

分类号_____ 密级 _____

UDC _____

学 位 论 文

基于小波变换的显著性模型及其在腰椎间盘突出 分割中的应用

作 者 姓 名： 赵丹

指 导 教 师： 陈东岳 副教授

东北大学信息科学与工程学院

申请学位级别： 硕士 学科类别： 工学

学科专业名称： 模式识别与智能系统

论文提交日期： 2012 年 6 月 论文答辩日期： 2012 年 6 月

学位授予日期： 答辩委员会主席：

评 阅 人：

东 北 大 学

2012 年 6 月

A Thesis in Pattern Recognition and Intelligent System

The Wavelet-based Saliency Model and Its Application in the Lumbar Disc Segmentation

By Zhao Dan

Supervisor: Associate Professor Chen Dongyue

**Northeastern University
June 2012**

独创性声明

本人声明，所呈交的学位论文是在导师的指导下完成的。论文中取得的研究成果除加以标注和致谢的地方外，不包含其他人已经发表或撰写过的研究成果，也不包括本人为获得其他学位而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示谢意。

学位论文作者签名：

日 期：

学位论文版权使用授权书

本学位论文作者和指导教师完全了解东北大学有关保留、使用学位论文的规定：即学校有权保留并向国家有关部门或机构送交论文的复印件和磁盘，允许论文被查阅和借阅。本人同意东北大学可以将学位论文的全部或部分内容编入有关数据库进行检索、交流。

作者和导师同意网上交流的时间为作者获得学位后：

半年☐ 一年☐ 一年半☐ 两年☐

学位论文作者签名：

导师签名：

签字日期：

签字日期：

基于小波变换的显著性模型及其在腰椎间盘分割中的应用

摘 要

人类视觉系统的视觉选择注意机制使人们能从大量的视觉信息中自动选择一部分相关信息进行后续的加工处理。随着多媒体技术的发展,图像和视频已经逐步替代文本成为主要的信息载体。如果计算机能模拟人的视觉系统对信息自动进行有选择的加工处理,不仅能节约大量的计算资源,还能大大提高处理效率。这使得视觉注意选择机制的研究成为当前计算机视觉领域的研究热点之一。本文对视觉显著性计算模型及其在腰椎间盘的分割中的应用进行了研究,主要研究内容如下:

(1) 本文在深入分析现存多种显著性计算方法的优劣之后,从新的角度提出了基于小波特征的概率评价显著性计算模型。该模型以 YC_bC_r 色彩空间为基础,结合小波分解实现图像的特征提取,并提出了基于独立元分析的特征显著性估计,最后由自信息表征显著度进行显著图的计算。实验结果表明该模型具有较好的显著性检测能力,能够为注意力选择模型提供准确快速的显著性估计。在与现有算法的对比实验中,该模型不仅在心理学图像测试中表现出一定的优越性,自然图像的实验结果也比其他显著性模型更符合人眼的测试结果。

(2) 利用本文提出的显著性计算模型对腰椎 CT 图像进行显著性计算。算法中对图像先后进行边界剪切,金字塔增强,方向信息测度边缘提取,各向异性扩散滤波等操作,最后对滤波结果进行显著性计算。显著图中腰椎间盘区域表现出了较高的显著性,因此可以对腰椎 CT 图像进行粗提取进而得到间盘区域图像。

(3) 在通过显著性计算对间盘区域进行粗提取之后,结合 Hough 森林检测对该区域图像进行间盘中心检测。以该中心位置为中心,两个间盘中心连线的垂直方向为倾斜角度提取间盘区域进行分割。该方法不仅利用了 CT 图像在骨骼成像方面的优势,也有效去除了软组织区域无关信息对分割结果的影响。实验结果表明,通过显著性计算、腰椎中心检测以及阈值分割可以准确地分割出腰椎图像中的间盘区域。

关键词: 视觉显著性; 小波变换; Hough 森林; 腰椎间盘分割

The Wavelet-based Saliency Model and Its Application in the Lumbar Disc Segmentation

Abstract

The visual attention of human visual system enables people to select a portion of relevant information from a large of information to perform further processing. With the development of multimedia technology, image and video becomes the main information carrier instead of the text gradually. If the computer can imitate the human visual system to process the information selectively, it not only saves large amounts of computing resources, but also improves the processing efficiency. This makes the research of visual attention become a research hotspot. In this thesis, the visual saliency model and its application in lumbar disc segmentation are researched. The main contents of this thesis are as follows:

(1) The wavelet-based feature probability evaluation model is proposed after summarized the inferiors and superiors of the existing models. In the proposed model, the visual saliency of a pixel is defined as the self-information of the local features. The model applies the multi-channel structure based on the YC_bC_r color space. The feature vector is obtained using the wavelet decomposition, and the joint probability density of feature vectors is evaluated with the ICA method. The experimental results demonstrate that the model has the superior ability to detect saliency, can provide the accurate and fast evaluation of saliency. In the experiment compared with existing algorithms, the proposed model not only demonstrates the certain superiority in psychology image test, but also obtains the results accord with human's visual attention in the natural image test.

(2) The saliency map of lumbar CT image is computed by the wavelet-based feature probability evaluation model. In the algorithm flow, the image is processed by boundary remove, multi-scale image enhancement, edge detection with orientation information measure, anisotropic diffusion. Finally, the saliency of filter result is computed by the proposed saliency model. The saliency of lumbar disc region is higher than other regions, therefore the lumbar disc region image can be extracted based on the saliency.

(3) The lumber disc region is extracted roughly based on saliency, then the center point of lumbar disc is detected by Hough forest. Consider this point as the center and use the vertical direction to the line of two disc center as the inclination angle to extract a fixed size

region as the disc region image which is to be segmented. Finally, this disc region image is segmented in order to get the contour profile of the lumbar disc. The method not only takes advantage of the superiority of CT in bone imaging, but also removes the influence of soft tissue area in the segmentation. The results show that the disc can be segmented precisely from the lumbar image through the saliency computation, the disc center detection and threshold segmentation.

Key words: Visual saliency; Wavelet transformation; Hough forest; Lumbar disc segmentation

目录

独创性声明	I
摘 要	II
Abstract	III
第 1 章 绪论	1
1.1 引言	1
1.1.1 视觉显著性研究背景与意义	1
1.1.2 腰椎 CT 图像分割研究背景与意义	2
1.2 国内外研究现状及发展趋势	3
1.2.1 显著性计算模型研究现状	3
1.2.2 腰椎间盘突出研究现状	5
1.3 本文主要工作内容与安排	6
第 2 章 视觉选择注意机制及其代表模型	9
2.1 视觉选择注意机制	9
2.2 视觉选择注意模型	10
2.2.1 视觉选择注意心理学模型	11
2.2.2 视觉选择注意计算模型	13
2.3 本章小结	19
第 3 章 基于小波特征的的概率评价显著性模型	21
3.1 WFPE 模型结构	21
3.2 色彩空间选择	22
3.3 基于小波变换的多尺度特征提取	23
3.4 以自信息表征显著度	25
3.5 注意力转移与显著区域提取	26
3.6 实验结果与分析	27
3.6.1 心理学图像实验结果	27
3.6.2 自然图像实验结果	30
3.6.3 显著性评价准则	34
3.7 本章小结	36
第 4 章 基于注意力选择的间盘区域提取	37

4.1 预备知识	37
4.1.1 CT 图像及其窗口技术	37
4.1.2 腰椎知识	38
4.2 腰椎 CT 图像显著性计算流程	39
4.3 图像预处理	40
4.3.1 Gauss-Laplace 金字塔增强	40
4.3.2 基于方向信息测度的边缘提取	42
4.3.3 基于各向异性扩散的图像滤波	45
4.4 基于显著性检测的间盘粗提取	48
4.5 本章小结	49
第 5 章 基于 Hough 森林的间盘分割	51
5.1 间盘分割整体流程	51
5.2 基于 HOUGH 森林的间盘检测	52
5.2.1 构建 Hough 森林	52
5.2.2 腰椎间盘检测	54
5.2.3 间盘检测结果	56
5.3 基于 OTSU 的间盘区域二值分割	58
5.3.1 提取间盘区域	58
5.3.2 间盘区域分割	59
5.3.3 间盘分割结果	60
5.4 本章小结	62
第 6 章 总结与展望	63
6.1 论文总结	63
6.2 研究展望	64
参考文献	65
致谢	69
硕士期间发表的论文	71
作者简介	73

第1章 绪论

1.1 引言

1.1.1 视觉显著性研究背景与意义

“注意看那个戴帽子的人”，“我没有注意到他”，日常生活中我们会经常接触“注意”这个词。教育家乌申斯基曾经说过：“注意是我们心灵的惟一窗户，意识中的一切，只有经过注意才能进来。”对心理学家来说，注意意味着为了有效地处理事物而摆脱其他无关事物的影响。人类具有的可以实时高效地注意周围事物的能力就称之为注意力。

研究表明，在人脑所接受的外界信息（其中包括听觉、视觉、触觉、嗅觉等）中，视觉信息占到所有信息的80%以上。当我们睁开双眼，眼前呈现的是承载着丰富信息的场景，我们的视觉系统会只针对某一显著目标进行分析，而忽略其它不相关的事物。实现注意的过程中，人类视觉系统首先根据图像中的特征来提取图像的某些特定区域，经过快速的眼动扫描之后，该区域会被移到具有高分辨率的视网膜中央凹区，与此同时实现了对该区域的注意，接下来可以进行后续的观察与分析。正是因为人类视觉系统中视觉的选择性注意机制(Visual Selected Attention Mechanism)的存在，才使得我们能处理大量的视觉信息并及时做出反应。选择性是指人类的视觉注意系统从大量信息中筛选出与任务相关的部分信息；而注意是指忽略无关信息的干扰而仅关注筛选出的部分相关信息。综上，视觉选择性注意机制是指观察者的视觉系统具有的选择视觉场景中最“明显”区域进行处理的能力。某一区域是否“明显”是由它的视觉显著性高低决定的，即场景信息中可以引起观察者注意的物体所具有的特性。研究者普遍认为，物体的视觉显著性可由颜色、纹理、亮度、对比度、运动等特征决定。

视觉选择性注意机制是一个由认知心理学、神经学、计算机视觉与模式识别等多学科交叉的研究领域，它吸引了众多研究者参与对其研究。所以说这个研究领域的发展不论是对于生理学、认知学还是计算机视觉、人工智能等学科的发展都具有重要的意义。

如何快速准确的选择感兴趣区域，高效地完成视觉搜索任务一直以来都是认知神经科学及心理学关注的问题。在信息技术领域，图像和视频已经逐步替代文本成为主要的信息载体。计算机处理图像信息时往往只关注图像中某一部分区域，如果将注意机制引入到图像信息处理领域中从而实现主动地获取所关注目标，优先分配计算资源给目标区域，不仅能大大降低计算量，还能提高计算机的利用效率。因此将注意力选择机制引入到计算机信息处理中，使得计算机系统具备类似人的视觉主动性和选择性具有十分重要

