示例学习算法的目标是要学习一个示例级别的分类器  $h(x_{ij})$ 。包级别的分类器  $H(x_i)$  可以通过示例级别的分类器表示出来：

$$H(x_i) = \max_j(h(x_{ij})) \quad (2-4)$$

**MILBoost** 算法利用提升方法来解决多示例学习问题，它遵循 AnyBoost 框架<sup>[38]</sup>，通过梯度下降法不断优化损失函数  $\mathcal{L}(h)$  来训练出一系列弱分类器  $h_t$ ，并将这些弱分类器组合而成强分类器  $h$ ：

$$h(x_{ij}) = \sum_{t=1}^n \alpha_t h_t(x_{ij}) \quad (2-5)$$

上式中的  $\alpha_t$  是弱分类器的权重。根据 AnyBoost 框架，MILBoost 算法通过梯度下降法来寻找使损失函数最小的最优的弱分类器，所使用的损失函数是负对数的形式：

$$\mathcal{L}(h) = - \sum_{i=1}^n (\mathbf{1}(y_i = 1) \log p_i + \mathbf{1}(y_i = -1) \log(1 - p_i)) \quad (2-6)$$

其中， $p_i$  是包  $i$  为正的概率，它可以通过广义均值 *softmax* 函数计算出来： $p_i = \frac{1}{m}(\sum_{j=1}^m p_{ij}^r)^{\frac{1}{r}}$ 。MILBoost 算法中使用 *softmax* 函数作为对 *max* 函数的一种近似，这是因为梯度下降法需要计算损失函数的导数，因此需要找到 *max* 函数可微分版本的近似。上式中的  $p_{ij}$  是包  $i$  中示例  $j$  为正的概率，它可以由  $h_{ij}$  通过 *sigmoid* 函数计算出来： $p_{ij} = \frac{1}{1+\exp(-2h_{ij})}$ 。其中的  $h_{ij} = \sum_t \alpha_t h_t(x_{ij})$  是上一次分类器的输出结果。接着 MILBoost 采用梯度下降法来计算示例的权重：

$$w_{ij} = -\frac{\partial \mathcal{L}}{\partial y_{ij}} = \begin{cases} 2 \frac{p_{ij}^r - p_{ij}^{r+1}}{\sum_{j=1}^n p_{ij}^r} & \text{if } y_i = 1 \\ -\frac{2p_i}{1-p_i} \frac{p_{ij}^r - p_{ij}^{r+1}}{\sum_{j=1}^n p_{ij}^r} & \text{if } y_i = -1 \end{cases} \quad (2-7)$$

在得到权重后，算法继续利用这些权重的绝对值来找出最优的弱分类器：

$$h_t = \arg \min_h \sum_{i,j} \mathbf{1}(h_{candidate}(x_{ij}) \neq y_i) |w_{ij}| \quad (2-8)$$

并通过线性搜索计算出弱分类器的权值  $\alpha_t$ ：

$$\alpha_t = \arg \min_\alpha \mathcal{L}(h + \alpha h_t) \quad (2-9)$$

最后，将新得到的弱分类器加到之前的强分类器，得到更新之后的强分类器。当达到收敛条件或预设的弱分类器个数时，算法的训练过程结束，可以得到最终的示例级别的分类器：

$$h(x_{ij}) = \sum_{t=1}^T \alpha_t h_t(x_{ij}) \quad (2-10)$$

### 2.3.2 骨扫描图像特征提取

为了能够实现对骨扫描区域的热点检测，设计一套能够反映局部图像信息的特征至关重要。前面我们采用卷积稀疏编码的方法完成了对骨扫描图像的细节描述，但是由于提取的深度特征维数较大，而区域热点识别需要利用分类器对图像进行扫描，运算量很大。由于实时性的要求，深度特征无法直接应用在区域级别热点的检测上。为此，在本节中，我们设计了一套能够有效反映骨扫描图像细节的特征，它对于热点区域和非热点区域有很好的区分性。

由于原始图像是全身骨扫描图像，因此本节首先阐述分割出胸廓区域的方法。同时也将介绍人体中心线的计算方法，准确的人体中心线对于计算之后要提到的对称对比度

特征有很大的帮助。按照 Wang 等<sup>[9]</sup>提出的方法，首先对全身骨扫描图像进行预处理，先将其二值化，将灰度值小于 6 的像素设置为 0，接着通过中值滤波去除二值化过程中产生的噪声。接着根据二值化后的图像计算出头部和双脚的纵轴坐标，并画出头部和双脚两条水平线，分别称为头部基准线和双脚基准线，将两条线之间的距离作为病人身高。最后根据人体解剖学的知识，将头部基准线以下 0.17-0.25 倍的病人身高范围作为肩膀点纵轴区间，将头部基准线以下 0.38-0.45 倍的病人身高范围作为盆骨上端纵轴区间。在区间中寻找每个纵轴位置最左和最右第一个不为零的坐标，计算出左右两边的宽度，对于前后宽度变化最大的位置，分别作为最终的肩膀点和盆骨上端点，从而完成胸廓区域的提取。对于人体中心线，本文利用如下方法进行计算。对之前的二值化图像自上而下进行逐行的水平扫描，记录每一行左边和右边第一个不为 0 的像素点的横坐标，将其作为每一行人体轮廓的左边界点和右边界点，分别记为： $left(i)$  和  $right(i)$ 。最后将左右两个横坐标取平均，得到每一行中心点的纵坐标： $c(i) = \frac{left(i)+right(i)}{2}$ 。

本文对  $4 \times 4$  大小的区域提取 29 维的特征向量，包含了直方图特征、对比度特征和纹理特征三个部分：

直方图特征包括了区域中像素灰度值的平均值、最大值、最小值、最大值与均值的比值和中位值，一共 6 个特征。

对比度特征包含了邻域对比度和对称对比度两个部分。邻域对比度是当前区域与其上下左右四个同样大小区域的卡方距离，一共 4 个特征，其中  $M$  和  $N$  分别是当前区域和四个邻域的灰度直方图统计信息：

$$Neighbor\ Contrast(M, N) = \frac{1}{10} \sum_{i=1}^n 2^i (M_i - N_i)^2 \quad (2-11)$$

而邻域对比度是当前区域与其关于人体中心线对称区域的卡方距离：

$$Symmetry\ Contrast(M, S) = \sum_{i=1}^n \frac{(M_i - S_i)^2}{M_i + S_i + 1} \quad (2-12)$$

对于纹理特征，本文采用局部二进制模式 (Local Binary Patterns, LBP)<sup>[44]</sup> 描述骨扫

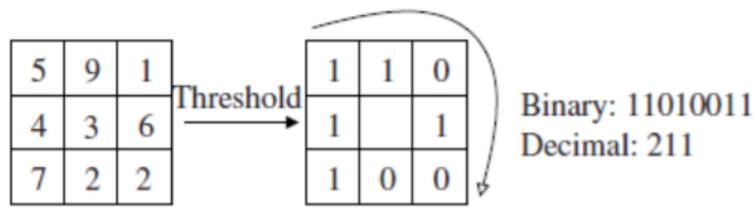


图 2-6 局部二进制特征示意图

Fig. 2-6 Illustration of local binary pattern.

描图像的纹理。传统的 LBP 算子采用区域中心的灰度值作为阈值对整个区域进行二值化，再按某一方向如顺时针对邻域灰度值进行编码，具体原理如图2-6所示。

由于骨扫描图像存在大量的噪声，因此采用区域中心的灰度值作为阈值描述效果并不好，本文对传统方法进行改进，采用区域的均值作为二值化的阈值，这样得到的 LBP 响应矩阵能够清晰地重现图像的纹理。本文最后将得到的响应矩阵进行量化，得到一个 18 维的纹理特征。

所有的 29 维骨扫描图像特征如表2-1所示：

表 2-1 抽取的 29 维骨扫描图像特征

Table 2-1 29-dimension features extracted from bone scan images

特征名称	维数	描述
直方图特征	6	包含均值、标准差、最大值、最小值、最大值均值比、中值
邻域对比度特征	4	所选区域与上、下、左、右四个区域的对比度
对称对比度特征	1	所选区域与关于人体中心线对称区域的对比度
局部二进制特征	18	以整幅图像灰度均值作为阈值并经过量化得到的纹理特征

### 2.3.3 区域级别的诊断

本节将介绍区域级别热点诊断的整体框架和并演示具体的诊断效果。区域级别的热点诊断框架如图2-7所示。为了从骨扫描图像中提取训练数据集，我们首先从图像中

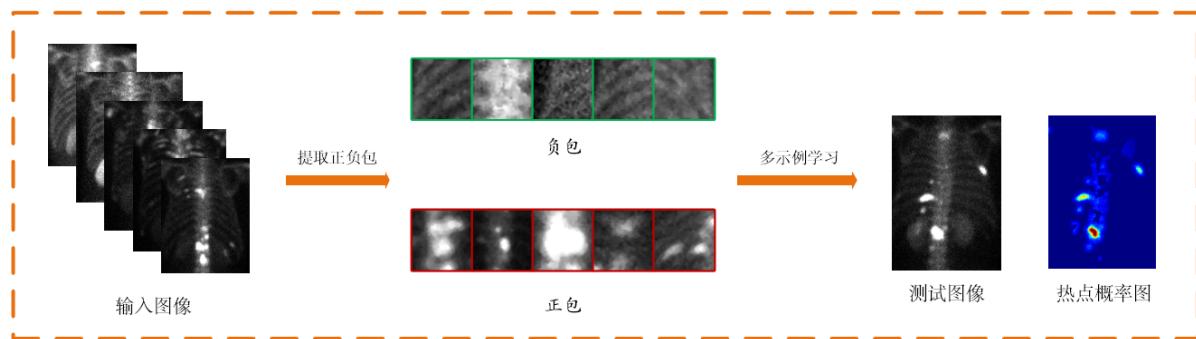


图 2-7 区域级别诊断框架

Fig. 2-7 Framework of Patch-level Diagnosis.

选择较大的区域并把它们作为多示例学习中的包，并对这些包进行了人工的标注。接着从这些包中以 2 个像素的步长紧密采样  $4 \times 4$  大小的区域作为包中的示例。本文一共从骨扫描图像库中提取了 39 个正包和 33 个负包，这些包一共包含 18947 个示例。我们用 MILBoost 算法对训练数据集进行学习，得到了区域级别的分类器，训练时弱分类器的个数被设置为  $T = 50$ 。该分类器可以用来识别一个区域中是否包含热点，也可以对区域进行整体扫描，生成热点的概率图。

图2-8给出了本文训练的区域分类器对于热点区域的识别例子，图中的百分比代表了分类器的置信度。从图中可以发现，该分类器对于脊柱和肋骨位置有热点的区域都能很好地检测出来。对于没有热点的区域，该分类器也有很好的特异性。

本文利用训练出的分类器对胸廓骨扫描图像进行全局扫描，并将分类器的输出作为热点的概率，绘制出胸廓骨扫描图像的热点概率图。利用生成的热点概率图，可以完成后续的热点自动分割步骤。图2-9给出了一些用上述方法生成的热点概率图。图中 I 部分是两张异常图像的热点概率图，可以发现，热点概率图可以很好地反映热点的分布位置和存在可能性。图中 II 部分是两张正常图像的热点概率图，可以看出，两张图都几乎没有亮的区域，说明该方法能够很好地识别区域中是否包含热点。

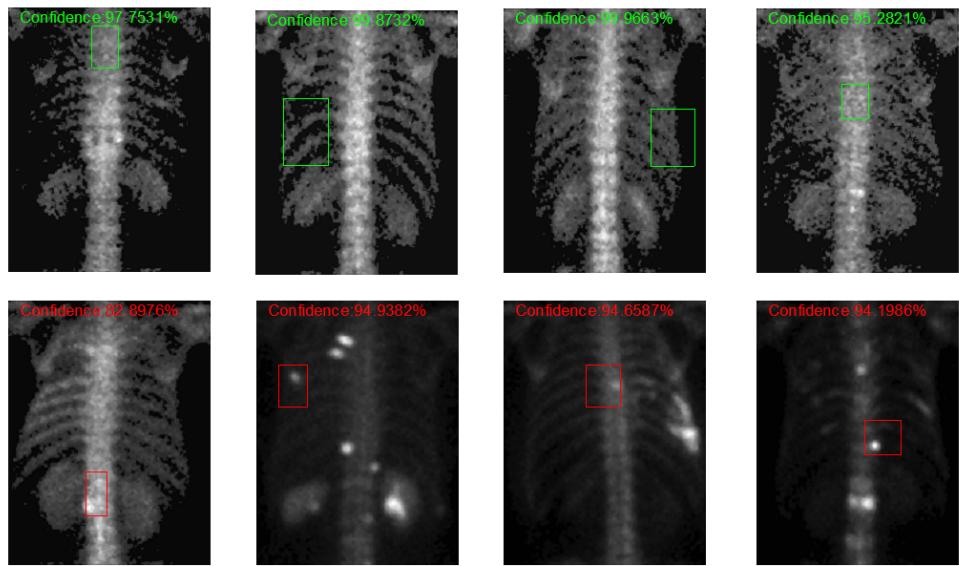


图 2-8 区域识别示例

Fig. 2-8 Examples of Patch-level Diagnosis.

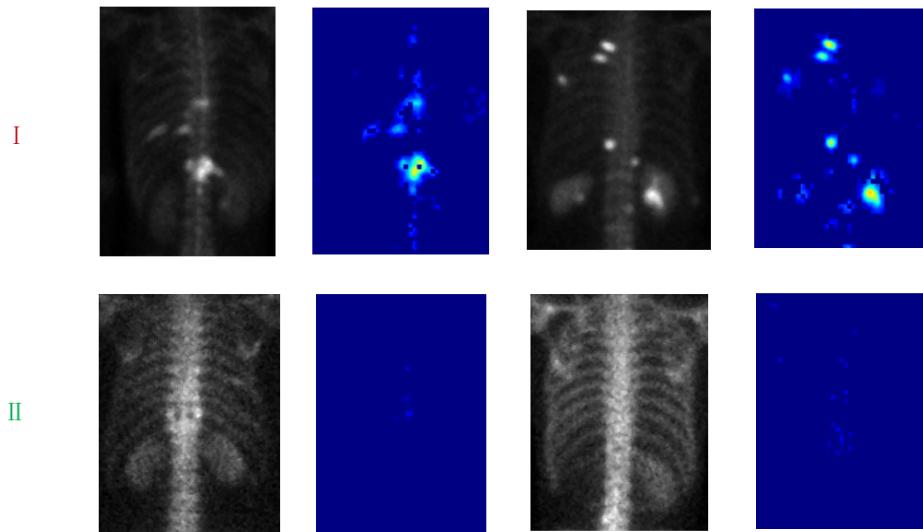


图 2-9 热点概率图示例

Fig. 2-9 Examples of hotspot probability map.