的意义，它在图像检索、图像压缩、目标检测、主动视觉、图像分割等领域也有重要的应用价值。

对于图像检索与分类，可以利用注意力选择机制提取图像中一些具有较强区分能力、与语义紧密结合的区域来代表图像的内容作为图像的标签进行分类与检索^[1]。

在图像压缩领域，人们在提高压缩比的同时还希望保证图像质量。研究图像压缩编码的研究者将注意力选择机制引入到图像压缩中调和这对矛盾，最大程度地满足人们的需求。基于注意力选择的图像压缩的基本思想是将视觉注意模型产生的显著图作为一种概率标准，指导压缩算法设置压缩比的空间分布，即对显著目标采用低压缩比进行压缩而对非显著目标采用高压缩比压缩^[2]。JPEG2000标准已经采用该技术区别对待显著性区域与非显著性区域，保留显著区的压缩算法已经成为JPEG2000的一项新增功能。

在目标检测中，目标之所以成为目标正是因为它和背景之间存在着各种特征上或尺度上的差别。利用注意力选择机制可以确定出目标位置进而提取目标区域。近年来很多的研究者开始深入研究这一课题并取得很大的进展，Itti(2001)^[3]，Frintrop S(2005)^[4]将注意机制应用到目标检测。

在主动视觉中，如果计算机能模拟视觉注意机制进行显著性区域检测，这样计算机就具备了类似于人眼的视觉机能^[5-6]。在机器人视觉研究领域，这种类人的视觉机能使机器人在环境感知、物体识别、目标跟踪等方面的能力都有大大的提高。

图像分割算法进行分割时分割结果总是会受到背景区域的影响，如果能引入注意机制提取出图像中的目标区域，在该区域中进行区域分割，分割结果也会更加精确。

1.1.2 腰椎 CT 图像分割研究背景与意义

伴随图形、图像技术的有力手段，医疗图像的显示方法和质量都有了较大程度的改善，因此也就兴起了医疗图像处理与分析这一热门交叉学科。该学科的发展不仅使医学临床诊断水平得到了相应的提高，同时也为医学的研究与发展打下了坚实的基础^[7]。

医疗图像主要分为功能过程图像和解剖结构图像两类。其中，前者主要用于描述人体在不同状态下组织器官的功能活动情况，包括PET，SPECT，fMRI等^[8]。而后者则主要用于描述人体的生理解剖结构，其来源包括X射线，CT，MRI及超声等。

与普通图像不同，医学图像本质上具有模糊性和不均匀的特点，具体可分为如下四种^[8]。

(1) 各种成像系统不可避免地存在热噪声、成像数值算法以及活动器官如心脏、肺的周期运动，人体不自觉的运动也会造成图像在一定程度上的模糊效应。

(2) 局部体素效应。在一个边界上的体素中，常常同时包含边界和物体两种物质：

图像中物体的边缘、拐角及区域间的关系都难以精确地描述；一些病变组织由于侵袭周围组织，其边缘无法明确界定。

(3) 灰度上的模糊性。同类组织的CT值会出现大幅度的变化，如骨骼中股骨、鼻窦骨骼和牙齿的密度就有很大的差别，在同一个物体中CT值也不均匀，如股骨外表面和内部的骨髓的密度。

(4) 病理情况下的模糊性，例如，在病变情况下原来正常的组织的灰度发生了变化，如脏器表面的肿块，骨骼表面的骨刺等等，这些病理组织和正常组织之间互相浸润，边界模糊且灰度值比较接近。

正是医学图像的这些特点给医学图像处理带来了一定的困难。随着计算机视觉在图像处理方面的应用越来越多，以及临床应用对计算机辅助治疗的需求越来越高，很有必要对医疗图像进行更深入的研究工作。

据卫生部门统计表明：目前的腰椎疾病患者已经突破2亿，其中腰椎间盘突出症患者占总人口的15.2%，且有不断上升之势。曾有数据表明，在我国的50-60岁的人群中，患腰椎疾病患者比例高达71%；60岁以上人群中的比例更攀升至82%。由此可见，腰椎疾病已经成为困扰人类健康的一大问题。由于CT图像在骨成像上有着一定的优势加之费用偏低，所以目前腰椎CT图像广泛应用于腰椎疾病诊断领域，对腰椎疾病诊断有着重要的意义。

1.2 国内外研究现状及发展趋势

1.2.1 显著性计算模型研究现状

无论国内还是国外，关于视觉选择注意机制的研究都是由认知心理学的研究人员先展开的。之后随着计算机处理能力的迅速提升，人类视觉系统的这种注意选择机制受到信息处理领域的关注，于是越来越多的计算机视觉领域的研究人员也展开了对视觉注意机制模型及其应用的研究。但对于视觉选择注意机制目前的研究来说，不论是在计算机视觉领域还是心理学领域都是亟待完善的。

Cave和Wolfe于1990年指出选择性注意机制既使用来自图像的自底向上的数据信息，也使用来自高层视觉结构组织的自顶向下的任务信息^[9]。据此，注意可以划分为两类：一种是由底层数据驱动的自底向上(bottom-up)的注意，比如绿色的草地上开出一朵红色花朵一定会吸引你的注意。这种基于视觉输入的显著性，采用自底向上控制策略的注意机制属于低级的认知过程。另一种是与任务、知识等有关的自顶向下(top-down)的注意，比如你找一本黑色的本子时一定会专注看黑色的物品来寻找。这种基于任务、知识等信息，采用自顶向下控制策略的注意机制属于高级的认知过程。但不论哪种方式

在视觉领域的应用还不是很成熟，相对于后两者来说，人们在基于数据驱动方式方面取得了一些成果，提出了一些较为成功的模型。

在 20 世纪 70 年代末，David Marr 提出了计算视觉理论框架，该框架是第一个较为完善的视觉系统框架。对于视觉的信息处理过程，他提出了包含计算理论层、表征与算法层以及实现层三个层次的研究方法。同时又将获取过程分为三个阶段，即原始基元图(primal sketch)，2.5 维图(2.5 dimensional sketch)，3 维模型(3 dimensional model)表示。迄今为止，研究者们所进行的大量研究以及取得的成果，都是在 Marr 的计算视觉理论框架下完成的^[10]。

Treisman & Koch 于 1980 年提出了著名的特征综合理论^[11]。该理论指出，在前注意阶段会对各种不同的特征并行的处理和编码。每一个特征维度包含几张特征图，最后的每张特征图的显著性在最终的显著图中进行编码。如果一个目标和干扰物在一个特征维度上不同，并且所有干扰物的特征都是同性的，那么它可以被简单，快速，并行的检测出来。如果存在两个以上的维度是不同的，那么它需要线性的搜索时间，这完全符合心理学的结论。

随后 Wolfe 提出了一个在心理学领域十分有影响力的 Guided Search 2.0 模型^[12]。这个模型的主要目的是解释和预测视觉搜索实验的结果。与特征综合理论不同的是，Wolfe 理论中一种特征维度（如颜色等）只有一个自下而上的特征图，在这个特征图中，不同的特征（如红色，绿色等）被表达出来。而且每个特征维度不但包含自下而上的特征图，而且还包含自上而下的特征图，这使得目标很容易从干扰物中区分出来。

在特征综合理论和 Wolfe's Guided Search2.0 的基础之上，Itti & Koch 于 1998 年提出基于显著性特征的注意力选择模型，称为 NVT^[13-15]。输入模型的图像先通过多个通道的处理，经过多尺度的滤波分析后会获得一个二维的显著图，这个显著图中显著度最大的区域为注意力首次选择的注意焦点，其次强的区域为第二选择，依次转移下去就得到了多个显著区域。之后，Walther 等人对 NVT 模型进行改进后用于目标识别，提出 Saliency Toolbox (STB)^[16]。

2005 年，Bruce 首次从信息论的角度提取视觉注意特征和分析图像的显著性^[17]。此后，他提出一种基于局部统计的图像显著性算法，认为图像中最显著的区域应该是信息量最大的区域。在此理论之下，Bruce 等人提出了基于信息最大化的显著图获取方法，该方法结合香农的自信息度量和神经回路，建立自下而上的视觉注意模型，最大化场景信息来获取图像显著图。

2007 年，Hou 从残留谱的角度提出了一种基于频域的显著性计算方法^[18]。从信息论

对编码的要求来看,单幅图像的信息可以看作由先验信息和新颖信息两部分组成,而新颖信息的部分可能就是输入图像的显著图。通过计算图像的新颖信息部分即可得到图像的显著图。

在SR算法的基础上,Guo通过研究发现用于获取图像特征显著信息的不是SR算法中提出的新颖信息部分而是图像频率域相位信息。由此他们提出了利用傅立叶相位谱来获取显著图的方法(PFT)和时空显著图算法(PQFT)^[19]。2010年,Chen在总结了现有的基于傅里叶变换的显著性计算模型之后,提出了频率谱调制模型(FSM),模型中引入了多特征通道与横向竞争来模拟人的视觉系统^[20]。

虽然视觉注意机制的研究有着不容小觑的理论和应用价值,但事实上在这方面的研究也存在着一些困难。目前神经生理学,尤其是实验学方面的研究还十分有限,尚处于探索阶段;认知心理学中有关视觉选择性注意的理论描述也大多为假说性观点,尚需物理验证;计算机视觉方面的现有研究成果也往往是面向具体应用任务提出的一些普适性有限的计算模型或方法。

1.2.2 腰椎间盘突出研究现状

目前医疗领域的影像设备种类较多,不同的设备有着不同的成像原理,也就对应拥有不同的优势。对于CT图像来说,在骨骼的显示上取得优于其它任何设备的效果,在成像速度与扫描费用上也有一定的优势,这些优点使其广泛应用于腰椎疾病的检测。但是受CT成像技术、原理的限制,腰椎CT图像的分割也具有不容忽视的困难。

腰椎图像分割是一个由人工分割发展到半自动分割再到自动分割的逐步发展过程。早期的椎体分割完全是靠人工完成分割,操作人员在原始图像上手动画出期望的边界。这种完全手动的分割方法不仅费时费力,而且分割结果完全依赖分割者的解剖经验与知识。伴随计算机技术的发展分割方法由人工分割发展到半自动分割方法,操作人员结合自己的知识经验等,利用计算机来完成对腰椎图像的人机交互分割。与人工方法相比,半自动的分割方法虽然分割速度快,精度高,但图像分割结果仍然受操作者的知识与经验影响较大。随着一些新兴技术的发展如人工智能技术和模糊技术,近年来图像分割领域中也涌现出一些全自动的分割技术。全自动分割方法则完全无需操作人员干预,但也正是因为整个分割过程全部由计算机完成,因此对分割算法的精度要求也就随之提高。

在医学图像分割领域,相关研究人员已经尝试过多种分割方法。从处理对象来看,医学图像分割可分为面向整体区域的整体分割与面向边缘的部分分割两种类型^[21]。整体分割法通过计算同一个图像区域内的灰度均匀性来识别图像中的不同区域,如阈值分割、区域生长和分裂等分割方法,但这些分割方法通常会造成过度分割;而部分分割法

则根据区域间的不同性质（如灰度不连续性）来划出各区域间的分界线，如微分算子、边界曲线拟合法等，但通常分割线不完整会导致分割结果会得到错误的区域轮廓或轮廓存在间断现象。

基于形变模型的分割方法根据图像中区域与边界信息并结合物理学、几何学和近似理论多方面知识进行分割。算法利用两种信息来进行图像中的目标分割，分别是从小图像中提取的自底向上的信息和一些自顶向下的先验知识，如目标位置、大小、形状等。从物理学角度看，形变模型可以看做一个在施加外力和内部约束条件下自然反应的弹性物体。形变模型可以直接产生闭合的曲线或曲面而且并对伪边界和噪声都有很好的鲁棒性。但是它对边界的初始位置比较敏感，有时还需要人工设置合适的参数。随着医学影像设备的发展，医疗图像的空间分辨率和软组织分辨率都有所提高，基于形变模型的方法开始大量应用于医学图像，并获得了一定程度的成功^[22-24]。

但对于本文中的腰椎CT图像分割椎间盘的任务来说，形变模型并不适用，因为腰椎CT图像中的间盘在图像中表现的并不是闭合的轮廓，而只是通过上下椎骨的边界体现出来的，这就无法控制闭合曲线使其在到达椎间盘的边缘时停止演化。

1.3 本文主要工作内容与安排

视觉显著性计算的研究是当前视觉领域的热门课题，也是一个漫长的研究过程。在对当前研究现状进行深入分析之后，本文对视觉显著性的计算模型及其应用进行了深入研究，提出了一些新的思路，同时也进行了大量的实验进行验证。

目前的显著性计算模型大致可以分为两种，即基于空域的计算模型，基于频域的计算模型。前者可以计算局部显著性，但该类模型算法复杂，运算时间长，同时其运算结果严重依赖于参数的选取。后者运算速度上有优势，但是只适合计算全局显著性。为了调和这两种模型之间的矛盾，本文提出一种基于小波特征的概率评价模型(WFPE)，模型采用小波变换和独立元分析来定量的计算视觉显著性。视觉显著性计算在多种领域有着广泛的应用价值，本文结合显著性计算与Hough森林对腰椎CT图像进行间盘检测。

本文一共分为六个章节，各章节的内容安排如下：

第一章介绍了视觉注意选择机制的研究背景及意义，并概括介绍了目前视觉选择注意机制在国内外的研究现状。

第二章对当前最有代表性的心理学模型和计算模型进行介绍。

第三章将会介绍我们提出的基于小波特征的概率评价显著性计算模型，并在其中对模型中的色彩空间的选择，小波特征的获取，如何用概率表征显著性进行详细解释。

第四章将视觉显著性计算模型应用到腰椎间盘的粗检测中。首先对原始腰椎CT图像

进行窗口调节并转存为灰度图像，接下来剪去无关边界以去除其对分割任务的影响。剪切后的图像进行金字塔增强，边缘提取以及滤波等处理后进行显著性计算，在得到的显著图中间盘表现出了较高的显著性，所以可根据显著度的高低来粗略估计腰椎间盘的大概位置。

第五章主要介绍了通过显著性计算得到间盘的预估区域后利用Hough森林进行间盘检测，进而进行间盘分割。首先由显著度引导提取出间盘预估区域，将其作为待分割间盘区域图像，这样不仅提高了缩小了检测范围，提高了处理效率，与此同时也大大降低了间盘误检情况的发生。然后利用通过有监督训练生成的携带有间盘相关信息的Hough森林对待分割间盘区域图像进行检测，Hough投票结果得到待检测图像中间盘的中心位置。取以间盘中心为中心点，以两间盘中心点连线的垂直方向与水平方向的夹角为倾斜角度来提取腰椎间盘区域图像作为进行区域分割得到间盘的边缘。算法不仅很好地利用了CT对骨成像的优势，又大大减少了其他无关软组织区域的影响。

第六章是总结和展望，这一章会对硕士阶段的所有工作进行总结，同时对将来的研究方向进行展望。

第 2 章 视觉选择注意机制及其代表模型

在绪论中已经对视觉选择注意机制的研究背景，意义以及研究现状进行了简要介绍。本章我们会在 2.1 节详细介绍一下视觉选择注意机制，并在 2.2 节中介绍当前主流的心理学模型，以及一些具有代表性的计算模型。最后，在 2.3 节对本章内容进行小结。

2.1 视觉选择注意机制

在人类视觉处理信息的过程中注意机制是一项重要的心理调节机制。研究表明，对视觉信息进行筛选是人类视觉系统的基本特征，而这种选择往往是由注意机制实现的。在观察日常场景的时候，我们的眼睛时刻处于运动状态，即眼球的扫视运动，平均每秒钟眼球扫视达5次左右。人类视觉系统的注意选择机制，正是通过眼球的扫视将感兴趣的区域移动到具有高分辨率的视网膜中央凹区，从而对该区域实现注意并在各个感兴趣区域进行逐个转移。生理学上认为，大脑中的某些相通区域控制眼动与注意力转移，这是形成眼动的根本原因。而控制机制可以分为基于任务驱动的自上而下的控制和基于数据驱动的自底而上的控制。观察图像时，人类的视觉系统采取的是并行的处理方式，而分析场景图像时采取的却是一种串行化的处理策略，前者产生的信息量远远大于后者，视觉选择注意机制就起到了连接这个失调过程的作用。

图 2.1 描述了自底而上的数据驱动的视觉选择注意机制流程^[28]。通过图像采样、特征提取、注意目标搜索三个模块的操作来从输入图像中发现注意目标，形成了计算速度较快，可操作性较强的数据驱动的注意焦点检测方法。

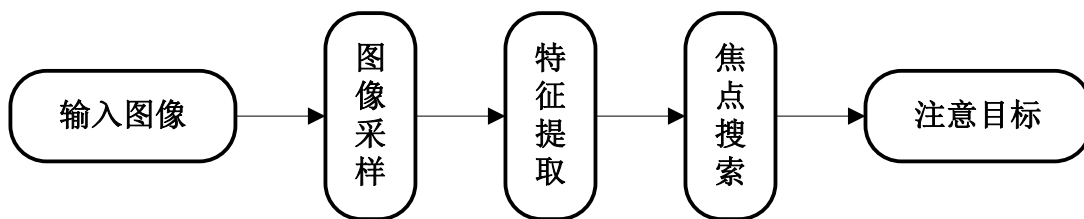


图2.1 数据驱动的视觉注意机制
Fig. 2.1 Data-driven model of visual attention

图 2.2 描述了自下而上的任务驱动的视觉选择注意机制流程^[28]。与图 2.1 不同的是，视觉信息处理过程增加了高层知识的驱动，其中包含任务、知识库和视觉期望。图 2.3 中列举了三个关于视觉注意的示例图，观察这些图片的时候就是实现注意机制的一个过程。其中图(a)中的三角形，图(b)中的实心五角星，图(c)中的向右下角倾斜的直线，都会第一时间吸引我们的注意。从图像自身来说，图片中的这几个物体自身与周围差异较

大，导致显著性高，自然就会吸引我们的注意。

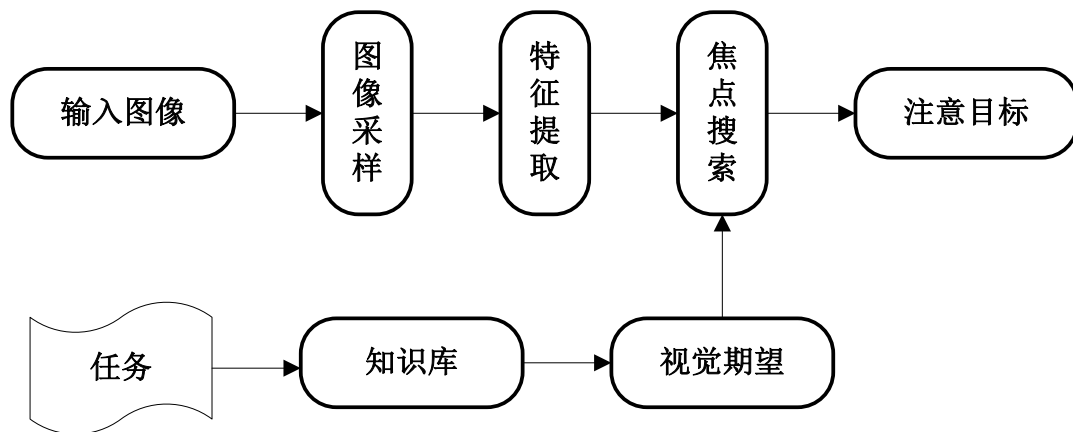


图2.2 任务驱动的视觉注意机制
Fig. 2.2 Task-driven model of visual attention

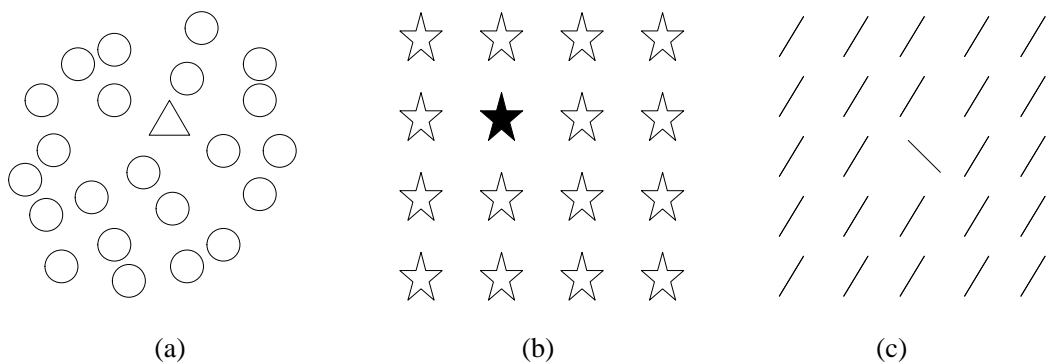


图2.3 视觉注意的示例图
Fig 2.3 The example images of visual attention

2.2 视觉选择注意模型

由于自顶向下的模型中的任务，知识，经验等千变万化，只能具体问题具体分析，所以研究人员主要对如何建立统一的自底向上的模型进行研究。通常情况下，一个视觉选择注意的计算模型中包含三个阶段：预注意阶段，显著图生成阶段以及注意力转移阶段。

(1) 预注意阶段

预注意阶段主要对各种早期的视觉特征（如颜色、运动和方向等）进行计算。在预注意阶段，这些特征在整个图像上进行计算，不同的特征对注意引导的作用取决于它们在图像中的显著性，然后通过竞争来体现。

(2) 显著图生成阶段

接下来如何利用预注意阶段的计算结果来引导注意呢？现有的模型几乎都沿用了Koch 和 Ullman 于1985在文献[25]中提出的基于显著图(Saliency Map)的引导方式。显著图是一幅二维图像，图像中描述了输入原图中每点的视觉显著性，它融合了多种特征