## 2.4 局部符号差异水平集分割

### 2.4.1 水平集方法

以前的骨扫描计算机辅助诊断系统主要采用 OTSU 自适应阈值分割<sup>[21]</sup>、区域生长法<sup>[10]</sup>这两种方法完成热点的分割。OTSU 自适应阈值分割根据图像的灰度统计特征，寻找使分割的两部分类间方差最大的灰度值，作为最优阈值。但该方法的效果依赖于前期对骨扫描图像每个部分的分割，并且容易受到噪声的干扰。区域生长法利用用户给定的种子点，将原有区域向四周的相似区域不断扩展。但该方法需要用户给定生长种子点和增长系数，对于不同的热点，由于最优的参数不同，因此分割的效果不稳定，容易出现泄漏现象。水平集方法是一种数值方法，由 Osher 和 Sethian<sup>[45]</sup>于 1988 年首次提出，它将界面表示为高一维空间中函数  $\phi$  的零水平集，并通过  $\phi$  的不断迭代演化，实现能量函数的最小化，从而获得界面的形状。由于水平集具有很好的拓扑特性和亚像素精度，它被广泛应用在图像分割，尤其是医学图像分割领域。

水平集分割可以分成基于边界的水平集和基于区域的水平集两大类。基于边界的水平集方法利用图像的边缘信息吸引主动轮廓曲线向被分割物体的边界进化。这类方法对曲线初始化位置和噪声都很敏感，因此并未得到广泛的使用。和基于边界的水平集方法相比，基于区域的水平集有效地克服了以上的问题，并且在分割弱边界目标上的表现更胜一筹。根据颜色模型建立方法的差异，基于区域的水平集也可以分为两类：全局方法和局部方法。Chan 和 Vese<sup>[46]</sup>在 2001 年提出的 CV 模型是最经典的全局方法，它假设每个被分割区域都具有均匀的统计分布，并可以自动检测出图像中的所有边界。用  $\Omega$  表示二维的图像域， $C$  表示在  $\Omega$  上进化的一条曲线， $I$  表示给定的灰度图像， $I(x)$  表示在  $x$  位置的图像灰度值。**CV** 模型试着把图像分成两个互不重叠的两部分，分别将曲线内和曲线外的图像区域记为  $\Omega_f$  和  $\Omega_b$ ，则 **CV** 模型的全局拟合能量函数可以表示为：

$$\mathcal{E}_{CV}(C) = \lambda_f \int_{\Omega_f} |I(x) - c_f|^2 dx + \lambda_b \int_{\Omega_b} |I(x) - c_b|^2 dx \quad (2-13)$$

其中的  $c_f$  和  $c_b$  分别是两个区域的均值， $\lambda_f$  和  $\lambda_b$  是正常数。但是由于 **CV** 模型比较

简单，在分割一些灰度不均匀的图像时，就会产生不准确的分割结果。局部方法通过考虑局部的灰度信息来控制曲线的进化，从而解决了全局方法对于灰度不均图像分割效果欠佳的问题。Li 等<sup>[47]</sup>于 2008 年提出的基于核函数的局部二元拟合（LBF）能量函数就是典型的局部方法。用  $x$  表示局部区域的中心， $y$  表示局部区域内部的像素， $I(y)$  表示在  $y$  位置的图像灰度值，则 LBF 模型的局部拟合能量函数可以表示为：

$$\mathcal{E}_{LBF}(x) = \lambda_f \int_{\Omega_f} K_{\sigma,x}(y) |I(y) - C_f^x|^2 dy + \lambda_b \int_{\Omega_b} K_{\sigma,x}(y) |I(y) - C_b^x|^2 dy \quad (2-14)$$

其中的  $K_{\sigma,x}(y) = \frac{1}{2\pi\sigma^2} \exp\left\{-\frac{\|x-y\|^2}{2\sigma^2}\right\}$  是所使用的高斯核函数，而  $C_f^x$  和  $C_b^x$  分别是局部的前景簇和背景簇。将局部的能量进行累加，可以算出整体的能量：

$$\mathcal{E}_{total}(C) = \int_{\Omega} \mathcal{E}_{LBF}(x) dx \quad (2-15)$$

然而，虽然解决了灰度不均匀的问题，局部方法却有对初始轮廓敏感的问题。为了解决这一问题，很多方法尝试结合全局和局部两种方法的优势。Zhang 等<sup>[48]</sup>提出了一种全新的局部压力函数，可以保证水平集函数停留在模糊的弱边界处，同时允许用户选择进行局部分割还是全局分割。Wang 等<sup>[49]</sup>提出了局部符号差异（LSD）能量，它综合考虑局部可分性和全局的连贯性，达到同时兼顾局部和整体的分割效果。该方法通过局部符号项将局部区域的顺序引入能量函数，有效提升了分割效果的鲁棒性。由于局部符号差异能量函数具有的优良特性，本文将采用局部符号差异水平集实现热点的自动分割。

## 2.4.2 局部符号差异水平集

局部符号水平集将局部方法和全局方法的优势相结合，其能量函数包含局部符号项和局部差异项两个部分。局部符号能量的定义如下：

$$\mathcal{E}_{LSD}(x) = sgn(C_b^x - C_f^x) |C_b^x - C_f^x| \quad (2-16)$$

上式中的  $sgn(C_b^x - C_f^x)$  是局部符号项，它可以表达局部前景簇和背景簇之间的顺序关系，从而反映图像的全局连贯性。而  $|C_b^x - C_f^x|$  则代表了局部差异项，它的大小决定了局部前景簇和背景簇之间的差异，反映了图像的局部可分性。局部符号能量可以化简成： $\mathcal{E}_{LSD}(x) = C_b^x - C_f^x$ 。根据梯度下降法，可以表示出局部前景簇和背景簇的值：

$$C_f^x = \frac{\int_{\Omega_f} K_{\sigma,x}(y)I(y)dy}{\int_{\Omega_f} K_{\sigma,x}(y)dy}, \quad C_b^x = \frac{\int_{\Omega_b} K_{\sigma,x}(y)I(y)dy}{\int_{\Omega_b} K_{\sigma,x}(y)dy} \quad (2-17)$$

而 LSD 能量可以表示为：

$$\mathcal{E}_{LSD}(x) = \frac{\int_{\Omega_b} K_{\sigma,x}(y)I(y)dy}{\int_{\Omega_b} K_{\sigma,x}(y)dy} - \frac{\int_{\Omega_f} K_{\sigma,x}(y)I(y)dy}{\int_{\Omega_f} K_{\sigma,x}(y)dy} \quad (2-18)$$

全局的能量可以表示成局部能量的积分形式，即  $\mathcal{E}_{total}(C) = \int_{\Omega} W(x)\mathcal{E}_{LSD}(x)dx$ ，其中  $W(x)$  是位置  $x$  下的 LSD 能量系数，其定义为：

$$W(x) = \int_{\Omega_b} K_{\sigma,x}(y)I(y)dy \int_{\Omega_f} K_{\sigma,x}(y)I(y)dy \quad (2-19)$$

最终的全局能量可以表示为：

$$\begin{aligned} \mathcal{E}_{total}(C) = & \int_{\Omega} \left( \int_{\Omega_b} K_{\sigma,x}(y)I(y)dy \int_{\Omega_f} K_{\sigma,x}(y)I(y)dy - \right. \\ & \left. \int_{\Omega_f} K_{\sigma,x}(y)I(y)dy \int_{\Omega_b} K_{\sigma,x}(y)I(y)dy \right) dx \end{aligned} \quad (2-20)$$

通过引入轮廓的周长项作为先验知识，并利用 Lipschitz 函数  $\phi : \Omega \rightarrow R$  来表示 LSD 全局能量，用  $H$  表示 Heaveside 函数，全局能量可以表示成：

$$\begin{aligned} \mathcal{E}(\phi) = & \int_{\Omega} \left( \int_{\Omega} K_{\sigma,x}(y)H(\phi(y))I(y)dy \int_{\Omega} K_{\sigma,x}(y)dy - \right. \\ & \left. \int_{\Omega} K_{\sigma,x}(y)I(y)dy \int_{\Omega} K_{\sigma,x}(y)H(\phi(y))dy \right) dx + \mu L(\phi) + v P(\phi) \end{aligned} \quad (2-21)$$

其中  $L(\phi) = \int_{\Omega} |\nabla H(\phi(x))| dx$  为周长项,  $P(\phi) = \int_{\Omega} \frac{1}{2}(|\nabla \phi(x)| - 1)^2 dx$  为正则项。接着利用标准的梯度下降法解如下的梯度流方程, 使得全局能量  $\mathcal{E}(\phi)$  最小化:

$$\frac{\partial \phi}{\partial t} = -\frac{E(\phi)}{\partial \phi} = -\delta_\epsilon(\phi)DF + \mu\delta_\epsilon(\phi)div\left(\frac{\nabla \phi}{|\nabla \phi|}\right) + \nu\left(\nabla^2 \phi - div\left(\frac{\nabla \phi}{|\nabla \phi|}\right)\right) \quad (2-22)$$

其中的 DF 表示数据压力项, 它来自于局部符号能量:

$$DF = \int_{\Omega} \left( K_{\sigma,x}(y)I(y) \int_{\Omega} K_{\sigma,x}(y)dy \right) dx - \int_{\Omega} \left( K_{\sigma,x}(y) \int_{\Omega} I(y)K_{\sigma,x}(y)dy \right) dx \quad (2-23)$$

### 2.4.3 热点自动分割

在生成骨扫描图像热点概率图后, 本文将热点概率图  $\mathcal{P}(x)$  以  $\rho$  作为阈值进行二值化, 并将其作为水平集方法的初始轮廓  $\phi_0$ , 具体的方法可以表示为:

$$\phi_0 = \phi(x, t=0) = \begin{cases} -c, & x \in \{x | \mathcal{P}(x) > \rho\} \\ 0, & x \in \{x | \mathcal{P}(x) = \rho\} \\ c, & x \in \{x | \mathcal{P}(x) < \rho\} \end{cases} \quad (2-24)$$

本文将  $\phi_0$  的值设置为 0.6。接着通过多次迭代, 使水平集函数  $\phi$  不断进化, 完成对热点的分割:

$$\phi^{n+1} = \phi^n + \Delta t \frac{\partial \phi^n}{\partial t} \quad (2-25)$$

其中  $\Delta t$  为进化的时间步长, 在本文中我们将  $\Delta t$  设置为 0.12, 正则项系数  $\nu$  设置为 100, 周长项系数  $\mu$  设置为 1, 高斯核窗的大小设置为 5。传统的水平集需要设置循环次数作为终止条件, 本文对终止条件进行了改进, 算法不断比较相邻两次的进化结果, 如果迭代次数大于 200 次且相邻两次轮廓面积的差小于 2, 就终止水平集的迭代。这样可以避免水平集不断进化, 但轮廓基本不变的问题, 从而可以更快地完成分割过程。

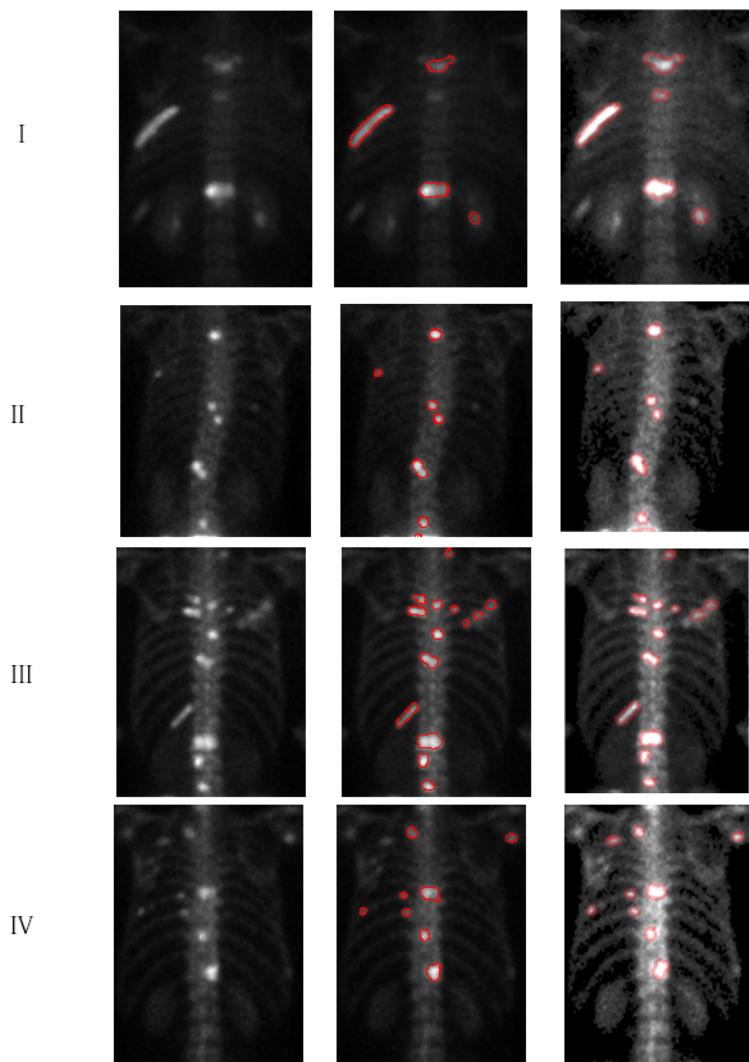


图 2-10 热点自动分割示例

Fig. 2-10 Examples of automatic hotspot segmentation.

图2-10给出了一些热点自动分割的例子，其中第一列是胸廓骨扫描图像的原图，第二列是本文提出的骨扫描自动分割算法的分割结果，第三列是医生手动进行标注的金标准。从中可以看出，本文提出的方法可以有效地从胸廓骨扫描图像中把热点分割出来。在分割的同时，本文的方法保证了分割出的边界尽量贴近热点的真实边界，在第2.5小节的实验中，我们将给出具体的定量结果来衡量本文提出的热点自动分割算法。

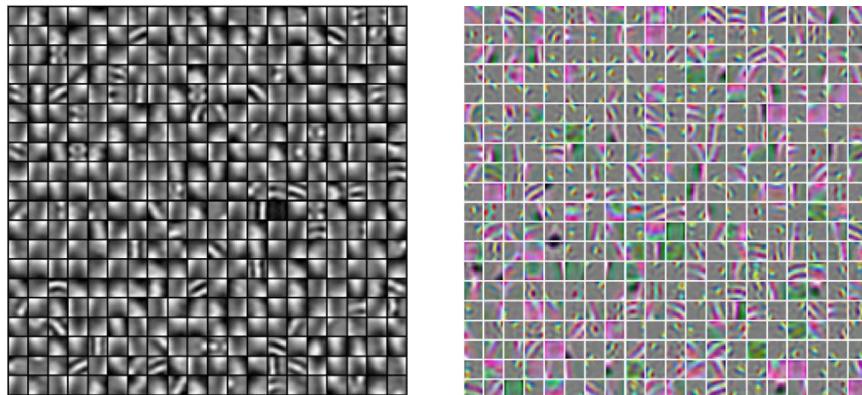


图 2-11 采用和不采用密度分割训练出的基

Fig. 2-11 Trained Bases with or without density slicing.

## 2.5 实验结果

在这一小节，我们将通过三个定量实验来证明本章所提出方法的有效性。首先将阐明使用密度分割后的图像进行卷积稀疏编码的原因。其次将对本章提出的图像级别诊断分类器的准确率进行测试。最后将对比本章提出的热点自动分割算法与之前方法的分割准确率。

**实验 I** 在本实验中，我们选择 1030 张胸廓骨扫描图像，将它们缩放成  $100 \times 100$  的大小。并从这些图像中随机抽取 100000 个  $11 \times 11$  的小区域来进行卷积稀疏编码的训练。我们分别将使用密度分割的彩色图像和不使用密度分割的灰度图像作为输入。通过稀疏编码训练得到的字典如图2-11所示。为了比较它们对骨扫描图像细节的表达能力，我们利用支持向量机对它们抽取出来的特征进行训练，并利用 10 折交叉验证测试训练出分类器的识别准确率，得到的结果如表2-2所示：从表2-2中可以看出，将密度分割作

表 2-2 采用密度分割前后交叉验证准确率对比

Table 2-2 Cross-validation accuracy with or without density slicing

训练数据集	彩色图像（密度分割）	灰度图像（未密度分割）
交叉验证准确率	0.947	0.895

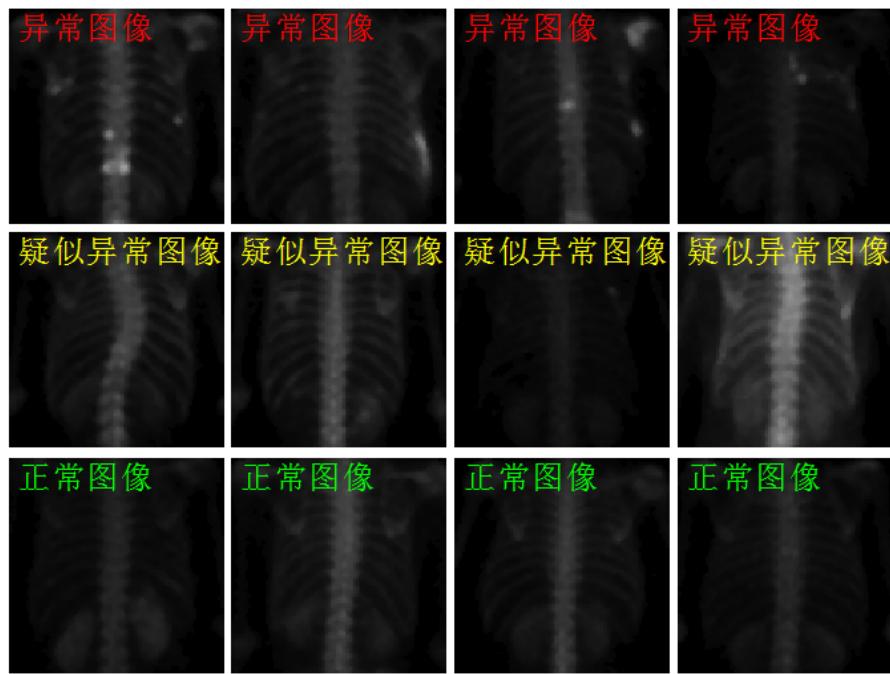


图 2-12 图像级别诊断的例子

Fig. 2-12 Examples of Image-level Diagnosis.

为预处理的步骤，确实可以增强训练出深度特征的区分能力。

**实验 II** 在本实验中，我们选择 111 张胸廓骨扫描图像，由医生对其进行 3 级别标注，根据医生的金标准，一共包含 42 张级别 3 的图像、30 张级别 2 的图像和 39 张级别 1 的图像。接着，利用本章提出的图像级别诊断分类器对这些图像进行 3 个级别的分类，图2-12给出了图像级别诊断的一些结果。从图中可以看出，本文提出的图像级别诊断分类器能够对胸廓骨扫描图像实现正确的诊断，同时完成对图像的 3 级别分类。我们将本文分类器的诊断结果与医生的金标准进行了对比，具体的结果如表2-3所示。从表中可以发现，本文提出的图像级别分类器对于级别 3 图像的分类准确率为 90.5%，假阴率为 2.4%。对于级别 2 的图像，分类准确率为 86.7%。对于级别 1 的图像，分类准确率为 92.3%，假阳率为 0%。本文分类器的整体准确率为 90.1%。如果分类器将级别 3 和级别 1 的图像分类成级别 2，则代表分类器对于当前图像是否含有热点并不是很确定，若不把这种情况当成错分，则分类器的整体失误率仅有 0.9%。May Sadik 等<sup>[5]</sup>在文中提

到，根据医生经验的不同，金标准的实际准确率为 82% 到 98%。根据上述数据，可以认为本文提出的图像级别诊断分类器取得了不错的诊断效果。

表 2-3 图像级别诊断和金标准分类结果对比

Table 2-3 Classifications Made by CAD System, Compared with Gold Standard

图像诊断	金标准			合计
	级别 3	级别 2	级别 1	
级别 3	38	1	0	39
级别 2	3	26	3	32
级别 1	1	3	36	40
合计	42	30	39	111

**实验 III** 在本实验中，我们将本章提出的热点自动分割算法与之前骨扫描辅助诊断系统常用的两种方法进行了比较。我们选择了 68 张含有热点的胸廓骨扫描图像作为测试集，并邀请专业的医生对热点区域进行了标注，其中一共含有 572 个热点。在图像分割领域，最常用的衡量分割准确程度的指标是 Jaccard 指数和 Dice 指数，设  $S$  为算法分割的结果， $G$  为标准的结果，则两种指标可以表示为：

$$J = \frac{|S \cap G|}{|S \cup G|} , \quad D = \frac{2|S \cap G|}{|S| + |G|} \quad (2-26)$$

表2-4给出了三种方法的分割效果和两种指标的值。从表中我们可以发现，本章提出的热点自动分割方法比传统的方法分割地更精确，分割准确率也更高。图2-13给出了分割效果的对比图，其中 (a) 是骨扫描原图，(b) 是本文提出的方法的分割效果，(c) 是自适应阈值法的分割效果，(d) 是区域生长法的分割效果，(e) 是医生标注的热点轮廓金标准。从图中可以看出，本文的方法在分割效果上也比之前的方法要好，而且分割出的热点边界光滑，和金标准非常接近。

通过以上的三个定量实验，我们证明了本文提出的热点自动分割方法的有效性。

表 2-4 各种分割方法准确率对比

Table 2-4 Accuracy comparison of different segmentation methods

方法	提出的方法	区域生长	阈值分割法
Jaccard 指数	0.8051	0.6505	0.5183
Dice 指数	0.8887	0.7613	0.6574

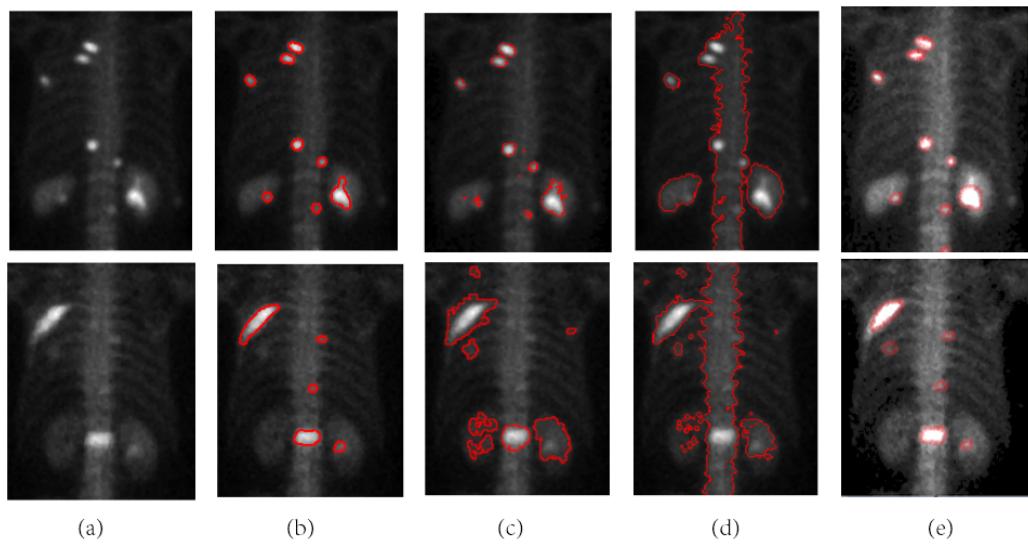


图 2-13 热点自动分割方法对比

Fig. 2-13 Comparison of different automatic hotspot segmentation methods.

## 2.6 本章小结

本章提出了一套全新的骨扫描图像热点自动分割框架。首先通过卷积稀疏编码提取骨扫描图像的深度特征，并利用 AdaBoost 算法训练图像级别的分类器，实现了图像级别的诊断。接着，提出了一套 29 维的骨扫描区域特征，并利用多示例学习算法训练了区域级别的热点分类器，完成区域级别的热点诊断并生成热点的概率图。最后，利用热点概率图对局部符号水平集进行初始化，并实现对图像中热点的自动分割。本章通过后续的三个定量实验，证明了提出的热点自动分割框架的有效性和准确性。



## 第三章 基于平均场和水平集的热点交互式分割

上一章提出了一种热点的自动分割方法，能够快速而有效地将热点从骨扫描图像中分割出来。然而，自动的分割方法不可能对所有的图像都有满意的分割结果，必然会产生一些未能分割出来或分割效果不好的热点。对于这种情况，就需要一种交互式的分割算法，使得用户可以通过提供最少量的信息，实现对热点的分割和局部修饰。本章提出了一种热点的交互式分割算法，它基于平均场和水平集完成对热点的精确分割。同时，还提供了几种不同的局部修饰方法，可以为医生分割热点提供很大的方便。

### 3.1 问题概述

医学影像引导治疗 (Image-guided Therapy)，简称 IGT，它利用各种模态的医学影像进行外科手术流程的规划、评估和干预。在 IGT 治疗过程中，对手术目标的识别和定位至关重要，而这都依赖于前期对目标的精准分割。IGT 中的分割方法主要分为 4 类：人工分割、半自动分割、交互式分割和自动分割。人工分割虽然准确程度最高，但是需要花费医生大量的时间，而且实际操作非常不方便。半自动分割通常需要用户人为地调整一系列参数，从而完成对目标的分割。虽然可以在一定程度上减少人力负担，但调参的方法对于终端用户却非常难学。全自动分割能够为医生的诊断提供便利和参考信息，但是它却并不能每次都给出完美的结果。交互式分割算法可以利用用户提供的少量信息，完成对特定目标的分割，由于其分割过程简便且效果好，因此在 IGT 中应用的最多。

很多学者提出了他们的医学图像交互式分割方法。Ben-Zadok 等<sup>[50]</sup> 提出了一种交互式分割框架，将离散的用户反馈表示为一个连续能量项，并作为附加项融合到水平集函数中。他们用此方法去分割子宫肌瘤的核磁共振图像，取得了很好的效果。Zhu 等<sup>[51]</sup> 提出了一种结合信度传播和水平集的交互式分割方法，这种方法不断迭代，计算每个像素的置信度，使最终的分割结果达到全局最优。由于利用了局部的用户信息和全局的图

像特征，此方法对弱边界和噪声具有较强的鲁棒性。Wang 等<sup>[52]</sup> 利用在线增强学习实现了基于上下文的交互式图像分割。此方法建立了一个隐式模型，可以根据用户的输入理解用户的意图，并通过在线学习实现自适应的分割。能够显著减少用户交互次数，同时保持分割的准确性和连贯性。

本文作者所在的课题组曾对上海交通大学仁济医院核医学科的医生进行了关于骨扫描图像辅助诊断系统的调研。医生们普遍反映，他们需要一种交互式的热点分割工具，通过尽量少的输入，精准地分割出热点的轮廓，同时给出热点的亮度分布以及面积等参考信息。本章就将实现这样一套热点交互式分割算法，它通过平均场推断将用户信息进行传播，并利用水平集完成对热点的精确分割。同时，本章还通过图割算法等方法实现了热点分割结果的局部修饰，能够完成热点轮廓的扩张、缩小和消除。

## 3.2 平均场推断

在计算机视觉的研究中，经常需要对现实世界进行建模。概率图模型是一种简洁而明确的语言，我们可以用它对多个变量进行建模。通过它，我们可以将已知信息和未知变量相关联，并通过已知信息去推断未知变量<sup>[53]</sup>。概率图模型在计算机视觉领域有着广泛的应用，图像分割、图像配准、目标跟踪、图像去噪等问题都可以通过建立概率图模型来解决。概率推断是概率图模型最重要的应用，它主要用来计算隐含变量的边缘概率以及可能性最大的状态。对于一部分图模型，如树状概率图、单环概率图和二值状态的概率图，能够通过精确推断解出边缘概率<sup>[54]</sup>。但是，对于大部分的图模型，求解它们的精确解是一个 NP-Hard 问题，无法在多项式时间内解出。因此，一般采用近似推断的方法来求解边缘概率。近似推断方法可以分为两种，一种是变分法，这种方法将推断问题转化为近似优化问题，主要包括平均场推断算法<sup>[55]</sup>、置信度传播算法<sup>[56]</sup>；另一种是随机法，这种方法将推断问题转化为采样问题，主要包括马尔科夫链蒙特卡洛算法 (MCMC)。采样算法计算量大的特点，使得随机方法比较耗时<sup>[57]</sup>，因此变分推断在概率图近似推断中应用的更多，效果也更好。本文将使用平均场推断算法实现未知像素置信

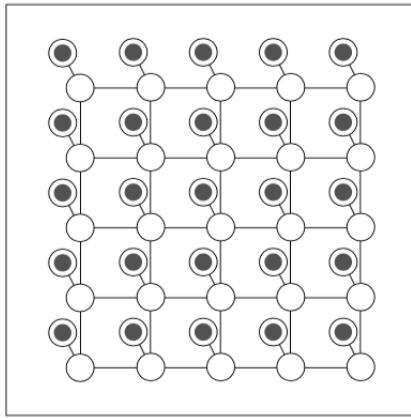


图 3-1 马尔科夫随机场示意图

Fig. 3-1 Illustration of Markov random field model.

度的近似推断。

平均场推断<sup>[55]</sup> (Mean Field Inference) 最早来源于统计物理学中的平均场理论，它利用简单的模型研究复杂的随机系统。它假设系统内的大量个体间存在相互作用，对任意个体而言，其它个体都对其形成了一种近似的平均效应，从而将含有数量巨大相互作用的多体问题转换为每个个体附近弱场内的单体问题。

对于一张二维图像，我们首先构建无向图模型  $G = \{V, E\}$ ，其中图像中所有的像素点构成了顶点集  $V$  的元素，而每两个相邻的顶点构成了图模型中边的集合  $E$ 。这样就建立了如图3-1所示一个马尔科夫随机场 (MRF) 的模型，图中的白色圆圈代表隐含节点，黑色圆圈代表观测值。若以  $\psi_{ii}(x_i)$  代表节点势能 (即隐含节点  $x_i$  与观测值  $y_i$  具有相同状态的置信度)， $\psi_{ij}(x_i, x_j)$  代表边界势能 (即相邻两个隐含节点  $x_i$  和  $x_j$  具有相同状态的置信度)，则在给定观测变量  $y$  的条件下隐含变量  $x$  的条件概率可以表示为：

$$P(x|y) = \frac{1}{Z} \prod_{i,j} \psi_{ij}(x_i, x_j) \prod_i \psi_{ii}(x_i) \quad (3-1)$$

其中  $Z$  代表归一化系数，将条件概率的值限定在  $[0, 1]$  之间。接着我们利用全因子分布  $q(x)$  去近似条件概率  $P(x|y)$  的概率分布，并通过平均场推断的迭代来解出  $q(x)$  中

所有的变分参数。 $q(x)$  具有如下的形式：

$$q(x) = \prod_i b_i(x_i) \quad (3-2)$$

上式中的变分参数  $b_i$  是隐含节点  $x_i$  的边缘概率。*Kullback–Leibler* 散度，又称相对熵，是用来衡量两个概率分布相似程度的指标。通过平均场推断使全因子分布  $q(x)$  近似于条件概率分布  $P(x|y)$  的过程，实际上就是最小化两个分布的 *K-L* 散度的过程，当迭代完成，变分参数  $b_i$  也同时被解出。对于平均场推断，最小化  $P$  与  $q$  之间的 *K-L* 散度等价于在  $\sum_i b_i(x_i) = 1$  的约束条件下求解如下平均场自由能的最小值：

$$F_{MF}(\{b_i\}) = - \sum_{i,j} \sum_{x_i, x_j} b_i(x_i) b_j(x_j) \ln \psi_{ij}(x_i, x_j) + \sum_i \sum_{x_i} b_i(x_i) [\ln b_i(x_i) - \ln \psi_i(x_i)] \quad (3-3)$$

将上式对  $b_i(x_i)$  求导，并设导数为 0，可以求解出关于  $b_i(x_i)$  的表达式：

$$b_i(x_i) = \alpha \psi_{ii}(x_i) \exp \left( \sum_{j \in N_i} \sum_{x_j} b_j(x_j) \log \psi_{ij}(x_i, x_j) \right) \quad (3-4)$$

其中  $\alpha$  是归一化系数， $N_i$  代表节点  $x_i$  所有邻域节点的集合。最后平均场推断算法通过不断迭代，根据上式计算出新的  $b_i$  值，使平均场自由能不断降低。直到算法收敛或达到最大迭代次数，就得到了稳态的变分参数  $b_i$ 。整个的训练过程可以看成一个无向图中的并行信息传递过程，每个节点  $x_i$  都将一条信息  $b_i$  传递给它相邻的节点。而该节点本次传递出去的信息是根据其上一次收到的信息计算而成。而信息  $b_i$  需要在训练开始时进行初始化，在本文中，我们将用户提供的交互信息  $\psi_{ii}(x_i)$  作为信息的初始值，并设  $b_i(x_i) = \psi_{ii}(x_i)$ 。

### 3.3 基于平均场和水平集的交互式分割

本节中我们通过把平均场推断与水平集方法相结合，来实现热点的交互式分割。所提出的交互式分割算法仅需用户通过简单的点击提供关键点的位置信息，就可以方便地完成分割过程。使用户通过较少的输入获得满意的分割结果。作为预处理的步骤，该算

法首先对图像的所有像素点建立无向概率图，得到顶点集合  $V$  和边集合  $E$ 。在得到用户提供的种子点之后，平均场推断算法利用用户输入信息对变分参数  $b_i$  和水平集函数  $\phi$  进行初始化，然后利用 Heaviside 函数对  $\phi$  进行二值化得到  $H(\phi)$ ，计算出每个节点的节点势能和每条边的边界势能，并利用平均场推断得到新的节点置信度。最后水平集算法利用得到的节点置信度不断进化，从而分割出热点。

本文中使用的水平集算法是最经典的全局方法——Chan-Vese 模型<sup>[46]</sup>，它的能量函数可以表示为：

$$\begin{aligned} \mathcal{E}_{CV}(c_1, c_2, C) = & \mu \cdot Length(C) + \nu \cdot Area(inside(C)) \\ & + \lambda_1 \int_{inside(C)} |u(x, y) - c_1|^2 dx dy \\ & + \lambda_2 \int_{outside(C)} |u(x, y) - c_2|^2 dx dy \end{aligned} \quad (3-5)$$