在多尺度上的局部对比度。显著性图能否准确地引导注意可作为视觉显著性模型优劣的评判标准。

(3) 注意力转移阶段

生成显著图之后, 根据胜者为王机制(Winner Take All, WTA), 通常取显著度最大的点作为首次注意点, 接下来对注意力进行转移。通过认知实验可知, 注意力按照返回抑制机制(Inhibition of Return, IOR)进行转移。也就是说被注视过的物体的显著性将会被抑制, 由此保证该物体不会被重复注意。现有大多数模型都是根据胜者为王和返回抑制机制进行注意力获取与转移, 即寻找当前显著图中显著度最大的位置作为注视目标, 而一旦该位置被注视过, 其显著度降为零。

2.2.1 视觉选择注意心理学模型

心理学领域中存在着许多与视觉选择注意有关的模型, 这些模型可以用来更好的解释人的感知能力。本节中将介绍目前为止两个最有影响力的模型, Triesman 提出的特征综合理论和 Wolfe 提出的 Guided Search 2.0 模型。

2.2.1.1 Tresiman & Koch's Model

1980年, Tresiman提出了著名的特征综合理论, 该理论是目前普遍为人们所接受的视觉注意的心理学模型^[11]。图2.4所示为该模型的总体概况。首先将看到的场景图像作为stimulus输入, 其中颜色、方向等不同的特征在前注意阶段被自动并行地处理。

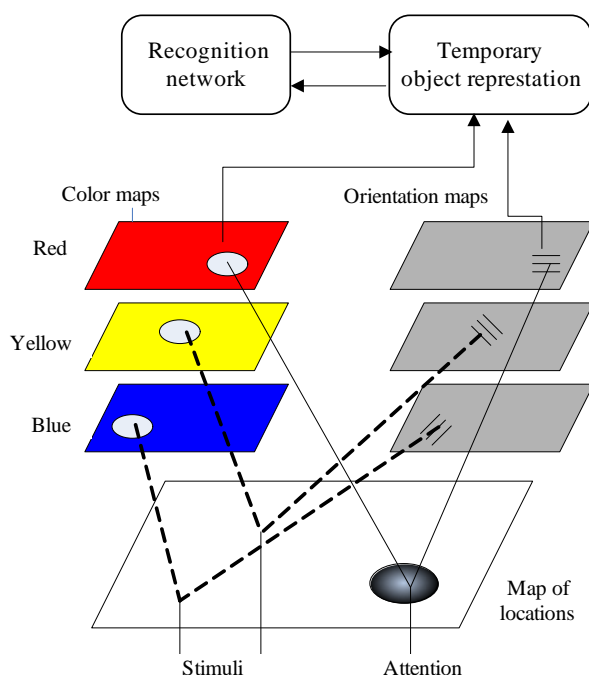


图 2.4 Treisman 的特征综合理论的模型

Fig. 2.4 Treisman's features comprehensive theory model

这里的每个特征维度（如颜色）包含有多种特征（红色，黄色，蓝色）。提取特征后会形成颜色特征图和方向特征图，随后将各特征图的显著性综合编码得到最终的显著图，图中仅根据不同位置的显著度不同表现出不同的灰度。扫描并提取出含感兴趣目标的场景区域，接下来模型将根据取出来的感兴趣区域，利用不同的特征进行编码，并获得物体的表示（图2.4中的Temporary representation模块）。最后这个物体的表示将会被一个神经网络所学习，进而实现目标识别。

特征综合理论认为，如果目标和干扰物只在一个特征维度上不同，并且所有干扰物的特征都是同性的，那么目标可以被简单，快速，并行的检测出来。如果存在两个以上的特征维度是不同的，那么它需要线性的搜索时间来检测，这是完全符合心理学的结论。

2.2.1.2 Wolfe's Guided Search Model 2.0

除了上面介绍的特征综合理论之外，Wolfe 提出的指导搜索模型(Guided Search Model)也在心理学领域有一定的影响力。和特征综合理论的模型一样，指导搜索模型也从1989年提出第一个版本(Guided Search Model 1.0)之后一直在改进。1994年，Wolfe在文献[12]中提出了 Guided Search Model 2.0。随后在1996年和2007年相继提出了 Guided Search Model 3.0^[29]和 Guided Search Model 4.0^[30]。这里针对 Guided Search Model 2.0 进行介绍。

图2.5描述了指导搜索模型的基本结构。将场景图像信息作为stimulus信息输入，不同的特征滤波器（比如颜色、方向）对输入进行滤波，可以获得相应的特征图。模型中不仅包含基于数据驱动的自底而上的特征图，对应地还包含有和任务有关的自上而下的特征图。通过加权方式对这些特征图进行融合可得到一张最终的综合显著图。

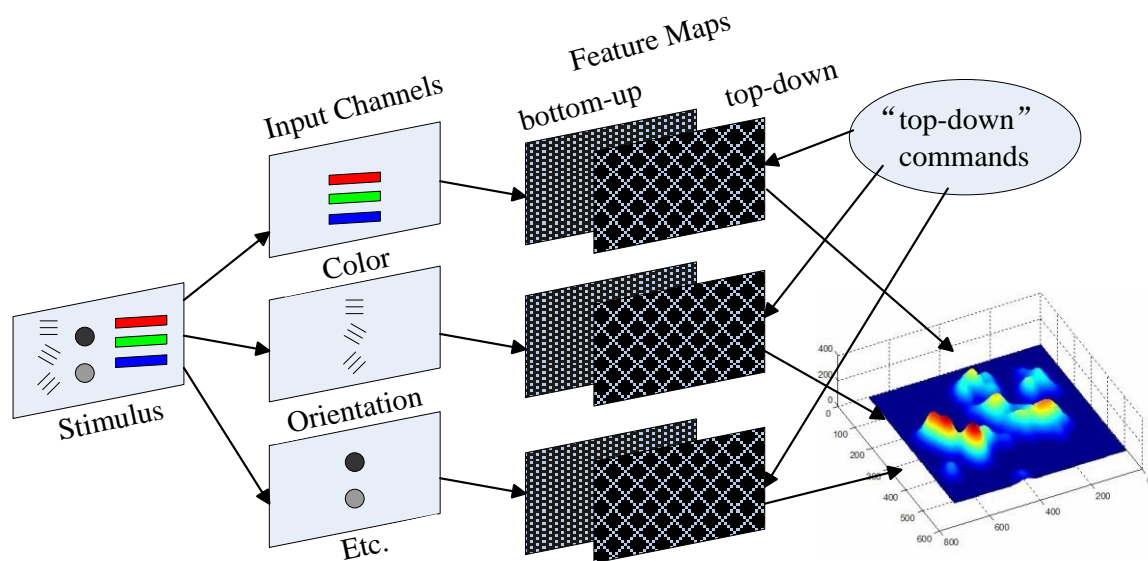


图 2.5 Wolfe 提出的指导搜索模型
Fig. 2.5 Wolfe's Guided Search Model

Wolfe 的模型中的很多概念都是和特征综合理论一致的，但对有些利于计算机实现的方面进行了更详细的设计。比如特征综合理论中每一种特征有不同的特征图，而 Wolfe 理论一种特征维度只有一个自下而上的特征图，其中包含了该特征的不同类型。另一方面不同的是，模型中除了自下而上的数据特征图，在对应的特征维度里还包含自上而下的任务特征图。

以上只是心理学中的视觉注意理论模型，它们只是将人在处理注意力选择的工作机理模块化，流程化，但并不能用来计算输入图像中的注意目标的具体位置。

2.2.2 视觉选择注意计算模型

上一节中介绍了两种当前主流的心理学模型，即特征综合理论和指导搜索模型。在它们的基础之上，计算机领域的研究人员提出了许多可计算的注意力选择模型。这些模型大致可以分为两种：基于空域计算的模型和基于频域计算的模型。基于空域计算的模型当中最有影响力的当数 Itti 提出的 Neuromorphic Vision Toolkit (NVT) 模型与 Bruce 提出的 Attention of Information Maximization(AIM)。而频域中较有名的是张立明和过晨雷提出的 PFT 以及 PQFT 模型。下面就对这几种模型进行详细介绍。

2.2.2.1 基于空域的计算模型

(1) Neuromorphic Vision Toolkit (NVT)模型

1998年，在 Koch 和 Ullman 提出的视觉注意的计算结构基础上，Itti 提出了 NVT 计算模型。从提出至今，Itti 带领的团队一直在对该模型进行完善。

NVT 模型以特征综合理论为基础设计，图 2.6 描述了模型的基本结构及流程^[31]。先通过线性滤波器对输入图像进行去噪，随后分别计算颜色、亮度和方向三个通道的特征。对于各通道数据进行逐层低通滤波处理和降采样，建立多尺度的图像金字塔。接下来对每个特征通道的不同尺度特征图之间的“中心点-周边区域”差异进行计算并归一化。然后合并归一化每通道的各尺度特征图，从而获得颜色、亮度和方向三个通道各自对应的综合特征图。

通过金字塔高斯滤波器对输入图像进行逐层低通处理和降采样，得到一个含有 9 个不同尺度图像的金字塔，分别是 1:1（第 0 层）到 1:256（第 8 层）的尺度。设 r, g, b 分别代表红、绿、蓝三通道，则亮度 I 可由式(2.1)获得。

$$I = (r + g + b)/3 \quad (2.1)$$

红、绿、蓝、黄四种颜色通道分别由式(2.2)的四个公式获得，若有输出结果为负值则均置为 0。

$$\begin{aligned}
 R &= r - (g + b)/2 \\
 G &= g - (r + b)/2 \\
 B &= b - (r + g)/2 \\
 Y &= (r + g)/2 - |r - g|/2 - b
 \end{aligned} \tag{2.2}$$

假设 W_{lpf} 为金字塔形高斯低通滤波器，通过 W_{lpf} 对 R, G, B, Y 以及 I 五个通道进行处理。设原始五个通道的图像分别为 R_0, G_0, B_0, Y_0 与 I_0 ，通过低通滤波器滤波后，降采样取得低一个分辨率的图像，然后不断地滤波和降采样。最终可以得到 9 个不同尺度的四个颜色通道和亮度信息的金字塔 $R_\sigma, G_\sigma, B_\sigma, Y_\sigma$ 和 I_σ ，其中 $\sigma \in \{0, 1, 2, \dots, 8\}$ 。

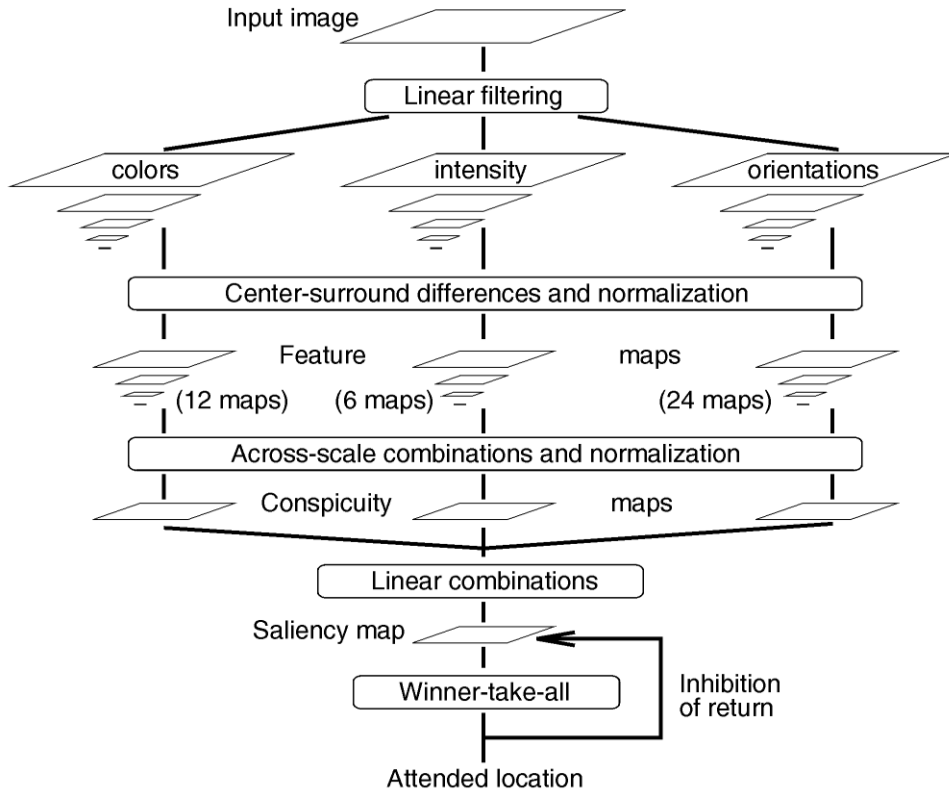


图 2.6 Itti 等人提出的 NVT 模型

Fig. 2.6 NVT model by Itti et al.

方向信息由 Gabor 滤波器获得，将输入图像经过不同尺度的 $W_{bpf}(\sigma; \theta)$ 进行滤波后得到输出为：

$$O_\sigma(\theta) = W_{bpf}(\sigma; \theta) * I \tag{2.3}$$

其中 $\sigma \in \{0, 1, 2, \dots, 8\}$ 代表相应尺度的图像， $\theta \in \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$ 对应 4 个不同的角度方向。

中心边缘差由不同尺度图像之间的差异求得。设 c 对应滤波后产生的不同尺度的图像，选取图像 $c \in \{2, 3, 4\}$ 的像素点作为中心点，而边缘区域是对应的在 $s = c + \delta$ 的图像层，这里 $\delta \in \{3, 4\}$ 。两层之间的差值是通过拉伸低尺度的图像至高尺度图像同样大小后对应像素点相减获得。如此生成一系列特征映射图，实现了不同尺度的图像层特征的提

取。亮度信息通道中可得到 6 幅特征映射图 $I(c, s)$ ，颜色通道中红绿特征、蓝黄特各 6 幅，而方向通道中包含 24 幅四个方向的特征映射图。

将三个通道的特征图线性组合得到最终的特征映射图即显著图，显著图中的显著度越高的区域相应更容易吸引注意力。但对于已经注意过的显著区域其显著性会受到抑制，这样注意力才能按照显著图中显著度的高低发生转移而不会停留在最显著的位置。

该模型以特征综合理论为基础，模仿人类视觉中自底而上的预处理过程，实现对场景图像的显著度的计算，目前得到了广泛的应用。但该模型的复杂度较高，运算效率低，而且运算结果严重依赖参数的选取（如各滤波器参数，不同尺度综合参数，线性迭加参数等）。在此之后，Walther 等人对 NVT 模型进行改进后提出 Saliency Toolbox (STB)^[16]。该模型主要用于目标识别，但是模型结果对参数依赖性高以及算法的复杂度高仍然是这类方法的弱点。

(2) Attention of Information Maximization (AIM)模型

2005 年，Bruce 在文献[17]中首次提出从信息论的角度提取视觉注意特征和分析图像的显著性。在视觉选择注意就是获取最大场景信息的假设下，他提出了一种基于局部统计的图像显著性算法。在此基础上，Bruce 和 Tsotsos 提出基于自信息最大化的显著图计算方法^[32]，并结合香农的自信息度量和神经回路，建立自下而上的视觉显著性模型，通过最大化场景信息获取图像显著图。模型实现框架如图 2.7 所示，下面我们介绍一下 AIM 算法流程中各模块及其相关概念。

研究表明人类动物的视觉系统建立在图像统计的稀疏编码原理上，在低级视觉注意中，训练学习一幅自然图像统计量稀疏编码会产生与之类似的简单细胞感受野。而独立元分析(ICA)作为主元分析(PCA)的一种延伸，可直接利用所测数据的统计特性得到基向量或基函数，进而提取图像的内部特性。

基于此，Bruce 等人从 Corel stock photo 数据库的 3600 幅自然图像中，对每幅图像随机提取 100 个 $7 \times 7 \times 3$ 的 RGB 图像块，一共可得到 360000 个像素块，然后利用独立元分析从大量的自然图像中训练出合适的基函数 A。从大量的自然图像随机图像块中得到的基函数 A 是一个混合矩阵，通过计算 A 的伪逆得到解混矩阵 $\text{pinv}(A)$ 。解混矩阵可以分离源信号中的各分量，使其相互独立。由此可利用解混矩阵将局部区域像素矩阵的分量分离为独立分量。将图像进行分块处理，局部区域像素矩阵与解混矩阵相乘后得到局部区域像素矩阵的基系数。该基系数表示基函数在局部邻域表征时各向量的权重，基函数对应局部区域像素矩阵的分量。将图像局部区域投影到独立分量空间，得到基函数向量 $W = [w_1, w_2, \dots, w_n]$ ，各分量 w_i 之间相互独立，其中分量 w_i 的值为 v_i ，由于各分量彼

此之间相互独立,其联合概率密度函数可通过各分量概率密度函数的乘积求得,如式(2.4)所示。

$$p(w_1 = v_1, w_2 = v_2, \dots, w_n = v_n) = \prod_{i=1}^n p(w_i = v_i) \quad (2.4)$$

设当前的局部图像为 I , 其水平方向上左右相邻的两幅局部图像为 I_l 与 I_r 。将 I, I_l, I_r 分别与解混矩阵相乘,得到三个基系数矩阵,将这三个矩阵中对应的系数重新整合,然后基于邻域基系数用高斯核密度估计出当前局部图像的基系数概率。

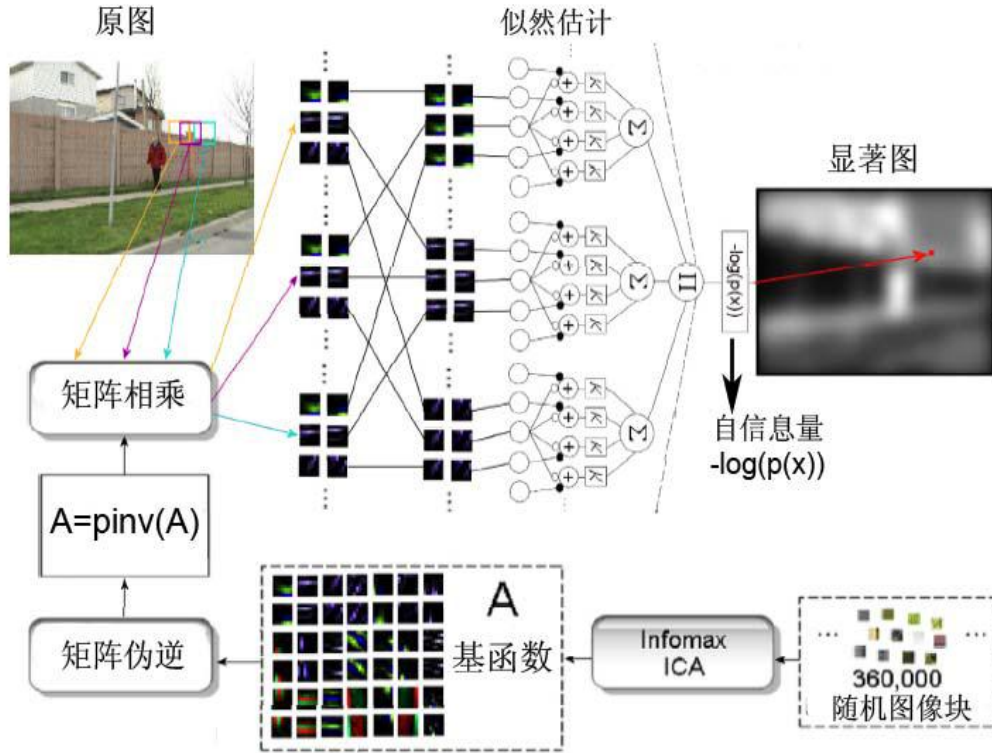


图2.7 AIM算法实现流程
Fig. 2.7 The flow chart of AIM algorithm

采用高斯核密度估计对 W 的各分量进行似然估计,令 $w_{i,j,k}$ 表示局部邻域独立基系数,局部邻域中心在 (j,k) ,式(2.5)表示基于高斯窗的估计。

$$p(w_{i,j,k} = v_{i,j,k}) = \frac{1}{\sigma\sqrt{2\pi}} \sum_{\forall s,t \in \Psi} \omega(s,t) e^{-\frac{(v_{i,j,k} - v_{i,s,t})^2}{2\sigma^2}} \quad (2.5)$$

其中 Ψ 表示整幅图像的范围, $\omega(s,t)$ 表示核密度函数中各高斯函数的权重。

显著度是通过量化每个局部图像区域的自信息量来测定,即使很小的图像区域其概率分布也分布在高维度空间中,但是局部图像像素矩阵与基函数的伪逆矩阵相乘后得到的基系数相互独立,因此概率分布也就从高维空间转变为一维空间。对于任一给定的图像,每一局部图像邻域的基系数概率分布估计可通过式(2.5)来获取,而局部图像的联合

密度函数就可以表示为式(2.4)，则局部图像的自信息量空间就可以表示成式(2.6)。

$$I(x) = -\log_2(p(x)) = -\log_2(p(w_1 = v_1, w_2 = v_2, \dots, w_n = v_n)) = -\log_2\left(\prod_{i=1}^n p(w_i = v_i)\right) \quad (2.6)$$

通过以上公式求取局部自信息量，其数值代表该局部图像的显著度，数值越大则自信息量越大，故该区域的显著性也就越高。

2.2.2.2 基于频域的计算模型

(1) Phase Spectrum of Fourier Transform(PFT)模型

基于频域研究的计算模型起步较晚，2007年Hou提出了残留谱显著性检测算法(SR)^[18]。算法中将图像的新颖信息部分作为图像的显著图进行计算。在此基础上，Guo经过研究发现图像中非周期性的纹理、边缘部分对应于相位信号恢复的输出图像中的数值都是较大的。同时这些部分也是蕴藏信息量比较大的位置，携有大量的新颖信息，因而也是人脑感兴趣的部分。所以可用相位信息恢复的图像作为是表征人眼注意力选择区域的特征显著图^[19]，接下来对PFT模型进行简单介绍。