式中  $u(x, y)$  代表图像中对应位置像素的灰度值， $c_1$  和  $c_2$  是分别代表轮廓内部和外部灰度分布的常数， $\lambda_1$  和  $\lambda_2$  是大于 0 的固定参数。除了全局能量拟合项，Chan-Vese 模型的能量函数中还加入了周长项和面积项对水平集函数进行正则化。Heaviside 函数  $H$  和 Dirac 函数  $\delta$  分别定义为：

$$H(\phi) = \begin{cases} 1, & \text{if } \phi \geq 0 \\ 0, & \text{if } \phi < 0 \end{cases}, \quad \delta = \frac{d}{df} H(\phi) \quad (3-6)$$

利用  $H$  和  $\delta$  可以将能量函数重写为：

$$\begin{aligned} E_{CV}(c_1, c_2, \phi) = & \mu \int_{\Omega} \delta(\phi(x, y)) |\nabla \phi(x, y)| dx dy \\ & + \nu \int_{\Omega} H(\phi(x, y)) dx dy \\ & + \lambda_1 \int_{\Omega} |u(x, y) - c_1|^2 H(\phi(x, y)) dx dy \\ & + \lambda_2 \int_{\Omega} |u(x, y) - c_2|^2 (1 - H(\phi(x, y))) dx dy \end{aligned} \quad (3-7)$$

对能量函数利用梯度下降法求解，可以得到如下的梯度流方程，并利用它完成水平

集的进化：

$$\frac{\partial \phi}{\partial t} = \delta_\epsilon(\phi) \left[ \mu \operatorname{div} \left( \frac{\nabla \phi}{|\nabla \phi|} \right) - \nu - \lambda_1(u - c_1)^2 + \lambda_2(u - c_2)^2 \right] \quad (3-8)$$

对于平均场推断，我们需要利用信息  $b_i$  和水平集函数  $\phi$  计算出每个节点的节点势能和每条边的边界势能。为了能够最准确地对图像的局部信息进行表达，本文通过综合考虑用户输入信息和图像信息来完成两种势能的计算。两种信息的系数分别表示为  $\lambda_U$  和  $\lambda_I$ 。我们用  $m$  表示算法当前的分割结果，其中  $m_i \in [0, 1]$ ，如果  $m_i = 1$  则像素  $i$  被标记为前景，如果  $m_i = 0$  则  $i$  为背景像素。并假设  $i$  和  $j$  分别是图像中两个像素， $j$  是  $i$  的相邻像素，则节点  $i$  的节点势能  $\psi_{ii}(x_i)$  可以表示为：

$$\psi_{ii}(x_i) = \exp(-\lambda_I |m_i - H(\phi)_i|) \quad (3-9)$$

如果信息由节点  $i$  向其相邻节点  $j$  传递，则它们构成边的边界势能  $\psi_{ij}(x_i, x_j)$  可以表示为：

$$\psi_{ij}(x_i, x_j) = \exp(-\lambda_U |m_j - H(\phi)_i|) \quad (3-10)$$

本文提出的交互式分割算法可以总结为如下具体的步骤：

1) 根据用户的输入对  $m$  进行初始化，如果  $i$  是用户选中的像素点，则设  $m_i = 1$ ，反之则设  $m_i = 0$ 。同时将水平集函数初始为：

$$\phi_0(x, y) = \begin{cases} c, & \text{if } m_i = 1 \\ -c, & \text{if } m_i = 0 \end{cases} \quad (3-11)$$

2) 通过公式3-9和公式3-10计算图像中所有的节点势能  $\psi_{ii}(x_i)$  和边界势能  $\psi_{ij}(x_i, x_j)$ ，并通过如下公式计算每个节点的置信度：

$$b_i(x_i) \leftarrow \alpha \psi_{ii}(x_i) \exp \left( \sum_{j \in N_i} \sum_{x_j} b_j(x_j) \log \psi_{ij}(x_i, x_j) \right) \quad (3-12)$$

3) 利用阈值  $\theta$  对得到的置信度图进行二值化，并利用  $\phi'$  表示  $b_i > \theta$  的那部分区域。对于区域  $\phi'$  中的像素点，利用公式3-8描述的梯度流方程进行进化：

$$\phi'_{n+1} = \phi'_n + \Delta t \frac{\partial \phi'_n}{\partial t} \quad (3-13)$$

并利用  $\phi'$  对  $\phi$  中满足  $b_i > \theta$  的像素点进行更新。

4) 如果达到了给定的终止条件，则终止迭代。将  $\phi_i > 0$  的像素设置为前景， $\phi_i < 0$  的像素设置为背景。否则将  $H(\phi_i) > 0$  的像素的  $m_i$  值设置为 1，并返回第 2 步继续进行迭代。

终止条件对于算法的分割效果有着重要的影响，如果选择不好会导致分割结果不完整或产生泄漏现象。然而，很难对不同背景条件下的热点找到一个统一且效果好的终止条件。因此，本文针对不同背景条件下的热点提出了两种不同的终止条件。分别是宽松终止条件和紧密终止条件，其中前者适用于热点所在局部区域前景和背景灰度分布差异较大的情况，而后者适用于局部区域前景和背景灰度分布差异较小的情况。

- 宽松终止条件：若迭代次数大于 10 且相邻两次前景区域面积不变，则终止迭代。
- 紧密终止条件：利用 OTSU<sup>[21]</sup> 方法寻找前景区域的最优阈值，并记大于此阈值的区域面积为  $A$ 。若迭代次数大于 10 且  $|A_{n+1} - A_n| = 0$ ，则终止迭代。

由于满足宽松终止条件的热点一般可以通过上一章提出的热点自动分割方法分割出来。因此，我们假定采用交互式方法分割的热点一般是属于比较难分的。并将紧密终止条件设置为默认的终止条件。如果用户认为分割结果不完整，可以点击“继续进化”的选项，并采用宽松终止条件完成分割。

在本文中，交互式分割算法的具体参数设置为： $c = 1, \epsilon = 1, \Delta t = 0.12, \lambda_1, \lambda_2 = 1, \theta = 0.45, \lambda_U, \lambda_I = 0.1$ 。图3-2给出了一个热点交互式分割的具体例子。图中 (a) 是用户给出种子点的位置信息，(b) 是进化过程中的热点轮廓，(c) 是采用紧密终止条件分割出的热点轮廓，(d) 是由 (c) 继续进化并通过宽松终止条件分割出的热点轮廓。在分割的同时，热点的面积、平均灰度、灰度范围和标准差等数据也被同时计算并显示出来，这些量化数据可以为医生的诊断提供帮助。从图中可以看出，本文提出的方法仅通过一个初始点，就可以将热点的轮廓比较完整地分割出来。

图3-3给出了两种不同的终止条件下成功的分割结果。图中的 I 部分是采用宽松终止条件的分割结果，II 部分是采用紧密终止条件的分割结果。从图中可以发现，正如我

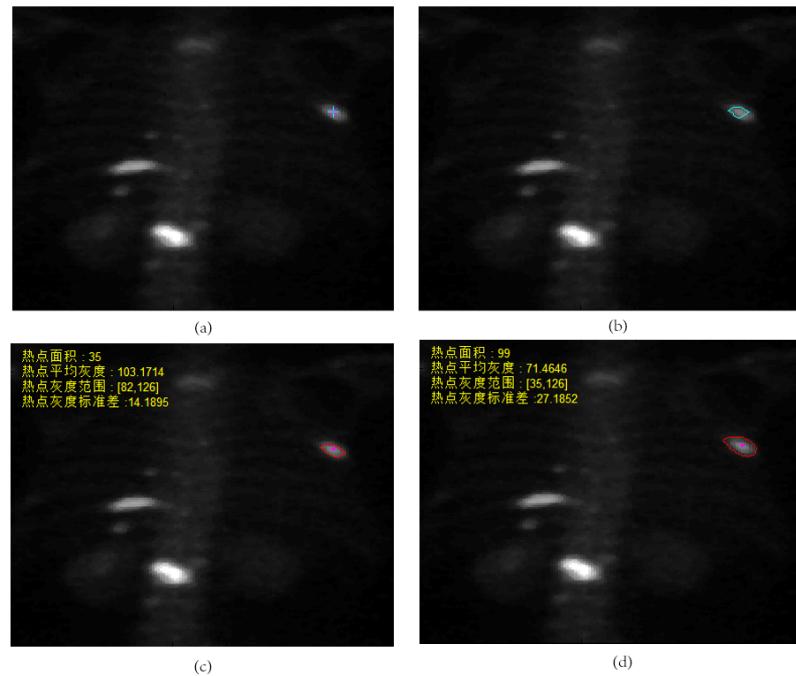


图 3-2 热点交互式分割示意图

Fig. 3-2 Illustration of hotspot interactive segmentation.

们之前的假设一样，宽松终止条件更适合于热点、背景灰度相差较大的情况，而紧密终止条件适合于热点、背景灰度相差较小的情况。

### 3.4 热点轮廓局部修饰

由于上一小节提出的两种方法分割的结果不可能十分完美，所以还需要后续步骤对热点轮廓进行局部修饰。在这一节，我们将利用图割法和阈值法完成热点轮廓的修饰。用户将继续在图像上进行点击，给出要修饰区域的边界关键点  $(bx, by)$ 。对于原轮廓外部的关键点，我们利用图割法对用户给出关键点局部区域的像素进行划分，并将前景像素的置信度设置为 0.5，通过水平集的继续进化完成局部扩张修饰。对于原轮廓内部的关键点，我们利用该点的灰度值进行局部阈值化，实现轮廓的局部缩小。对于距离较近的两个热点，在分割其中一个过程中，很容易将另一个也一起分出来，从而产生多个轮

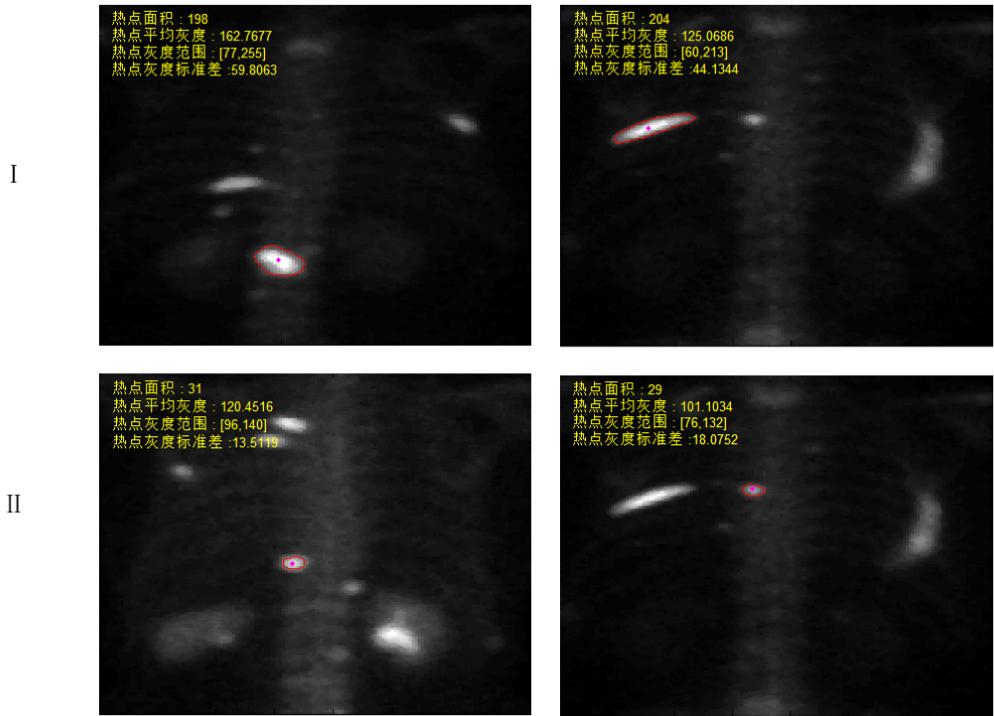


图 3-3 两种条件下的热点分割结果

Fig. 3-3 Segmentation result in two convergence condition.

廓。用户此时可以在需要消除的轮廓内部进行点击，从而实现多余轮廓的消除。

### 3.4.1 图割方法简介

本章我们采用图割方法更新前景像素的置信度，从而实现轮廓的局部外扩。Boykov 等<sup>[58]</sup>于 2001 年提出的图割算法 (Graph-Cut) 是一种非常有用的能量优化算法。它在计算机视觉领域的应用主要集中在图像分割和抠图<sup>[59]</sup>。其本质是把图像分割问题转化为能量的最小化问题，最终目标就是找到使能量函数最小的分割方式。图割方法的原理如图3-4所示。它首先对二维图像构建有向图模型  $G = \{V, E\}$ ，其中  $V$  是顶点的集合， $E$  是有向边的集合。并将用户提供的前景和背景像素点分别当作图模型中的源节点  $s$  和汇聚节点  $t$ 。 $s$  和  $t$  统称为终端节点。所有的未标注节点按照相互之间的位置关系构成了在源节点和汇聚节点之间的二维平面，平面中的节点记为  $x_i$ ，每个节点分别与两个终端节

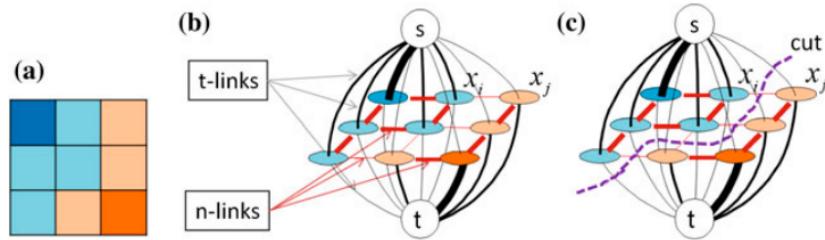


图 3-4 图割方法的示意图

Fig. 3-4 Illustration of graph-cut method.

点相连，且相邻的节点间也相互连接。如图3-4(b)所示，平面节点与终端节点之间的边被称为 **t-link**，而平面节点间连接形成的边被称为 **n-link**。该算法为每个边都分配了相应的代价，并利用代价函数来衡量各种分割方式的有效性，图3-4(c)就给出了一种可能的分割方式。一种分割方法的代价是图模型内所有边代价的总和，因此图割算法的任务就是寻找使代价函数最小的分割方式。根据 Ford-Fulkerson 定理，求图模型的最小割等价于求从源节点流向汇聚节点的最大流，这样就把最小割问题转化成了最大流问题。本文采用 Boykov 等<sup>[60]</sup>于 2004 年提出的 B-K 算法来有效地求解最小割问题。

根据图割算法，图模型的代价函数被定义为：

$$E(L) = \lambda \cdot \sum_{i \in V} D_i(L_i) + \sum_{(i,j) \in E} V_{i,j}(L_i, L_j) \quad (3-14)$$

其中  $L = \{L_i | x_i \in I\}$  代表一种分割方法对每个平面节点的标注。若  $L_i = 0$  则  $x_i$  是一个背景像素，若  $L_i = 1$  则  $x_i$  是一个前景像素。区域代价项  $D_i(L_i)$  代表将像素  $x_i$  标注为  $L_i$  的代价，而边界代价项  $V_{i,j}(L_i, L_j)$  则是对相邻像素  $x_i$  和  $x_j$  性质不一致的惩罚。在本文中，我们参考 Rother 等<sup>[59]</sup>提出的 GrabCut 方法，将区域代价项和边界代价项分别定义为：

$$\begin{aligned} D_I(L_i = 1) &= -\ln Pr(I(i)|F) \\ D_I(L_i = 0) &= -\ln Pr(I(i)|B) \end{aligned} \quad (3-15)$$

$$V_{i,j}(L_i, L_j) = \gamma |I(i) - I(j)| \exp(-\beta(I(i) - I(j))^2) \quad (3-16)$$

其中  $\beta = \left(2 \sum_{i,j} (I(i) - I(j))^2\right)^{-1}$ 。在本文中，我们将  $\gamma$  设置为 1，并通过 B-K 算法求解出区域内像素的最优二元划分。

### 3.4.2 局部修饰的实现

根据用户给出的关键点位置，我们对用户的意图进行判断。如果关键点位于原有轮廓的外部，则说明用户需要轮廓进行局部扩张。如果原轮廓仅包含一条曲线且关键点位于原有轮廓的内部，则说明用户需要轮廓进行局部缩小。如果原轮廓包含多个部分且关键点位于其中某一轮廓的内部，则说明用户需要对该轮廓进行去除。下面我们分别阐述实现这三种局部修饰的方法。

1) 局部扩张 为了实现轮廓局部扩张，本文采用图割方法对关键点局部区域的像素进行二元划分，并通过划分结果对置信度图进行更新，使水平集向关键点方向进化。由于图割方法是一种交互式的分割方法，需要指定源节点和汇聚节点，分别作为已知的前景点和背景点。因此本文提出了一种方法来计算源节点和汇聚节点的坐标位置。图3-5给出了局部扩张修饰的示意图，其中红色曲线代表原有轮廓，红色曲线内的绿色点  $(cx, cy)$  代表用户最初给定的种子点，原有轮廓外的黄色点  $(bx, by)$  代表用户给出的关键点，蓝色区域代表  $9 \times 9$  大小以关键点为中心的局部区域。我们将关键点与种子点进行连线，并沿着这条线，将距离关键点长度为 2 且在原有轮廓外的点  $(ox, oy)$  作为汇聚节点，类似地，将距离关键点长度为 2 且在原有轮廓内的点  $(ix, iy)$  作为源节点。我们假设该连线与水平方向的夹角为  $\alpha$ ，则可以通过下面的式子计算出源节点和汇聚节点的横坐标和纵坐标，进而获得它们的灰度信息：

$$\begin{cases} iy = by - \frac{2(by-cy)}{\sqrt{(by-cy)^2 + (bx-cx)^2}} \\ ix = bx - \frac{2(bx-cx)}{\sqrt{(by-cy)^2 + (bx-cx)^2}} \end{cases} \quad (3-17)$$

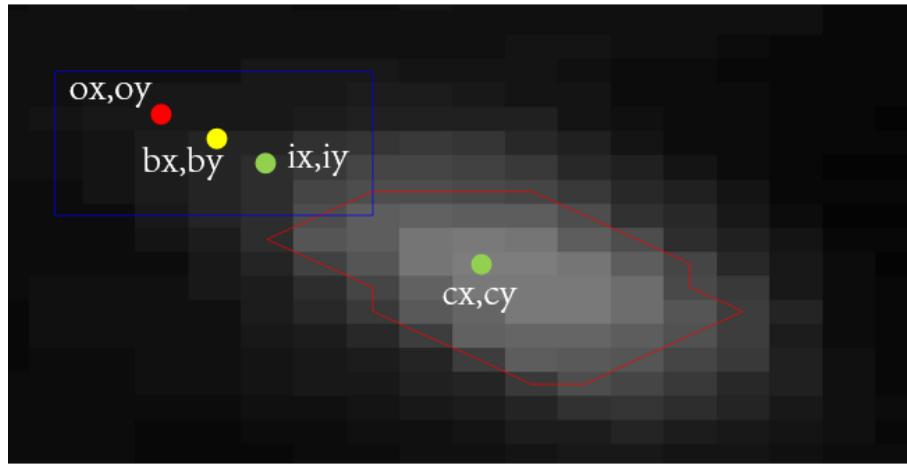


图 3-5 局部扩张修饰示意图

Fig. 3-5 Illustration of local outward modification.

$$\begin{cases} oy = by + \frac{2(by-cy)}{\sqrt{(by-cy)^2+(bx-cx)^2}} \\ ox = bx + \frac{2(bx-cx)}{\sqrt{(by-cy)^2+(bx-cx)^2}} \end{cases} \quad (3-18)$$

接着我们通过得到的源节点和汇聚节点建立图割模型，并利用 B-K 算法解出关键点局部区域像素的最优二元划分，将前景像素的置信度设置为 0.5，从而完成对置信度图的更新，实现轮廓的区域扩张。因为只是对原轮廓进行修饰，所以水平集迭代次数设为较小的值，本文中设为 5 次迭代。图3-6给出了向外扩张修饰的例子，从左到右依次展示了局部扩张修饰分割的步骤。图中的粉色点代表用户最初给定的种子点，黄色点代表用户给出的修饰关键点。从图中我们可以看出，采用图割方法可以有效地完成热点轮廓的局部扩张修饰。

2) 局部缩小 由于近似推断和水平集都有向外扩张的特性，因此无法用它们来完成对原有轮廓的局部缩小修饰。但由于热点本身具有灰度值比背景高的特性，本文采用局部阈值法对热点轮廓进行缩小修饰。该方法将用户给定的关键点的灰度值作为局部阈值，对以关键点为中心半径为  $R$  的区域进行阈值化，并将灰度值小于该阈值的像素排除在轮廓之外，从而获得局部缩小后的热点轮廓。在本文中，我们将局部修饰的区域半

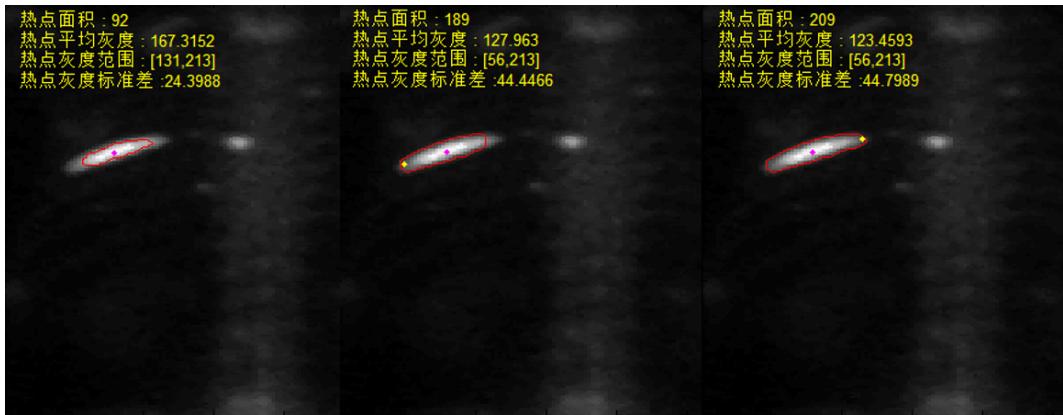


图 3-6 向外扩张修饰

Fig. 3-6 Outward expansion of boundary.

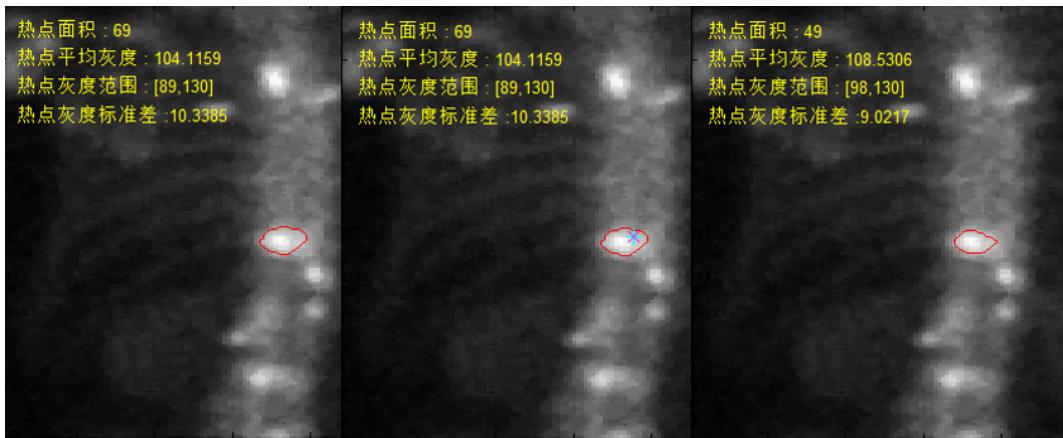


图 3-7 向内收缩修饰

Fig. 3-7 Inward shrink of boundary.

径  $R$  设置为 5。图3-7给出了向内收缩修饰的例子，从左到右依次展示了原有轮廓、用户给出的关键点和修饰后的热点分割结果。从图中我们可以看出，采用局部阈值法可以有效地完成热点轮廓的局部缩小修饰。

3) 多余轮廓去除 为了实现对多余轮廓区域的去除，本文首先将原有轮廓进行二值化，再标记出此二值图像中所有的连通区域。对于与用户给出的关键点具有相同标记的像素，本文将它们的水平集函数值设置为 0，从而完成对多余轮廓的去除。图3-8给出

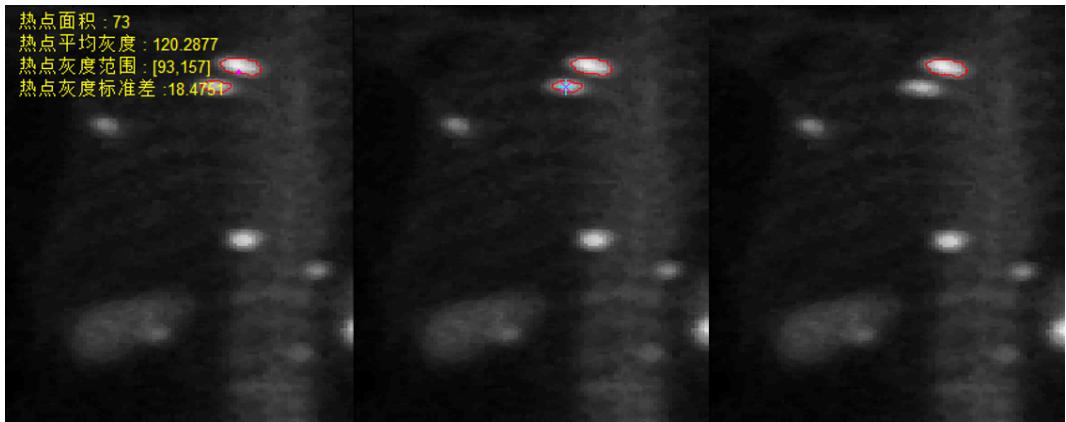


图 3-8 多余区域去除

Fig. 3-8 Removal of extra region.

了多余区域去除的例子，从左到右依次展示了原有轮廓、用户给出的关键点和去除多余轮廓后的热点分割结果。从图中我们可以看出，本文提出的方法可以有效地完成多余热点轮廓的去除。

### 3.5 实验结果

在本节中，我们利用定量实验来证明本章所提出的热点交互式分割方法的有效性。我们首先从 46 幅骨扫描图像中精心挑选了 100 个具有代表性的热点构成了热点数据库，这些热点具有不同的大小、形状和亮度。并邀请专业的医生对热点进行了人工的标注。本文构建的热点数据库如图3-9所示，其中包含了热点的原图和热点的金标准。

接着，我们利用本文提出的热点交互式分割方法对热点数据库中的 100 个热点进行分割。为了保证分割的公正性，我们邀请对本系统完全陌生的志愿者进行分割的操作，并要求除了种子点最多提供两个关键点的信息。对热点数据库的分割结果如图3-10所示，从图中可以看出，该方法可以很好地完成热点分割的任务，分割出的区域和金标准十分接近。

最后，我们利用上一章提到的 Jaccard 指数和 Dice 指数对分割的准确率进行量化的衡量。表3-1给出了本章提出的交互式分割方法的分割准确率，其中 Jaccard 指数超过了

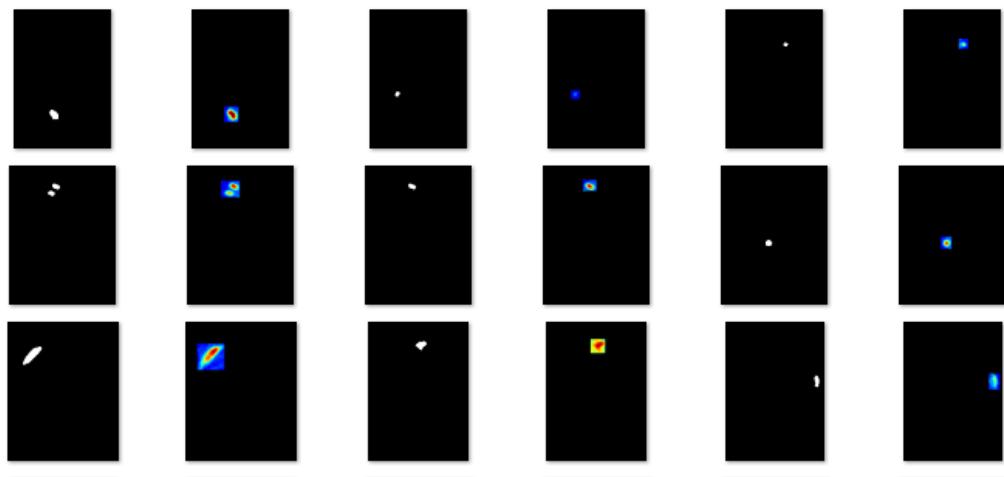


图 3-9 热点数据库

Fig. 3-9 Illustration of hotspot benchmark.

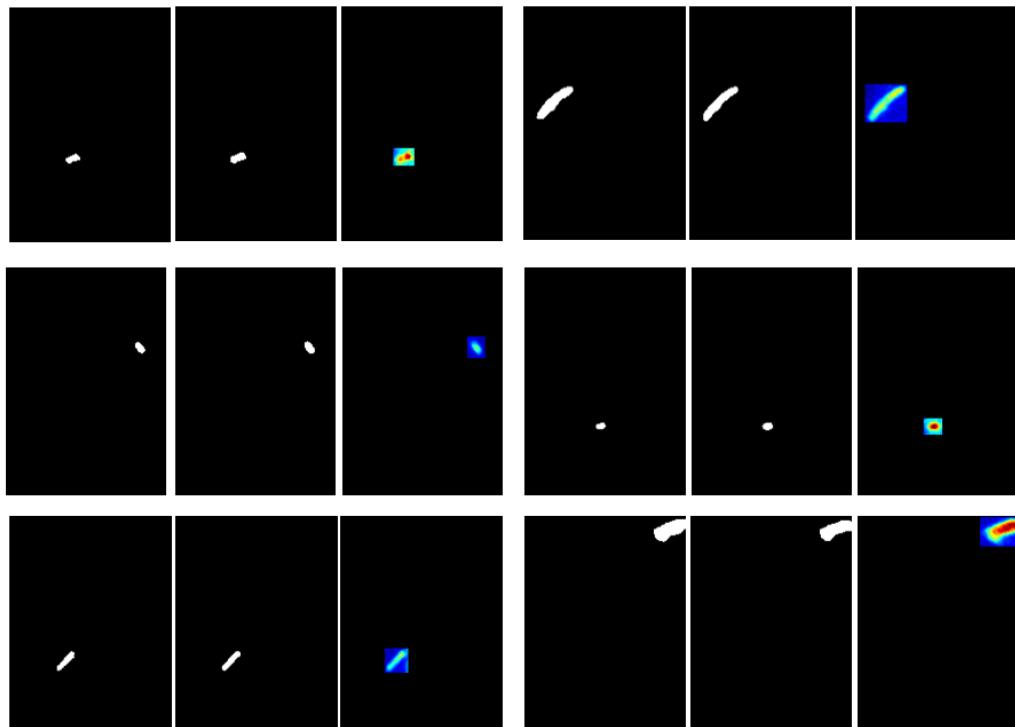


图 3-10 热点数据库分割结果

Fig. 3-10 Segmentation result of hotspot benchmark.

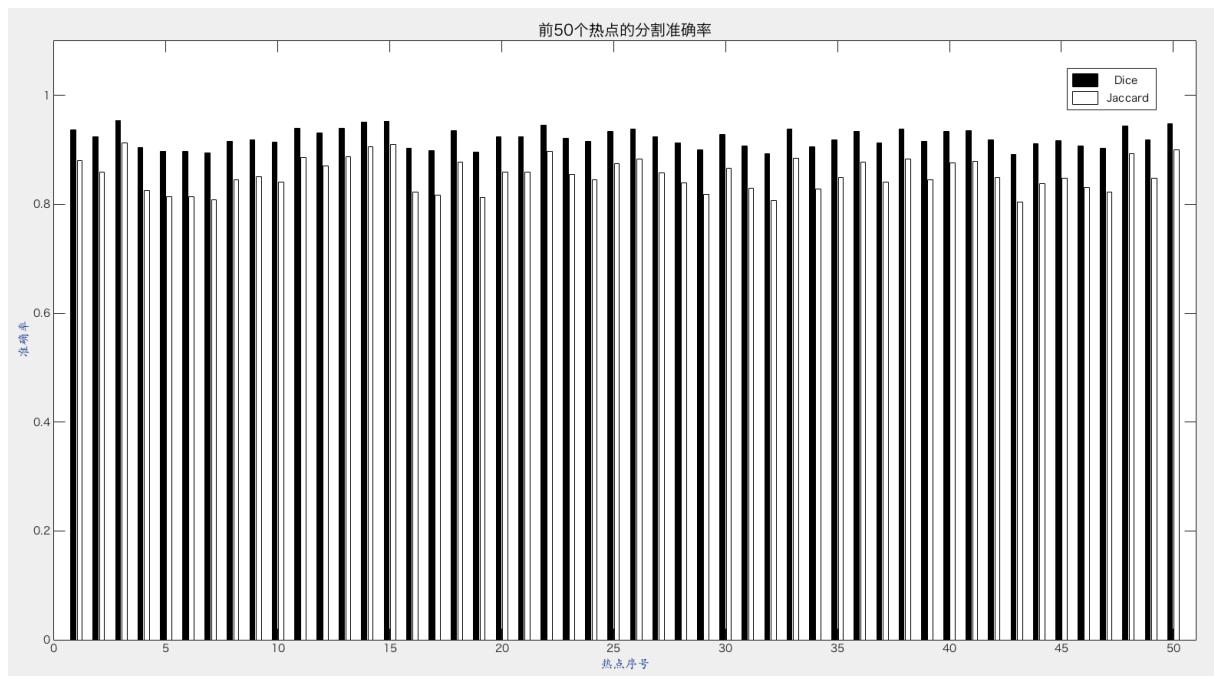


图 3-11 前 50 个热点的分割准确率

Fig. 3-11 Segmentation accuracy of the first 50 hotspots.