给定一幅图像 $I(x, y)$ ，对其进行傅里叶变换如式(2.7)所示。

$$f(u, v) = F(I(x, y)) \quad (2.7)$$

通过其频域结果与式(2.8)可计算获得其相位谱 $p(u, v)$ 。

$$p(u, v) = P(f(x, y)) \quad (2.8)$$

最后由相位谱进行傅里叶反变换，输出图像即为显著图，式(2.9)中 $sm(x, y)$ 为最后得到的显著图。

$$sm(x, y) = g(x, y) * \|F^{-1}[e^{i \cdot p(u, v)}]\|^2 \quad (2.9)$$

其中， F 和 F^{-1} 分别代表傅立叶变化的正变换和反变换。 $g(x, y)$ 是一个高斯滤波器。该方法计算简单，算法速度快，但抗噪能力差，而且忽略了彩色信息。

(2) 时空显著图(PQFT)模型

PFT模型提出之后，Guo等人又将其推广到四元数的傅里叶变换，从而获得时空显著图^[33]。模型可分为两个阶段，第一阶段，将输入图像表示为一幅四元数图像，它包含有四种特征，其中包括两种颜色特征、一种亮度特征以及一种运动特征。第二阶段，利用PQFT算法处理四元数图像，从而获得时空显著图。

首先建立一幅图像的四元数图像，令 t 时刻的输入视频帧为 $F(t), t = 1, \dots, N$ ，其中 N 为输入视频的总帧数。每帧包含有红、绿、蓝三个通道的信息，记为 $r(t), g(t), b(t)$ 。四个广义调制的颜色通道可以通过等式(2.10)-(2.13)来获得

$$R(t) = r(t) - \frac{g(t) + b(t)}{2} \quad (2.10)$$

$$G(t) = g(t) - \frac{r(t) + b(t)}{2} \quad (2.11)$$

$$B(t) = b(t) - \frac{r(t) + g(t)}{2} \quad (2.12)$$

$$Y(t) = \frac{r(t) + g(t)}{2} - \frac{|r(t) - g(t)|}{2} - b(t) \quad (2.13)$$

通常在有些神经元的感受野中，神经元会被一种颜色（比如红色）刺激而兴奋，却会被另外一种颜色（比如绿色）抑制^[34]，这种颜色的组合被称为颜色拮抗对。在人的视觉皮层中存在红/绿，绿/红，蓝/黄，黄/蓝4种颜色对。这样两个颜色通道可以通过式(2.14)和式(2.15)获得。

$$RG(t) = R(t) - G(t) \quad (2.14)$$

$$BY(t) = B(t) - Y(t) \quad (2.15)$$

图像的亮度通道和运动通道可以通过式(2.16)和式(2.17)获得：

$$I(t) = (r(t) + g(t) + b(t)) / 3 \quad (2.16)$$

$$M(t) = I(t) - I(t - \tau) \quad (2.17)$$

其中 τ 为时延，一般取为3。这样就获得了4个通道的信息：两个颜色通道，一个亮度通道以及一个运动通道。故一幅图像可以表示为形如式(2.18)的四元数形式。

$$q(t) = M(t) + RG(t)u_1 + BY(t)u_2 + I(t)u_3 \quad (2.18)$$

其中 $u_i, i=1,2,3$ 满足 $u_i^2 = -1$, $u_1 \perp u_2, u_2 \perp u_3, u_1 \perp u_3, u_3 = u_1 u_2$ 。

将 $q(t)$ 表示成超复数(symplectic)的形式：

$$q(t) = f_1(t) + f_2(t)u_2 \quad (2.19)$$

$$f_1(t) = M(t) + RG(t)u_1 \quad (2.20)$$

$$f_2(t) = BY(t) + I(t)u_1 \quad (2.21)$$

得到四元数图像之后利用PQFT来计算时空显著。一幅四元数图像 $q(n,m,t)$ 的傅立叶变换可以写成如式(2.22)所示：

$$Q[u, v] = F_1[u, v] + F_2[u, v]\mu_2 \quad (2.22)$$

$$F_i[u, v] = \frac{1}{\sqrt{MN}} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} e^{-j\mu_i 2\pi((mv/M) + (nu/N))} f_i(n, m) \quad (2.23)$$

其中， (n,m) 和 (u,v) 是每个像素点分别在时间域和频率域的位置， $f_i, i \in \{1, 2\}$ 由式(2.20)和式(2.21)计算得到。为了简化，将 t 忽略。

式(2.23)的反变换形式如式(2.24)所示：

$$f_i[n, m] = \frac{1}{2\pi j \sqrt{MN}} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} e^{j\mu_i 2\pi((mv/M) + (nu/N))} F_i(u, v) \quad (2.24)$$

我们采用式(2.19)-(2.23)来获得四元数图像 $q(t)$ 的频率域表示 $Q(t)$ ，将 $Q(t)$ 表示成极数

形式，形如式(2.25)：

$$Q(t) = \|Q(t)\| e^{u\phi(t)} \quad (2.25)$$

其中 $\phi(t)$ 是 $Q(t)$ 的相位谱， μ 是单位四元数。令 $\|Q(t)\|=1$ ，则 $Q(t)$ 只包含频率域的相位信息。我们采用式(2.24)来计算 $Q(t)$ 的反变换 $q'(t)$ ，结果如式(2.26)所示：

$$q'(t) = a(t) + b(t)\mu_1 + c(t)\mu_2 + d(t)\mu_3 \quad (2.26)$$

时空显著图可以通过式(2.27)获得：

$$sM(t) = g^* \|q'(t)\|^2 \quad (2.27)$$

其中， g 是一个两维的高斯滤波器， $\sigma=8$ 。

时空显著图综合考虑了颜色，亮度，运动等特征，它们被表示成一幅四元数图像，这意味着它们是并行处理的，降低了运算复杂度，使得 PQFT 可以实时运算。而且，PQFT 和 PFT 以及 SR 一样，不依赖于参数的选择和先验知识。

该方法使用颜色拮抗作为颜色通道，在获取显著图时无法得到整个物体的显著性，显著区域往往存在于图像的物体的边缘处。另外 PQFT 所加入的运动信息无法正确区分运动物体之间的显著性强弱。

2.3 本章小结

本章首先介绍了视觉选择注意机制的概念以及相关知识，接着简要介绍了注意选择模型的三个关键阶段。由此引出了当前著名的两个心理学模型——特征综合理论和指导搜索模型。最后对当前比较流行的空域计算模型——NVT，AIM 和频域计算模型——PFT，PQFT 进行了详细介绍。经过近四十年的研究发展，视觉注意机制已经取得了一定的进步。期间提出了一系列的注意机制模型，其有效性也在实际科研中得到了一定的验证。但这些成绩并不意味着视觉注意机制问题已经是一个被解决的成熟技术问题，相反，该领域仍然面临诸多挑战。如何理解人类视觉系统视觉注意的机理，如何建立更符合人类视觉系统的模型，仍将是相关领域内的研究重点和难点。

第3章 基于小波特征的概率评价显著性模型

上一章节介绍了视觉注意机制的相关知识、心理学模型以及计算模型。计算模型分为基于空域的计算模型和基于频域的计算模型。空域模型能够较好的计算局部显著性，但计算耗时，且特征空间完备性难以保证。频域模型虽然在计算速度上有明显的优势，但无法检测局部显著性，且在神经科学原理方面缺乏有效的支撑。尽管这两类模型在结构和算法上有较大的区别，但不论哪种类型的模型都需要对视觉刺激的显著性有一个数学上的定义。一般来说，独特的特征或者明显的特征更容易吸引人们的注意。特征“独特”或“明显”意味着该特征在特征空间里出现的概率较低。因此，我们可以将目标的显著性定义为该目标特征的概率密度的某个单调递减函数。根据熵的基本定义，若该区域的概率较低则其信息量会较高。所以，一幅图像的显著区域可以通过计算图像中具有较大信息量的区域获得。

基于上述思想，本文提出了一种基于小波特征的计算模型(The Wavelet-based Feature Probability Evaluation Model, WFPE)。该模型采用多通道色彩空间结合小波分解实现图像的特征提取，并提出了基于独立元分析的特征显著性估计，实验结果表明该模型兼具了空域模型和频域模型的优点，具有较好的显著性检测与预测能力，能够为注意力选择模型提供准确快速的显著性估计。

3.1 WFPE 模型结构

如图3.1所示，本文提出的WFPE模型，在结构上采用了类似于传统空域模型的多通道并行式结构。在色彩空间的选择上，与以往常见的RGB色彩空间或色彩拮抗通道模型不同^{[16][19][33]}，该模型采用了更合理的线性 YC_bC_r 色彩空间构建多特征通道。并通过对图像进行小波多尺度分解分别得到三个颜色通道的特征向量，而特征向量的联合概率密度由独立元分析进行估计。每一像素点的视觉显著性被定义为局部特征的自信息，自信息可通过计算特征向量的联合概率密度得到。算法的具体步骤如下：

- (1) 将输入图像分解到三个不同的颜色通道；
- (2) 通过小波变换的多分辨率分析将各个通道中的信息分解为 $3l+1$ 个特征图，其中 l 表示小波分解的级数；
- (3) 缩放这 $3 \times (3l+1)$ 个特征图至输入原图像大小，并利用独立元分析来计算每张特征图中的联合概率密度；
- (4) 融合所有的概率密度图产生显著性图；