0.85，Dice 指数超过了 0.9，这表明该算法具有很高的分割准确率。图3-11给出了前 50 个热点的 Jaccard 和 Dice 指标，从图中可以看出，本文提出的方法对大部分热点都能获得很好的分割效果，且具有很好的稳定性。

表 3-1 所提出方法的分割准确率

Table 3-1 Segmentation accuracy of the proposed methods

指标	Jaccard 指数	Dice 指数
准确率	0.8547	0.9214

通过以上的定量实验，我们证明了本文提出的热点交互式分割方法的有效性。

### 3.6 本章小结

本章提出了一套全新的骨扫描热点交互式分割算法。首先阐述了开发交互式热点分割算法的具体原因，介绍了近似推断中常用的平均场推断法，并利用它结合水平集方法对用户提供的信息进行传递，通过不断迭代更新置信度图和演化水平集曲线，从而完成对热点的分割。对于迭代的终止条件，本章提出了宽松和紧密两种方法，并分别适用于热点、背景灰度分布差异较大和较小的情况。接着本章利用图割算法和局部阈值法完成了对热点分割轮廓的局部修饰，包含局部扩张、局部缩小和多余轮廓去除三种模式。最后，本章构建了包含 100 个热点的热点数据库，并利用 Jaccard 和 Dice 两种定量指标来衡量本章提出的交互式分割算法的分割准确率。本章通过后续的定量实验，证明了提出的热点交互式分割算法的有效性和准确性。



## 第四章 基于深度特征的有监督哈希骨扫描图像检索

本文在之前的两章分别提出了热点的自动分割和交互式分割算法，医生利用这两种算法可以将骨扫描图像中的热点准确地分割出来。然而在医生进行诊断的过程中，也经常会碰到不确定的病例，这时如果能为医生提供过往病例中相似的图像作为参考，就可以有效提升诊断的成功率。医院在平时的临床诊断和治疗过程中积累了大量的病例，这些病例构成的数据库中包含有病人的医学影像、个人信息以及诊断结果，但这些有价值的信息却没有得到很好的利用。因此，我们可以通过建立医学图像检索系统，帮助医生快速地从数据库中找到所需的信息，从而为医生的诊断提供有价值的参考信息。本章构建了全新的基于内容的骨扫描图像检索系统。为了同时保证检索的准确性和快速性，我们采用了基于核函数的有监督哈希方法建立索引。由于传统的 SIFT 词袋模型对于骨扫描图像的描述效果一般，我们将其替换为第二章中通过卷积稀疏编码提取出的深度特征。利用该特征和哈希方法，我们构建了一套骨扫描图像检索系统，并取得了较好的检索效果。

### 4.1 哈希索引技术

为了构建基于内容的图像检索系统，首先要对数据库中图像的抽取区分度高的特征，再选取合适的检索方法从数据库中检索出与输入图像内容上最相似的图像。检索技术对于系统的性能至关重要，因此也一直是基于内容图像检索的研究重点。要实现对数据库的检索，最简单直接的方法就是线性搜索，即将查询图像的特征与数据库中每一张图像的特征进行一一比较，并返回特征最相似的几张图像。这种方法虽然准确度高，但却需要很大的计算量，实时性不高，尤其对于大规模的数据库，完成检索需要耗费大量的时间，基本不具有实用性。Nister 和 Stewenius<sup>[61]</sup> 在 2006 年提出了词汇树（Vocabulary tree）方法来实现对大量图像的检索。该方法通过对图像的关键点特征进行分层聚类来

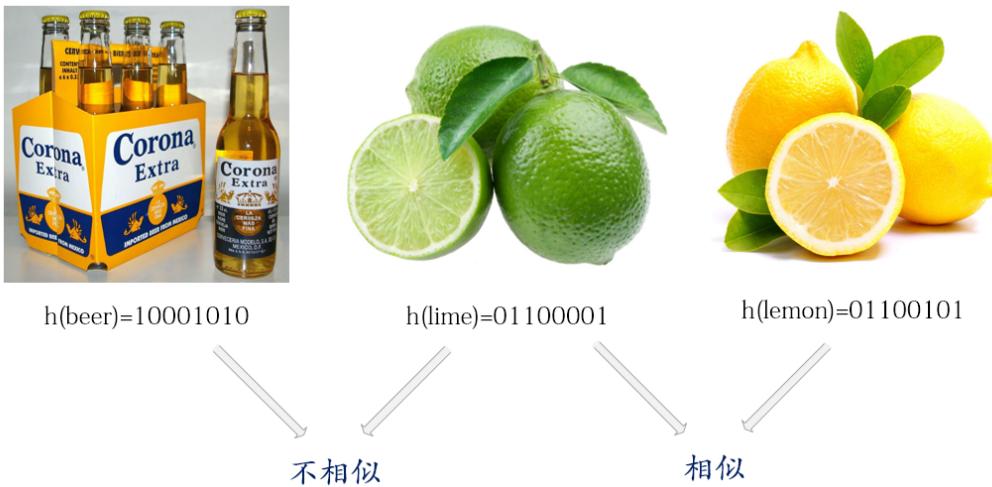


图 4-1 哈希方法示意图

Fig. 4-1 Illustration of hashing methods.

生成词汇树，并在每个节点处关联一个倒排文件。由于采用了树型结构和词汇倒排索引，词汇树方法具有很好的扩展性。同时在查询时，可以不用遍历所有词汇，因此具有很高的检索效率。虽然该方法节约了时间上的开销，却需要耗费较高的储存空间。比如建立一个 6 层，聚类 K 值为 10，采用 128 维 SIFT 特征的词汇树，就需要占用 143MB 的储存空间。

哈希检索技术恰恰可以解决这一问题，它通过一组哈希函数  $h : \mathbb{R}^d \mapsto \{1, 0\}^1$ <sup>1</sup> 将高维图像特征映射成 Hamming 空间内的低维哈希码，从而可以有效地节省储存空间。例如，8000 万张  $32 \times 32$  大小的图片需要消耗 600GB 的储存空间，但如果采用 64 位的哈希二进制码，就只需 600MB。同时哈希方法还利用近似最近邻 (ANNs) 搜索的思想，将搜索空间降低到较小的可接受范围，因此具有很高的检索效率。一个好的哈希算法能够保证原始特征空间距离相近的样本，在映射后的 Hamming 空间内依旧相近。即样本越相似，得到的哈希码也越有可能一样。图4-1给出了哈希方法的示意图，如图中所示，外表相似的图像具有相似的哈希码，而外表不相似的图像哈希码相差就较大。

局部敏感哈希 (LSH)<sup>[62]</sup> 是最经典的一种哈希算法，它按照一定的概率条件采用随

机方法生成哈希函数。由于采用了不依赖样本数据的随机方法，因此该算法的检索准确率不是很高。为了提高检索准确率，学者们提出了很多数据依赖型的算法。半监督哈希 (SSH)<sup>[63]</sup> 是其中有代表性的一种，它利用样本的标注信息，在最小化经验误差的同时最大化生成的未标注样本哈希码的信息熵，并通过较短的哈希码实现了优良的性能。数据依赖型哈希算法一般通过解复杂的优化问题生成哈希函数，因此训练时间较长，效率不高，限制了它们在大规模数据库上的使用。为了解决这一问题，Liu 等<sup>[64]</sup> 提出了核函数有监督哈希 (KSH)，它将 Hamming 距离的优化问题转化为编码内积的优化问题，从而降低了问题的复杂度并提升了优化的效率。同时，该算法通过引入核函数解决了样本的线性不可分问题，取得了很好的检索准确率。

## 4.2 核函数有监督哈希算法

核函数有监督哈希<sup>[64]</sup> 算法可以快速地完成哈希函数的生成，而且通过较短的哈希码就能达到足够的准确率，因此被很多医学图像检索系统<sup>[13,14]</sup> 所采用。本章构建的骨扫描图像检索系统也将采用核函数有监督哈希来实现。

给定数据库中图像的特征集合  $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\} \in \mathbb{R}^d$ ，并设未标注图像的  $d$  维特征向量为  $\mathbf{x} \in \mathbb{R}^d$ ，哈希函数的集合为  $H = \{h_1, \dots, h_r\}$ ，计算出的  $r$  位哈希码为  $\mathbf{y} = \{y_1, \dots, y_r\}$ 。则每位的哈希码可以表示为：

$$y_k = \text{sgn}(h_k(\mathbf{x})), \quad k \in \{1, \dots, r\} \quad (4-1)$$

针对样本的线性不可分问题，可以通过引入核函数来解决。引入核函数还有另一个优点，就是通过选择较少的锚点，降低需要学习的参数的数目。**KSH** 算法从标注样本中抽取  $m$  个锚点  $\{\mathbf{x}_{(1)}, \dots, \mathbf{x}_{(m)}\}$  来计算出所需的核函数，其中  $m \ll n$ 。这里将核函数表示为  $\kappa : \mathbb{R}^d \times \mathbb{R}^d \mapsto \mathbb{R}$ ，并将引入核函数后的哈希函数定义为：

$$h_k(\mathbf{x}) = \sum_{j=1}^m \kappa(\mathbf{x}_{(j)}, \mathbf{x}) a_j - b \quad (4-2)$$

上式中的  $a_j$  和  $b$  都是待定的系数。为了使生成的哈希码具有尽量大的信息量，其中的每一位哈希码  $y_k$  都应满足  $\sum_{i=1}^n y_k(\mathbf{x}_i) = 0$ ，因此将  $b$  设为：

$$b = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^m \kappa(\mathbf{x}_{(j)}, \mathbf{x}_i) a_j \quad (4-3)$$

将上式带入式4-2，可以得到：

$$\begin{aligned} h_k(\mathbf{x}) &= \sum_{j=1}^m \left( \kappa(\mathbf{x}_{(j)}, \mathbf{x}) - \frac{1}{n} \sum_{i=1}^n \kappa(\mathbf{x}_{(j)}, \mathbf{x}_i) \right) a_j \\ &= \mathbf{a}^\top \bar{k}(\mathbf{x}) \end{aligned} \quad (4-4)$$

上式中  $\bar{k}(\mathbf{x}) = [\kappa(\mathbf{x}_1, \mathbf{x}) - \mu_1, \dots, \kappa(\mathbf{x}_m, \mathbf{x}) - \mu_m]^\top$ ，是一个从  $\mathbb{R}^d$  到  $\mathbb{R}^m$  的向量映射，其中  $\mu_j = \sum_{i=1}^n \kappa(\mathbf{x}_{(j)}, \mathbf{x}_i)/n$ 。 $\mathbf{a}$  是核函数的系数，且  $\mathbf{a} = [a_1, \dots, a_m]$ ，它的选取决定了最终的哈希函数。因此，需要依赖标注的数据有监督地解出最优的  $\mathbf{a}$ 。为了建立监督信息矩阵  $S \in \mathbb{R}^{l \times l}$ ，从样本集中选取  $l$  个样本，如果其中任意两个样本  $i$  和  $j$  相似，则将  $S(i, j)$  设为 1，如果不相似则设为 -1。由于直接最小化 Hamming 距离  $\mathcal{D}_y(\mathbf{x}_i, \mathbf{x}_i)$  比较困难，故将其转换为优化编码内积的问题。两个编码  $code_r(\mathbf{x}_i)$  和  $code_r(\mathbf{x}_j)$  的内积可以表示为：

$$\begin{aligned} &code_r(\mathbf{x}_i) \circ code_r(\mathbf{x}_j) \\ &= |\{k | y_k(\mathbf{x}_i) = y_k(\mathbf{x}_j), 1 \leq k \leq r\}| \\ &\quad - |\{k | y_k(\mathbf{x}_i) \neq y_k(\mathbf{x}_j), 1 \leq k \leq r\}| \\ &= r - 2 |\{k | y_k(\mathbf{x}_i) \neq y_k(\mathbf{x}_j), 1 \leq k \leq r\}| \\ &= r - 2\mathcal{D}_y(\mathbf{x}_i, \mathbf{x}_j) \end{aligned} \quad (4-5)$$

因此，问题等价于求解编码内积的最大值，优化的目标函数定义为如下的最小二乘形式：

$$\min_{H_l \in \{1, -1\}^{l \times r}} \mathcal{Q} = \left\| \frac{1}{r} H_l H_l^T - S \right\|_F^2 \quad (4-6)$$

其中  $H_l = sgn(\bar{K}_l A) \in \mathbb{R}^{l \times r}$ , 代表  $l$  个监督样本的  $r$  位汉明编码。 $A = [\mathbf{a}_1, \dots, \mathbf{a}_r] \in \mathbb{R}^{m \times r}$ , 代表模型参数矩阵。 $\bar{K}_l = [\bar{k}(\mathbf{x}_1), \dots, \bar{k}(\mathbf{x}_l)]^\top \in \mathbb{R}^{l \times m}$ , 代表核函数矩阵。因此, 可以将式4-6转化为:

$$\min_{A \in \mathbb{R}^{m \times r}} \mathcal{Q}(A) = \left\| \frac{1}{r} sgn(\bar{K}_l A) sgn(\bar{K}_l A)^T - S \right\|_F^2 \quad (4-7)$$

最后, 利用贪心算法逐位地求解出每个哈希函数  $h_i$  的核函数系数  $\mathbf{a}_i = [a_{i1}, \dots, a_{im}]$ 。

## 4.3 骨扫描图像检索系统

在本节, 我们分别利用传统的 SIFT 词袋特征和卷积稀疏编码提取的深度特征, 结合核函数有监督哈希算法, 实现一套骨扫描图像检索系统。

### 4.3.1 SIFT 和词袋特征提取

SIFT 特征<sup>[65]</sup>, 又称为尺度不变性特征变换, 是一种很流行的局部特征。它对于图像中目标的旋转、尺度和亮度具有很强的不变性, 对于不同目标的区分度好, 且对于图像噪声、视角变化等也有很高的容忍度。因此在目标识别、图像配准、影像跟踪等计算机视觉领域被广泛使用。SIFT 算法对局部特征的提取主要分为以下几个步骤。首先通过高斯查分函数 (DoG) 对图像进行搜索, 找到极值点作为潜在的关键点。然后对检测到的所有潜在关键点进行筛选, 确定真正的关键点的位置和尺度。接着根据图像局部的梯度方向, 给每个关键点分配一个或多个主方向。最后对关键点邻域内的像素梯度方向进行直方图统计, 形成对关键点的特征描述。如果将关键点邻域分成  $4 \times 4$  的窗口, 并计算 8 个方向的梯度信息, 就可以得到最常用的 128 维 SIFT 关键点特征。在本文中, 我们采用 VLFeat<sup>[66]</sup> 工具箱实现 SIFT 特征的提取。

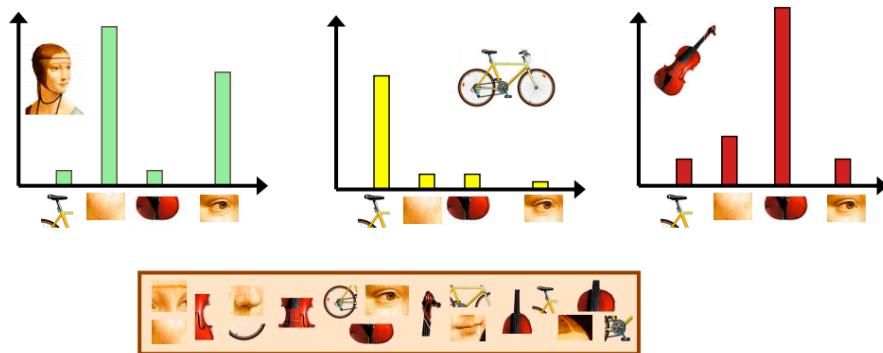


图 4-2 词袋方法示意图

Fig. 4-2 Illustration of bag of visual words method.

词袋模型 (Bag of Visual Words) 是一种特征的描述方法，它最早来源于文本信息检索领域，在计算机视觉领域也有着广泛的应用。如图4-2所示，它对所有通过 SIFT 提取出的关键点特征进行 K-means 聚类，并将得到的 K 个中心构成一个视觉词典，每个中心都代表视觉词典中的一个词汇。对于一张图像，词袋模型对其中含有的视觉词汇进行直方图统计，从而构成一个 K 维的向量，并用这个向量来作为该图像的特征。

本章中，我们利用 SIFT 提取 128 维的关键点特征，并通过词袋模型对每张图像的特征进行抽取，其中聚类参数 K 设置为 64。

### 4.3.2 深度特征提取和骨扫描图像检索的实现

我们按照第二章中提到的方法通过卷积稀疏编码训练出 400 个基底，并利用它们对骨扫描局部图像提取深度特征。为了能够构建骨扫描图像检索系统，我们需要首先建立一个骨扫描局部图像的数据库，并标注关于局部图像的相关信息，例如：病人信息、诊断结果、在原图中位置等。为此我们选择了 628 张异常胸廓骨扫描图像和 786 张正常胸廓骨扫描图像，并从所有图像中位置随机地裁剪  $49 \times 49$  大小的局部图像，一共抽取了 8091 张局部图像。其中裁剪自正常图像的被自动标注为不含热点，而对于来自异常图像的局部图像，我们利用第二章中提出的图片级别诊断分类器进行标注。最后一共有 3768 张局部图像被标注为含有热点，而不含热点的局部图像则有 4323 张。由于 DICOM

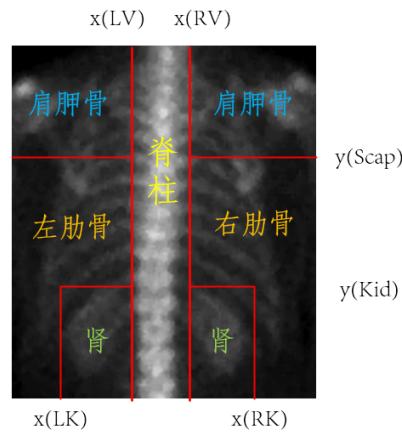


图 4-3 胸廓骨扫描图像分区示意图

Fig. 4-3 Image partition of thoracic bone scan images.

格式的原始骨扫描图像中含有病人的相关信息，因此我们直接将这些信息标注给从中剪裁出去的局部图像。对于局部图像在原图中的位置，我们采用如图4-3所示的方式对图像进行分区，并根据局部图像中心点在原图中的所在区域完成标注。为了实现分区，我们对大量胸廓骨扫描图像进行了统计，并结合解剖学知识，标定了 6 个关键的位置信息。设图像的水平长度和竖直长度分别  $M$  和  $N$ ，水平方向为  $x$  轴，竖直方向为  $y$  轴，左上角为原点。则这六个位置的坐标值可以表示为：

$$\begin{cases} x(LV) = 0.357 \times M, \\ x(RV) = 0.649 \times M, \\ y(Scap) = 0.239 \times N, \\ y(Kid) = 0.722 \times N, \\ x(LK) = 0.240 \times M, \\ x(RK) = 0.766 \times M \end{cases} \quad (4-8)$$

在完成了骨扫描局部图像数据库的建立后，我们分别利用 SIFT 词袋方法和卷积稀疏编码方法对其中的图像进行特征抽取。并将所有的样本划分为训练集和测试集。最

index	from	imname	pathname	position	ID	Name	Sex	Size	Weight
1494	'P547.bmp'	'1494_p.b...	'C:\Users\jey...	'Right Rib'	'08-1486'	'MEN RO...'	'M'	1.7196	56.0000
4216	'N233.bmp'	'448_n.bmp'	'C:\Users\jey...	'Right Rib'	'08-2570'	'huang ju'	'F'	1.6510	55
5839	'N720.bmp'	'2071_n.b...	'C:\Users\jey...	'Right Rib'	'05-1558'	'DONG MI...	'M'	1.7805	83
4855	'N425.bmp'	'1087_n.b...	'C:\Users\jey...	'Right Rib'	'08-3032'	'feng xian ...'	'M'	1.6993	60
3462	'P389.bmp'	'3462_p.b...	'C:\Users\jey...	'Scapula'	'08-0924'	'ZENG CH...	'M'	1.7501	84

图 4-4 检索出图像的信息

Fig. 4-4 Information from retrieved images.

后，我们利用核函数有监督哈希训练出 48 个哈希函数，并利用这些函数对数据库中的图像提取哈希码，从而搭建出一套骨扫描图像的检索系统。在利用检索图像对数据库进行检索时，图像检索系统首先对检索图像进行特征提取，接着将提取的特征利用哈希函数映射为哈希码，最后从数据库中选出和检索图像哈希码 Hamming 距离最小的 K 个图像，作为最终的结果图像。在检索出结果图像的同时，结果图像所包含的相关信息也可以一并被提取出来，如图4-4所示。这些信息可以为医生的诊断提供必要的帮助。

## 4.4 实验结果

在本节中，我们利用定量实验来证明本章所提出的骨扫描图像检索系统的有效性，并对根据两种不同特征构建的检索系统的性能进行了比较。我们从上一节建立的骨扫描局部图像数据库中随机选择了 428 张图像构成测试集。并利用两个检索系统——CSA+KSH 以及 BoW+KSH，从数据库中取出和检索图像外表最相似的 K 张结果图像。图4-5给出了两种方式检索结果的对比，每个子图中的第一个是查询图像，其余的 K 张则是结果图像。其中红色实线代表 BoW+KSH，蓝色虚线代表 CSA+KSH。从图中可以看出，采用深度特征的 CSA+KSH 具有更好的效果，而采用 SIFT 词袋特征的 BoW+KSH 没有能够找出相似的图像。

图4-6给出了更多 CSA+KSH 方法的检索结果，从图中可以看出，深度特征确实能够很好地描述骨扫描图像的细节，而根据它构建的检索系统，也能够从数据库中找到外表相似的结果图像。

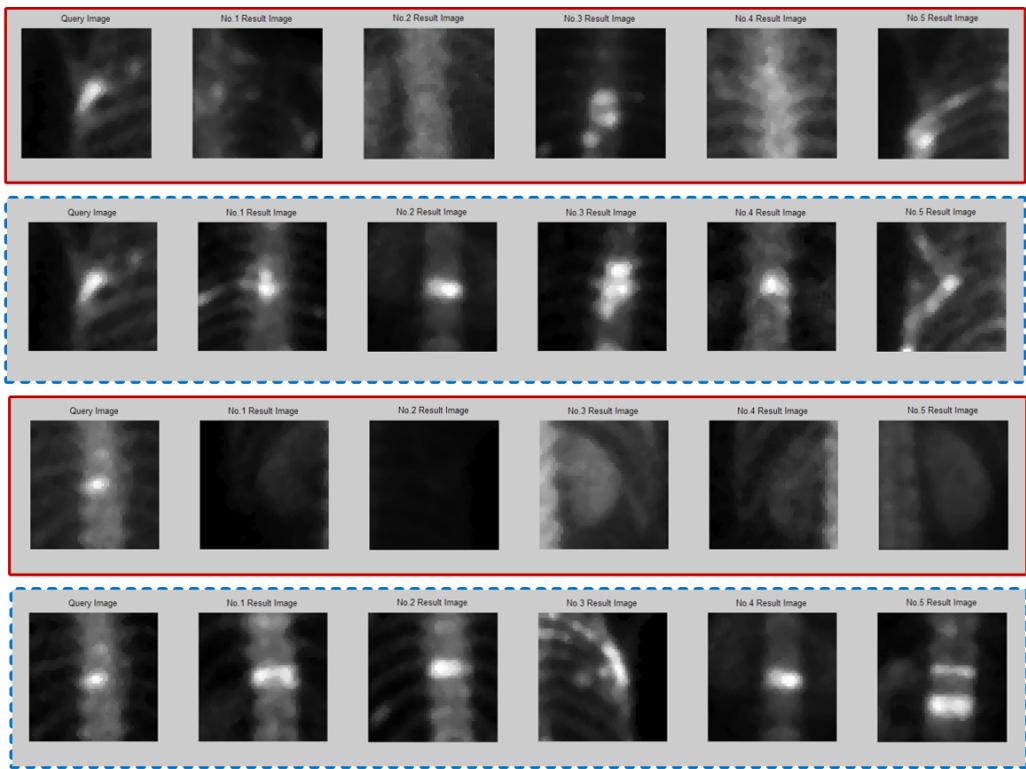


图 4-5 两种方法检索结果对比

Fig. 4-5 Comparison of two retrieval methods.

最后，我们利用诊断准确率和位置准确率两个指标对两个方法的检索效果进行比较。如果在结果图像中某类诊断标签最多，则将那一类的标签作为的检索系统给出的诊断结果。同样地，将类别标签最多的位置作为检索系统判定的位置结果。我们对 CSA+KSH 和 BoW+KSH 两种方法，分别测试了在  $K=3,5,11$  情况下的诊断准确率和位置准确率。具体的结果如表4-1所示，可以看出， $K=5$  时的 CSA+KSH 方法具有最高的诊断准确率，而  $K=11$  时的 CSA+KSH 方法具有最高的位置准确率。该结果说明了本章采用的深度特征结合核函数有监督哈希构建骨扫描图像检索系统的方法，能够从数据库中找出与检索图像外表相似的图像，并具有较好的诊断准确率和位置准确率。

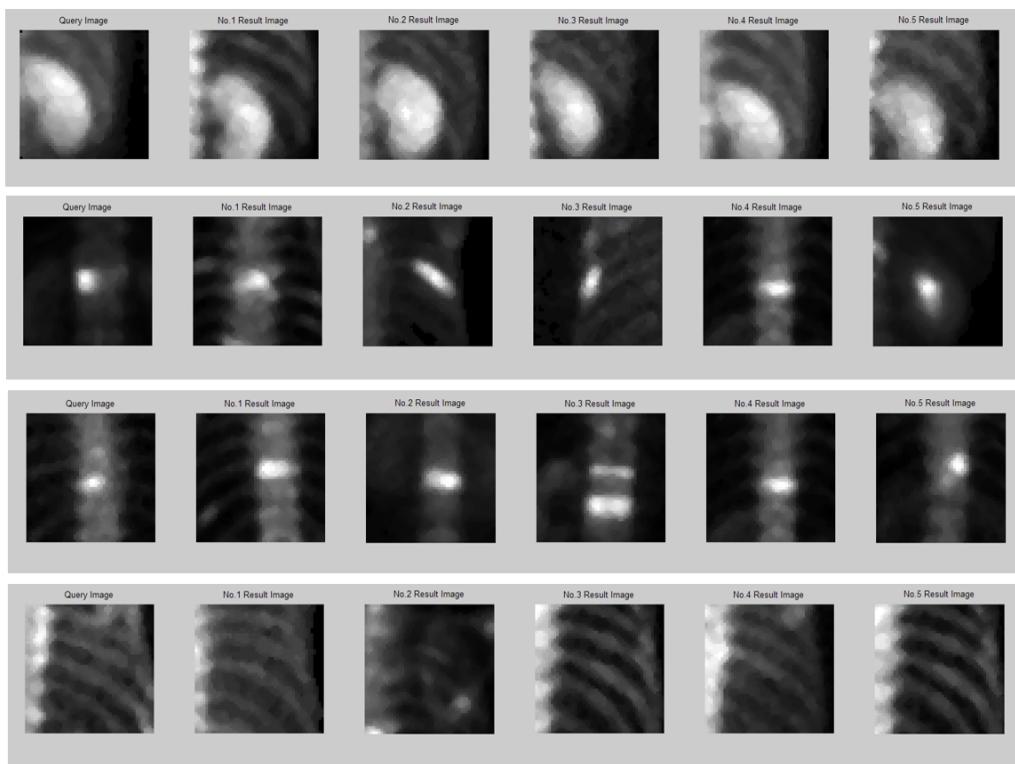


图 4-6 采用深度特征的检索结果

Fig. 4-6 Retrieval result using deep features.

表 4-1 两种方法检索结果比较

Table 4-1 Comparison of two retrieval methods

方法	检索数目 (K)	诊断准确率	位置准确率
CSA+KSH	3	0.7266	0.6729
	5	0.8178	0.6916
	11	0.7430	0.7126
BoW+KSH	3	0.5678	0.4860
	5	0.5724	0.5350
	11	0.6145	0.5187

## 4.5 本章小结

本章实现了一套全新的骨扫描图像检索系统。首先阐述了开发骨扫描图像检索系统的具体背景。之后介绍了图像检索中常用的哈希方法，以及本章所采用的核函数有监督哈希算法。接着构建了一套骨扫描局部图像数据库，并对其中的图像进行了标注。最后，分别用 SIFT 词袋特征、卷积稀疏编码特征与核函数有监督哈希相结合，设计了两种构建图像检索系统的方法。本章通过后续的定性实验，证明了深度特征对于描述骨扫描图像细节的能力。同时通过定量实验，证明了 CSA+KSH 方法的有效性和准确性。



## 第五章 总结与展望

### 5.1 主要工作

本文主要的工作是研究骨扫描图像的异常检测算法以及对检测出的异常区域进行定量分析的方法。为了解决医生在骨扫描诊断时的实际需求，本文根据骨扫描图像本身的特点和国内外相关领域现有的研究成果，从三个方面入手，分别提出了骨扫描图像的热点全自动检测和分割算法、热点交互式分割算法以及骨扫描图像检索系统。本文首先搭建了热点全自动分割的框架：使用卷积稀疏编码对骨扫描图像提取深度特征，并通过 AdaBoost 算法训练出图像级别的诊断分类器，实现骨扫描图像的初步诊断。之后，利用多示例学习算法训练区域级别的热点分类器，进而构建出热点的概率图，再通过局部符号水平集方法实现热点的全自动分割。后续的定量实验证明了本文提出的方法能够得到令人满意的分割结果。接着，针对医生的定量分析需求，本文提出了一种基于平均场推断和水平集的热点交互式分割算法，帮助医生通过简单的交互实现热点的分割和轮廓修饰，并获得热点的定量数据。后续的定量实验证明利用本文提出的交互式分割算法能够方便而准确地把热点从原图中分割出来。最后，本文搭建了基于深度特征和有监督哈希算法的骨扫描图像检索系统，它可以帮助医生在最快的时间内从数据库中找到相关的病例，并为医生的诊断提供必要的参考。

第一章阐述了本文研究工作的具体背景，介绍了国内外相关研究的现状，并给出了全文的整体框架。

第二章首先提出了基于卷积稀疏编码的图像级别诊断方法，接着通过多示例学习算法训练了区域级别的热点分类器，最后利用水平集方法实现热点的分割，并构成了一套热点的自动分割框架。

第三章首先介绍了概率图模型中的近似推断方法，接着基于平均场推断提出了一套

热点的交互式分割算法。最后，利用图割方法和局部阈值法实现了热点轮廓的后续局部修饰。

第四章介绍了哈希算法在图像检索中的应用，并基于核函数有监督哈希算法实现了一套全新的骨扫描图像检索系统，同时对于不同的特征提取方法进行了比较。

## 5.2 后续研究

随着图像处理和机器学习技术的进步，数字化医疗得到了快速的发展，针对各种医学影像的计算机辅助诊断系统被相继提出。由于它可以减轻人工读图的工作量，提高诊断的灵敏度和准确率，因此在临幊上得到了广泛的应用。同样地，骨扫描图像计算机辅助诊断系统对于医生的诊断也可以起到很大的帮助。它能够完成图像中热点的检测、分割和检索，并为医生提供定量分析的工具。本文针对骨扫描图像辅助诊断这一课题，提出了一系列的检测、分割、检索方法，取得了一定的成果。但是，由于骨扫描图像本身的低信噪比和弱边界特点，本文提出的方法在其中一些图像上，没能达到预期的效果。因此，本文所提出的算法仍有较大的提升空间。本文作者认为，可以继续从以下几个方向对本文提出的算法进行改进：