(5) 利用胜者为王(Winner Take All, WTA)和返回抑制(Inhibition of Return, IOR)机制进行注意力转移。

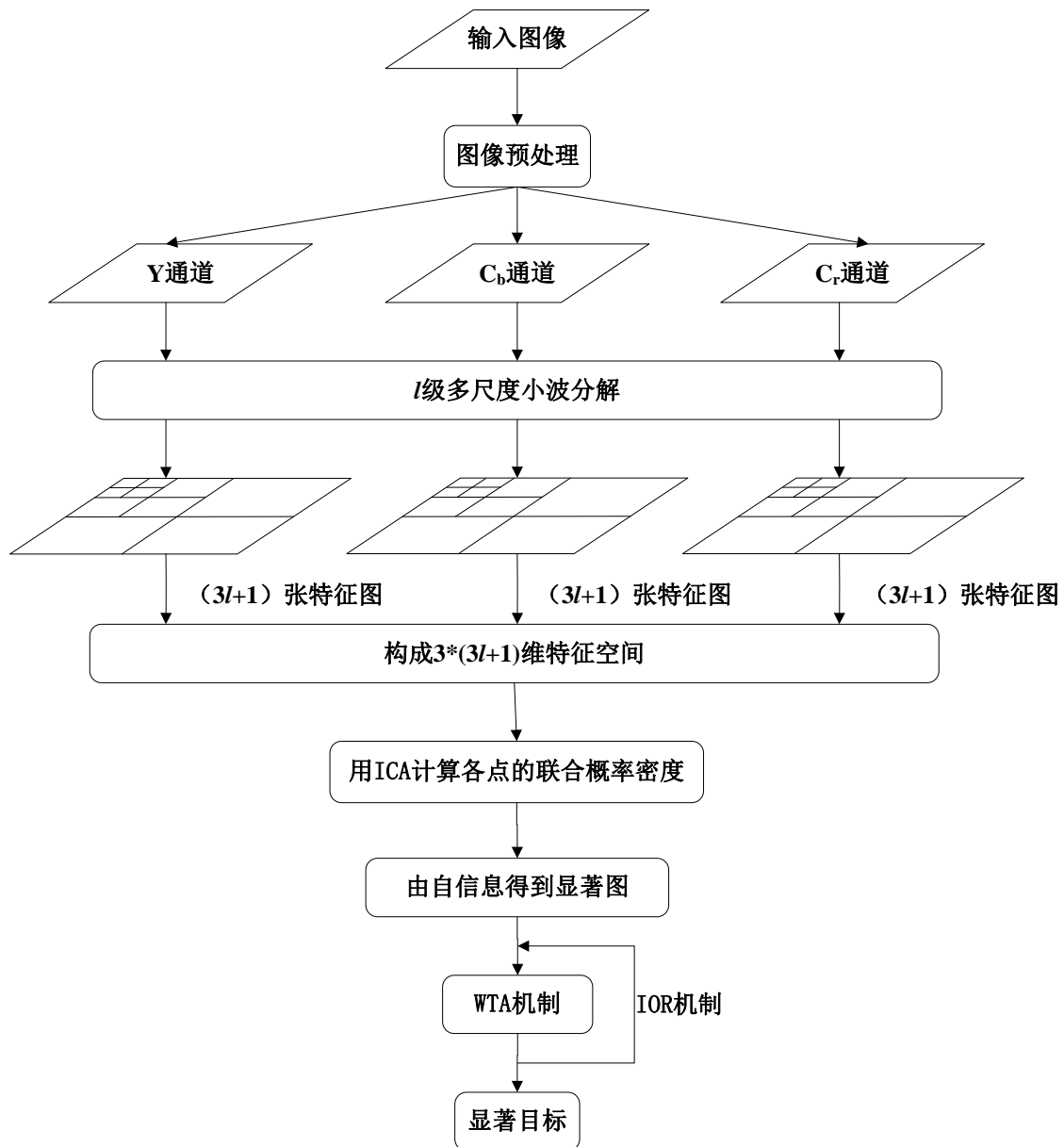


图3.1 基于小波特征的概率评价模型流程图

Fig. 3.1 The flow chart of the wavelet-based feature probability evaluation model

3.2 色彩空间选择

现有的视觉显著性计算模型多数都是基于RGB彩色模型来进行各通道特征的计算，所以颜色特征有红、绿、蓝三种基本分量。RGB色彩模型是基于笛卡尔坐标系统建立的，其色彩空间模型如图3.2(a)所示。模型为了方便起见，假定所有颜色值都归一化了，即所有 R , G , B 取值都在 $[0,1]$ 范围内。在RGB色彩模型中，所表示的图像由3个图像分量组成，每个分量都是其原色图像。每种颜色都可以由这三种颜色按照不同的比例线性组合，一共可呈现 $(2^8)^3=16777216$ 种颜色^[35]。

从神经生物学角度上来看^[36]，人类的视觉系统靠眼睛的锥状细胞来获取颜色。实验结果表明，人眼中的600~700个锥状细胞中可分为红、绿和蓝3个主要的感觉类别。大约65%的锥状细胞对红光敏感，33%对绿光敏感而对蓝光敏感的细胞只占2%的比例，但蓝锥状细胞对蓝光更敏感。所以对于 R, G, B 分量对最终显著图的贡献并非均等，应该分配三种分量以分配不同的权重来进行线性组合。

除了RGB模型外，一些非线性模型如HSV色彩模型也在一些计算模型中有过应用。HSV颜色空间中的 H 和 S 均为 R, G, B 的非线性变换，存在奇异点，在奇异点附近即使 R, G, B 的值有很小的变化也会引起变换值有很大的跳动。在亮度值和饱和度较低的情况下，采用HSV颜色空间计算出来的 H 分量是不可靠的，而且非线性函数有时会破坏色彩分布的连续性。

因此我们提出的模型采用由国际无线电咨询委员会(CCIR)提出的 YC_bC_r 彩色空间。该彩色空间由3个基组成，分别亮度分量 Y ，蓝色色度分量 C_b 和红色色度分量 C_r ^[37]，其色彩空间模型如图3.2(b)所示。从RGB颜色空间到 YC_bC_r 颜色空间的转换公式如式(3.1)所示。

$$\begin{aligned} Y &= 0.299R + 0.587G + 0.114B \\ C_b &= -0.168R - 0.3313G + 0.5B + 128 \\ C_r &= 0.5R - 0.4187G - 0.0813B + 128 \end{aligned} \quad (3.1)$$

基于 YC_bC_r 颜色空间中色度分量和亮度分量相互独立， YC_bC_r 颜色空间与 RGB 颜色空间存在一种线性变换关系且转换较为简单这些特点，因此选择 YC_bC_r 颜色空间。

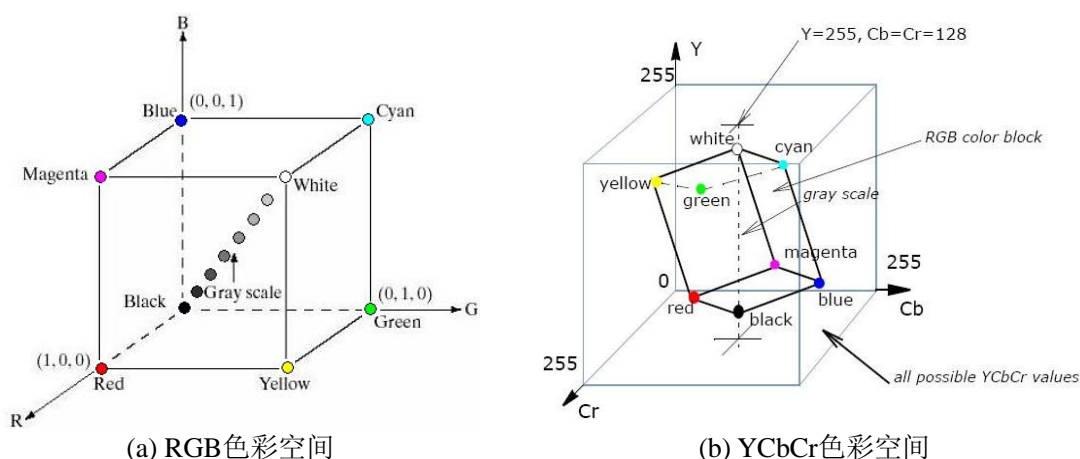


图3.2 色彩空间的彩色立方体示意图
Fig. 3.2 Schematic of the color cube

3.3 基于小波变换的多尺度特征提取

客观世界中存在着多种大小不同的物体，每种物体又同样存在着各种大小不同的局部特征。为了感知并识别这些特征，我们要在特定的尺度下感知某个特定场景下的特征。

人眼在观察某个场景时,当场景和观察者之间的距离不断变化时,视网膜感知到的图像信息也会不断变化。假想周末的时候和朋友出去踏青,从远处你会看到一大片养眼的绿色草坪,但当你走进草坪时你才会注意到草坪中竞相开放的颜色鲜艳的花朵,再走近观察花朵你还会看见在花上采蜜的蜜蜂。同样一幅场景只因尺度不同产生的注意焦点就不同。在日常生活里,我们周围的场景非常复杂,有各种大大小小的物体,事先我们又不知道他们的具体尺度,为了更好的理解图像的内容,我们需要从各个尺度上去对图像进行分析,从而建立图像的多尺度分析和处理机制^[38-39]。

算法中的多尺度分析采用小波变换来实现,具有“数学显微镜”之称的小波分析在时域和频域都具有良好的局部化性质。由于它对高频成分采用逐渐精细的时域或频域取样步长,从而可以聚焦到对象的任何细节。也正是基于这个优势,小波理论在近年来得到了迅猛的发展,也在许多领域得到了广泛而成功的应用。小波变换可以将信号分解到不同尺度的子空间上,这使得我们能够从不同的尺度来分析和处理信号,尤其是像图像这样的非平稳信号。与此同时,小波变换还具有快速性,可逆性,完备性等优点。这些优势使得图像中不同尺度内的高显著性物体都能以较小的计算量将其检测出来。

对图像 I 进行二维小波变换后图像被分解为各个分辨率上多种分量,如近似分量,水平方向细节分量,垂直方向细节分量和对角线方向细节分量。利用离散小波变换将图像分解三层,得到如图 3.3(a)所示的图像的多分辨率表示: $A_n, H_n, V_n, D_n (n=1,2,3)$ 。 A_3 表示最低频成分,是图像的一个低分辨(粗尺度)逼近, H_n 表示水平方向细节分量, V_n 表示垂直方向细节分量, D_n 表示对角线方向细节分量。图 3.3(b)为三层小波分解的示例图,其每块图像的频率成分与图 3.3(a)一一对应。

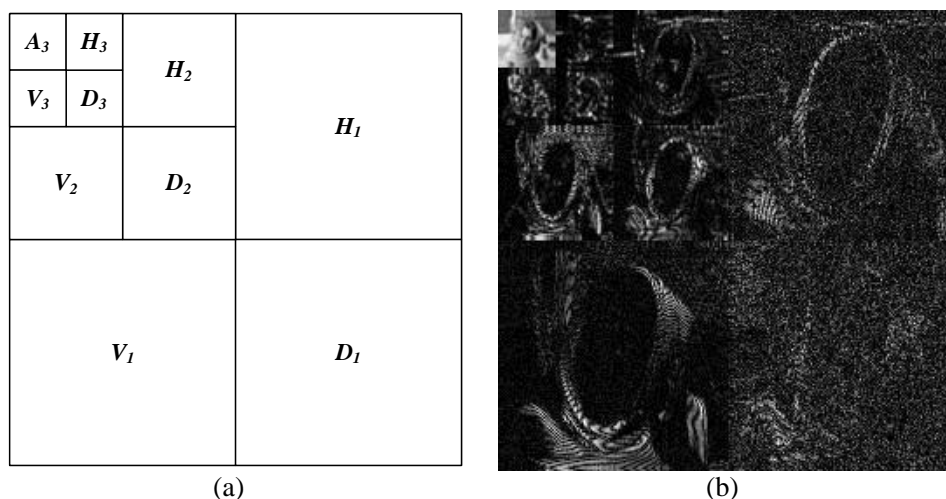


图 3.3 小波分解算法三层分解示意图

Fig. 3.3. The example images of three layers wavelet decomposition

综合考虑小波基的连续性、正交性、对称性、消失矩等特征,并结合实验效果算法

中选择 `rbio2.4` 作为多尺度分解的小波基。

3.4 以自信息表征显著度

视觉注意计算模型中，用显著图(Saliency map)来表示视觉场景的显著性。显著图中的每个像素点的值代表原图像对应点的显著性大小，且图像大小与输入图像相同。显著图不仅表示视觉场景中每个位置的显著性，并且可以通过显著性的空间分布来引导注意区域的转移。