1. 本文在第二章通过卷积稀疏编码提取了骨扫描图像的特征，虽然该特征能够在一定程度上描述骨扫描图像的细节，但却并非是高层的视觉特征。因此在后续的研究中，可以采用多层的深度神经网络，通过训练提取出高层的特征表达，从而实现对骨扫描图像细节更精准的描述。同时，可将提取出的高层特征用在第四章实现的骨扫描图像检索系统中，帮助提高检索的准确率。
2. 本文在第三章提出了热点交互式分割的两种终止条件。然而，在实际应用中，需要用户去选择使用哪种条件去终止分割进程，这会带来很大的不便。因此在后续的研究中，可以通过一定的条件判断方法，将两种终止条件合并，从而使分割的过程更加顺畅。
3. 本文在第四章实现了一套骨扫描图像检索系统，其中使用卷积稀疏编码来描述骨扫描图像的特征。然而，该系统仅仅采用了图像内容信息作为特征，却没有很好地利用

图像的标注信息。因此在后续的研究中，可以将标注信息和内容特征有机的结合，形成多模态特征并应用到检索系统中，从而提高检索的准确率。



## 参考文献

- [1] 王强. 骨扫描图像分割及配准算法研究 [mathesis], **2012**.
- [2] M. Sadik *et al.* “A new computer-based decision-support system for the interpretation of bone scans”. *Nuclear medicine communications*, **2006**, 27(5): 417–423.
- [3] T.-K. Yin and N.-T. Chiu. “A computer-aided diagnosis for locating abnormalities in bone scintigraphy by a fuzzy system with a three-step minimization approach”. *Medical Imaging, IEEE Transactions on*, **2004**, 23(5): 639–654.
- [4] J.-Y. Huang, P.-F. Kao and Y.-S. Chen. “A set of image processing algorithms for computer-aided diagnosis in nuclear medicine whole body bone scan images”. *Nuclear Science, IEEE Transactions on*, **2007**, 54(3): 514–522.
- [5] M. Sadik *et al.* “Computer-assisted interpretation of planar whole-body bone scans”. *Journal of Nuclear Medicine*, **2008**, 49(12): 1958–1965.
- [6] M. Ohlsson *et al.* “Automated decision support for bone scintigraphy”. In: *Computer-Based Medical Systems, 2009. CBMS 2009. 22nd IEEE International Symposium on*, **2009**: 1–6.
- [7] D. Ulmert *et al.* “A novel automated platform for quantifying the extent of skeletal tumour involvement in prostate cancer patients using the Bone Scan Index”. *European urology*, **2012**, 62(1): 78–84.
- [8] Q. Chang *et al.* “Adaptive detection of hotspots in thoracic spine from bone scintigraphy”. In: *Neural Information Processing*, **2011**: 257–264.
- [9] Q. Wang *et al.* “Knowledge-based segmentation of spine and ribs from bone scintigraphy”. In: *Neural Information Processing*, **2011**: 241–248.

- [10] S. Hojjatoleslami and J. Kittler. “*Region growing: a new approach*”. *IEEE Transactions on Image processing*, **1998**, 7(7): 1079–1084.
- [11] K. Simonyan, A. Zisserman and A. Criminisi. “*Immediate structured visual search for medical images*”. In: *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2011*. Springer, **2011**: 288–296.
- [12] X. Zhang, W. Liu and S. Zhang. “*Mining histopathological images via hashing-based scalable image retrieval*”. In: *Biomedical Imaging (ISBI), 2014 IEEE 11th International Symposium on*, **2014**: 1111–1114.
- [13] X. Zhang *et al.* “*Towards large-scale histopathological image analysis: Hashing-based image retrieval*”. *Medical Imaging, IEEE Transactions on*, **2015**, 34(2): 496–506.
- [14] X. Zhang *et al.* “*Fine-grained histopathological image analysis via robust segmentation and large-scale retrieval*”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, **2015**: 5361–5368.
- [15] M. Jiang *et al.* “*Computer-aided diagnosis of mammographic masses using vocabulary tree-based image retrieval*”. In: *Biomedical Imaging (ISBI), 2014 IEEE 11th International Symposium on*, **2014**: 1123–1126.
- [16] M. Jiang *et al.* “*Computer-Aided Diagnosis of Mammographic Masses Using Scalable Image Retrieval*”. *Biomedical Engineering, IEEE Transactions on*, **2015**, 62(2): 783–792.
- [17] O. Pauly, D. Mateus and N. Navab. “*Building implicit dictionaries based on extreme random clustering for modality recognition*”. In: *Medical Content-Based Retrieval for Clinical Decision Support*. Springer, **2012**: 47–57.
- [18] S. Haas *et al.* “*Superpixel-Based interest points for effective bags of visual words medical image retrieval*”. In: *Medical Content-Based Retrieval for Clinical Decision Support*. Springer, **2012**: 58–68.

- [19] C. Kurtz *et al.* “On combining image-based and ontological semantic dissimilarities for medical image retrieval applications”. *Medical image analysis*, **2014**, 18(7): 1082–1100.
- [20] C. Kurtz *et al.* “A hierarchical knowledge-based approach for retrieving similar medical images described with semantic annotations”. *Journal of biomedical informatics*, **2014**, 49: 227–244.
- [21] N. Otsu. “A threshold selection method from gray-level histograms”. *Automatica*, **1975**, 11(285-296): 23–27.
- [22] R. Raina *et al.* “Self-taught learning: transfer learning from unlabeled data”. In: *Proceedings of the 24th international conference on Machine learning*, **2007**: 759–766.
- [23] H. Lee *et al.* “Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations”. In: *Proceedings of the 26th Annual International Conference on Machine Learning*, **2009**: 609–616.
- [24] A. Coates, A. Y. Ng and H. Lee. “An analysis of single-layer networks in unsupervised feature learning”. In: *International conference on artificial intelligence and statistics*, **2011**: 215–223.
- [25] A. Krizhevsky, I. Sutskever and G. E. Hinton. “Imagenet classification with deep convolutional neural networks”. In: *Advances in neural information processing systems*, **2012**: 1097–1105.
- [26] M. Kim, G. Wu and D. Shen. “Unsupervised deep learning for hippocampus segmentation in 7.0 tesla mr images”. In: *Machine Learning in Medical Imaging*. Springer, **2013**: 1–8.
- [27] Y. Xu *et al.* “Deep learning of feature representation with multiple instance learning for medical image analysis”. In: *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, **2014**: 1626–1630.

- [28] H. Chen *et al.* “Automatic Localization and Identification of Vertebrae in Spine CT via a Joint Learning Model with Deep Neural Networks”. In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015*. Springer, **2015**: 515–522.
- [29] H. R. Roth *et al.* “Deeporgan: Multi-level deep convolutional networks for automated pancreas segmentation”. In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015*. Springer, **2015**: 556–564.
- [30] Y. Freund and R. E. Schapire. “A decision-theoretic generalization of on-line learning and an application to boosting”. *Journal of computer and system sciences*, **1997**, 55(1): 119–139.
- [31] J. Friedman, T. Hastie, R. Tibshirani, *et al.* “Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors)”. *The annals of statistics*, **2000**, 28(2): 337–407.
- [32] P. Viola and M. Jones. “Robust real-time object detection”. *International Journal of Computer Vision*, **2001**, 4: 51–52.
- [33] A. Vezhnevets and V. Vezhnevets. “Modest AdaBoost-teaching AdaBoost to generalize better”. In: *Graphicon*, **2005**: 987–997.
- [34] T. G. Dietterich, R. H. Lathrop and T. Lozano-Pérez. “Solving the multiple instance problem with axis-parallel rectangles”. *Artificial intelligence*, **1997**, 89(1): 31–71.
- [35] Q. Zhang and S. A. Goldman. “EM-DD: An improved multiple-instance learning technique”. In: *Advances in neural information processing systems*, **2001**: 1073–1080.
- [36] Z.-H. Zhou, Y.-Y. Sun and Y.-F. Li. “Multi-instance learning by treating instances as non-iid samples”. In: *Proceedings of the 26th annual international conference on machine learning*, **2009**: 1249–1256.

- [37] J. C. P. Paul Viola and C. Zhang. “*Multiple instance boosting for object detection*”. In: *Advances in neural information processing systems*, **2005**: 1417–1424.
- [38] L. Mason *et al.* “*Boosting algorithms as gradient descent in function space*”. In: **1999**.
- [39] B. Babenko *et al.* “*Simultaneous learning and alignment: Multi-instance and multi-pose learning*”. In: *Workshop on Faces in’Real-Life’Images: Detection, Alignment, and Recognition*, **2008**.
- [40] J. Wu *et al.* “*MILCut: A Sweeping Line Multiple Instance Learning Paradigm for Interactive Image Segmentation*”. In: *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, **2014**.
- [41] R. Achanta *et al.* “*SLIC superpixels compared to state-of-the-art superpixel methods*”. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, **2012**, 34(11): 2274–2282.
- [42] M. Kandemir *et al.* “*Digital Pathology: Multiple instance learning can detect Barrett’s cancer*”. In: **2014**.
- [43] Y. Xu *et al.* “*Weakly supervised histopathology cancer image segmentation and classification*”. *Medical image analysis*, **2014**, 18(3): 591–604.
- [44] T. Ahonen, A. Hadid and M. Pietikainen. “*Face description with local binary patterns: Application to face recognition*”. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, **2006**, 28(12): 2037–2041.
- [45] S. Osher and J. A. Sethian. “*Fronts propagating with curvature-dependent speed: algorithms based on Hamilton-Jacobi formulations*”. *Journal of computational physics*, **1988**, 79(1): 12–49.
- [46] T. F. Chan, L. Vese, *et al.* “*Active contours without edges*”. *Image processing, IEEE transactions on*, **2001**, 10(2): 266–277.

- [47] C. Li *et al.* “*Minimization of region-scalable fitting energy for image segmentation*”. *Image Processing, IEEE Transactions on*, **2008**, 17(10): 1940–1949.
- [48] K. Zhang *et al.* “*Active contours with selective local or global segmentation: a new formulation and level set method*”. *Image and Vision computing*, **2010**, 28(4): 668–676.
- [49] L. Wang, H. Wu and C. Pan. “*Region-based image segmentation with local signed difference energy*”. *Pattern Recognition Letters*, **2013**, 34(6): 637–645.
- [50] N. Ben-Zadok, T. Riklin-Raviv and N. Kiryati. “*Interactive level set segmentation for image-guided therapy*”. In: *Biomedical Imaging: From Nano to Macro, 2009. ISBI'09. IEEE International Symposium on*, **2009**: 1079–1082.
- [51] Y. Zhu, S. Cheng and A. Goel. “*Interactive segmentation of medical images using belief propagation with level sets*”. In: *Image Processing (ICIP), 2010 17th IEEE International Conference on*, **2010**: 4113–4116.
- [52] L. Wang *et al.* “*A general framework for context-specific image segmentation using reinforcement learning*”. *Medical Imaging, IEEE Transactions on*, **2013**, 32(5): 943–956.
- [53] S. Nowozin and C. H. Lampert. “*Structured learning and prediction in computer vision*”. *Foundations and Trends® in Computer Graphics and Vision*, **2011**, 6(3–4): 185–365.
- [54] 程强 *et al.* “概率图模型中的变分近似推理方法”. *自动化学报*, **2012**, 38(11): 1721–1734.
- [55] E. P. Xing, M. I. Jordan and S. Russell. “*A generalized mean field algorithm for variational inference in exponential families*”. In: *Proceedings of the Nineteenth conference on Uncertainty in Artificial Intelligence*, **2002**: 583–591.
- [56] J. Pearl. “*Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*”. **1988**.
- [57] C. M. Bishop. *Pattern recognition and machine learning*. springer, **2006**.

- [58] Y. Y. Boykov and M.-P. Jolly. “*Interactive graph cuts for optimal boundary & region segmentation of objects in ND images*”. In: *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, **2001**: 105–112.
- [59] C. Rother, V. Kolmogorov and A. Blake. “*Grabcut: Interactive foreground extraction using iterated graph cuts*”. *ACM Transactions on Graphics (TOG)*, **2004**, 23(3): 309–314.
- [60] Y. Boykov and V. Kolmogorov. “*An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision*”. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, **2004**, 26(9): 1124–1137.
- [61] D. Nister and H. Stewenius. “*Scalable recognition with a vocabulary tree*”. In: *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, **2006**: 2161–2168.
- [62] M. Datar *et al.* “*Locality-sensitive hashing scheme based on p-stable distributions*”. In: *Proceedings of the twentieth annual symposium on Computational geometry*, **2004**: 253–262.
- [63] J. Wang, S. Kumar and S.-F. Chang. “*Semi-supervised hashing for scalable image retrieval*”. In: *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, **2010**: 3424–3431.
- [64] W. Liu *et al.* “*Supervised hashing with kernels*”. In: *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, **2012**: 2074–2081.
- [65] D. G. Lowe. “*Distinctive image features from scale-invariant keypoints*”. *International journal of computer vision*, **2004**, 60(2): 91–110.

- [66] A. Vedaldi and B. Fulkerson. “*VLFeat: An open and portable library of computer vision algorithms*”. In: *Proceedings of the international conference on Multimedia*, **2010**: 1469–1472.

## 致 谢

时光飞逝，两年多的研究生生活不知不觉就到了尾声。这是一段令人难忘的时光，充满了苦读的辛酸、失败的泪水、前行的坚毅、收获的激动。让我感到开心的是，我坚持了下来并走到了最后。在一路走来的过程中，我自己取得了很多成长和进步，而这一切都离不开所有支持和帮助过我的师长、亲人和朋友们，我要向你们表示最衷心的感谢。

我要真诚地感谢我的导师乔宇老师，在学习、科研和生活上了给予了我很多的帮助和指导。您渊博的知识、广阔的视野给了我很大的启迪，您严谨认真、一丝不苟的治学态度，给我树立了良好的榜样。您对学生方向和人生道路的指引都帮助了我极大的成长。每当我情绪低落时，您都鼓励我积极地去看待，给予我殷切的关怀。在您的悉心指导下，我的眼界更加开阔、科研能力、实践能力以及写作表达能力都有了很大的提高。在论文即将完稿之际，请允许我衷心地向您表达感谢和敬意。

真诚地感谢这两年多来与我朝夕相处的室友王剑桥同学，在这两年半与我每天分享喜悦和烦恼，一起探讨人生和学术，一起熬夜看球，在生活中给予我帮助。我还要感谢跟我同一导师的贾少勇同学、我的本科校友顾运同学在内的 504 套间的全体成员，和你们生活在一起真的十分幸运，我们一起度过了一段快乐的时光，你们都是我学习的榜样和好朋友。感谢我的好兄弟周利锋同学，能和你探讨人生和艺术，让我在这两年多中不感到孤单。感谢我的好兄弟张义卓同学，每当我感到迷茫的时候，你总是很有耐心的安慰我，去全家请我吃好吃的。我还要感谢南洋 FC 所有的兄弟们，和你们一起踢球的时光是我这两年半里最快乐的回忆。医学组已经毕业的博士周雷师兄在科研上给予我很多的启迪和帮助，我要对周雷师兄表示由衷的感谢。

真诚地感谢我的女朋友韩玲。读研期间，在我迷茫的时候，你总是毫无保留地理解和支持我，给我很多鼓励和帮助。有了你，我就不感到孤单，也有了不断前行的动力。

特别要感谢我的父母，感谢你们无私的爱，你们一直包容着我，在我求学和生活的

道路上，一直默默地支持我、鼓励我，你们就是我前进最大的动力。

还有很多我无法一一列举姓名的师长和朋友给予了我指导和帮助，在此一并表示衷心的感谢。最后，衷心的感谢在百忙之中抽出时间审阅论文的专家教授老师。

## 攻读学位期间发表的学术论文

- [1] 第一作者. Combining CNN and MIL to Assist Hotspot Segmentation in Bone Scintigraphy[C]. In the International Conference on Neural Information Processing (ICONIP),2015. (Oral , EI 检索)
- [2] 第一作者. A MIL-based Interactive Approach for Hotspot Segmentation from Bone Scintigraphy[C]. In the IEEE International Conference on Acoustics, Speech and Signal Processing(ICASSP),2016. (EI 检索)
- [3] 第二作者. NSLIC: SLIC superpixels based on nonstationarity measure[C]. In the IEEE International Conference on Image Processing(ICIP),2015. (EI 检索)