在上一节中，已经将输入图像分解为 $3 \times (3l+1)$ 个与其尺寸相同的特征图。也就是说，每一个像素点都对应着一个长度为 $3 \times (3l+1)$ 的特征向量。显著度已经被定义为特征的概率密度函数，由此可以将显著图看作是在整个特征空间上对概率密度进行评价，即联合概率密度。接下来的问题是如何计算特征空间中的联合概率密度？最大的挑战就是特征空间的维数过高以至于无法用传统的统计方法来计算。在概率论中，联合概率是指在多元的概率分布中每个随机变量分别满足各自条件的概率，并且如果各随机变量之间相互独立，那么联合概率等于各随机变量的概率的乘积，其表达式如式(3.2)所示。

$$p(x_1, x_2, \dots, x_n) = \prod_{i=1}^n p(x_i) \quad (3.2)$$

但是对于小波分解得到的分量是正交的，它们之间并不是相互独立的。因此，我们引入独立元分析(ICA)来计算联合概率密度。从数学的角度来说，用ICA方法进行特征提取就是寻找一种能使得多维数据的特征尽可能相互独立的变换。该方法的基本思路是在独立性假设的前提下，以非高斯信号为研究对象，对多路观测信号进行盲源分离^[40]。从统计分析的角度看，作为PCA变换的延伸，ICA也是一种多变量数据分析的线性方法。但经ICA处理得到的各个分量之间不仅去除了相关性，而且还是相互统计独立的，并且是非高斯分布的信号^[28]。因此，这里将小波特征向量投影到由独立分量支撑的特征空间中，这样就可以利用式3.3计算特征向量的联合概率密度。

设小波分解得到的特征向量由 $X(t) = [x_1(t), x_2(t), \dots, x_n(t)]$ 表示，其中 t 是图像中像素值的索引， $n = 3 \times (3l+1)$ 表示所有特征图的数量。根据ICA算法，特征向量 $X(t)$ 可以看做是源信号 $S(t)$ 的线性组合，如式(3.3)所示。

$$X(t) = AS(t) \quad (3.3)$$

其中 A 是混合矩阵，源信号 $S(t)$ 和混合矩阵 A 均为未知的。如式(3.4)描述，ICA的目的是找一个向量 $Y(t) = [y_1(t), y_2(t), \dots, y_n(t)]$ 来估计源信号 $S(t)$ 。

$$Y(t) = WX(t) = WAS(t) \quad (3.4)$$

求一个分离矩阵 W ，并通过 W 从特征向量 $X(t)$ 中恢复源信号 $S(t)$ 。其中， W 是分

离矩阵其理想值应该等于 A^{-1} ，但实际上我们不知道矩阵 A 的具体值，这样就不能直接通过 A 计算矩阵 W 。ICA的基本思想就是寻求一个矩阵 W 确保 $Y(t)$ 中所有的信号之间独立性最大^[41-44]。由线性代数可知，若矩阵 W 是满秩的，则有式(3.5)：

$$P(Y(t)) = P(X(t)) \quad (3.5)$$

将式(3.5)代入式(3.4)中可以得到：

$$P(X(t)) = \prod_{i=1}^n P(y_i(t)). \quad (3.6)$$

模型中的独立分量 $y_i(t)$ 可以由ICA计算得到，并且由一幅新的特征图 R_i 表示。由此， n 维空间中的一个像素点的联合概率密度就可以从 n 个一维空间中的概率密度图得到。我们用图像 R_i 的直方图来估计它的概率 $P(y_i(t))$ ，如式(3.7)所示。

$$P(y_i(t)) = P(y_i(t) = r_k) = n_k / n, k = 0, 1, 2, \dots, L-1. \quad (3.7)$$

其中， r_k 是 $y_i(t)$ 的值， n_k 为图像 R_i 中 $y_i(t)$ 的值为 r_k 的像素点的个数， n 代表图像中像素点的总数。至此， $p(r_k)$ 是灰度级 r_k 发生的概率^[45]。

正如上文中提到，某点的概率密度越低其显著度就越高。自信息与概率之间的关系可由式(3.8)描述。

$$I(X(t)) = -\log(P(X(t))). \quad (3.8)$$

模型中将图像中每一点的显著性定义为其自信息。香农信息论中描述到，一个事件 x 的自信息反比于它发生的可能性即概率。根据式(3.6)与式(3.8)可得，像素点 t 的显著性 $S(t)$ 可由式(3.9)计算。由此，通过计算每一点的显著性值即可得到图像的显著图。

$$S(t) = I(X(t)) = -\log(\prod_{i=1}^n P(y_i(t))) = -\sum_{i=1}^n \log(P(y_i(t))). \quad (3.9)$$

3.5 注意力转移与显著区域提取

显著图生成之后的问题是如何依照显著程度转移注意力的焦点(Focus of Attention, FOA)。根据显著图生成原则可知，显著图中的极大值对应输入图像中最显著的部分，注意焦点的转移也是根据显著图中显著度的依次递减而转移。

显著图中各目标的竞争采用WTA机制来实现，注意焦点的检测和转移需要的另一个机制是IOR。返回抑制现象最早是由Posner等人于1984年在文献[46]提出的，该现象是指注意再次回到先前注意过的位置时受到了抑制，因此形象的称之为返回抑制。返回抑制反映了人的心理机制的灵活性和适应性，使注意力不会一直停留在具有最大显著性的位置，而是迅速转向以前未被注意过的区域，以利于快速的搜索活动。本文提出的模型中同样采用WTA和IOR机制进行注意力转移。显著图产生之后将其最大值点即最显著点作为注意力转移的起始位置，该点为winner，接下来会在以winner为中心的区域产生一定

的抑制，注意力焦点转移到下一个极大值点。若存在两个显著度相近的目标待注意，则采取“就近原则”将注意力转移到距离当前winner最近的目标。

通过显著图可以找到一幅图像中最显著的点，但是更多的时候我们关注的不是这个点，而是该显著点所表征的具体的物体的信息，也就是我们希望通过该显著点获取一个感兴趣区域。本文采用区域生长的方法对感兴趣区域进行提取。区域生长顾名思义，它是一种根据事先定义的准则将具有相似性质的像素或子区域聚合为更大区域的过程。具体过程是先对待处理的图像选取一些生长种子作为生长的起点，然后根据一定的相似性生长准则，将周围领域中与当前种子点性质相同或者相似的像素点合并到当前区域。然后，将这些新添加到区域中的像素作为新的生长起点重复上述的生长过程，直到满足生长终止准则或者没有新的区域可以合并。模型中以显著图中的显著点作为初始生长点进行区域生长，采用阈值法确定是否将相邻像素并入当前区域，最终生成基于视觉注意的显著区域。

至此，我们已经介绍了完整的如图 3.1 所示的基于数据驱动的视觉选择注意计算模型。从早期视觉特征提取得到不同尺度的小波特征图，然后缩放各特征图至原图大小，接下来采用 ICA 计算每个像素点在特征空间中的联合概率密度，再以各点的概率密度表征其显著度得到显著图，最后根据显著图进行注意力转移与感兴趣目标提取。

3.6 实验结果与分析

3.6.1 心理学图像实验结果

心理学图像常用来检测视觉显著性计算模型的正确性与有效性。我们从 Guo 等人的论文^[33]数据库中选取了一些代表性的心理学图像进行算法有效性验证。心理学图像分为五种，其中包括不同颜色，不同密度，不同形状，不同颜色与方向，以及添加噪声的不同颜色和不同方向的五组图像。我们对这五组图像分别进行了显著性计算。

首先，对形状相同颜色不同的一组图片进行实验，实验结果如图 3.4 所示。在第一幅图中，因为圆点颜色差异较大四种算法全部得到正确的计算结果，其中 WFPE 的显著点只有红色圆点一个目标，而其它三种算法中其余目标也表现出了较低的显著性。对于第二幅颜色差异较小的图像来说，除了 WFPE 检测到浅色圆点外，PQFT 算法的显著图中浅色圆点的显著性也是最高的，但 STB 和 AIM 则完全没有检测出浅色圆点。

接下来又进行了一组形状不同的测试实验，如图 3.5 所示。两幅实验图像中分别有加号和五角星两个不同的形状存在，按照视觉注意机制来说这两个目标的显著性最高。实验结果中可以看到，只有本文提出的 WFPE 算法得到了两幅图像中的显著目标。而 STB 和 AIM 只检测出减号中的加号，而丢失了弧线中的五角星。而对于 PQFT 来说显

著图结果中的五角星的显著度较高，但对于第一幅图像则没有检测到任何显著目标。

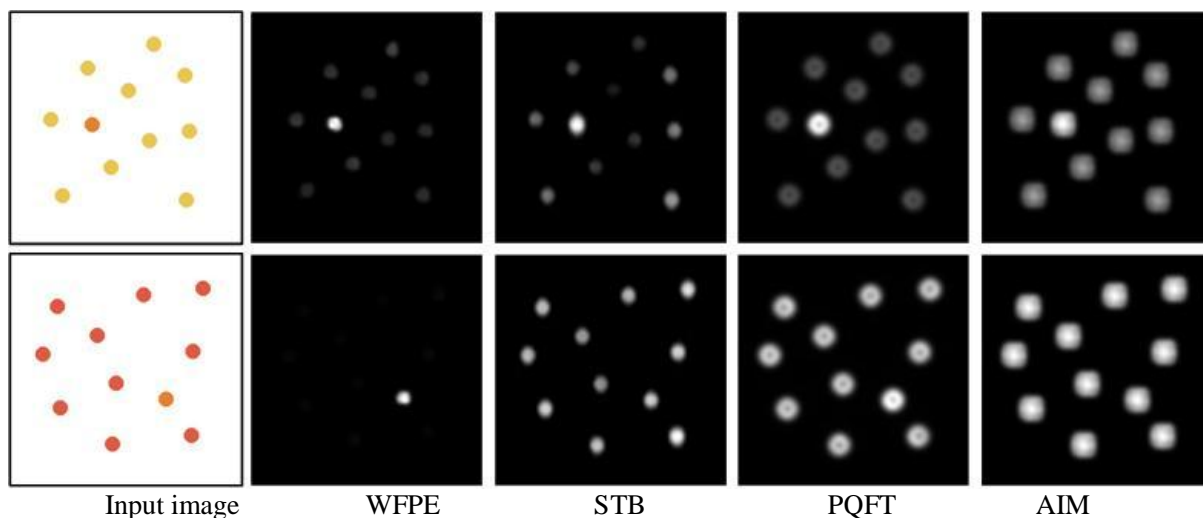


图 3.4 不同颜色显著性测试结果
Fig. 3.4 The saliency maps of different colors

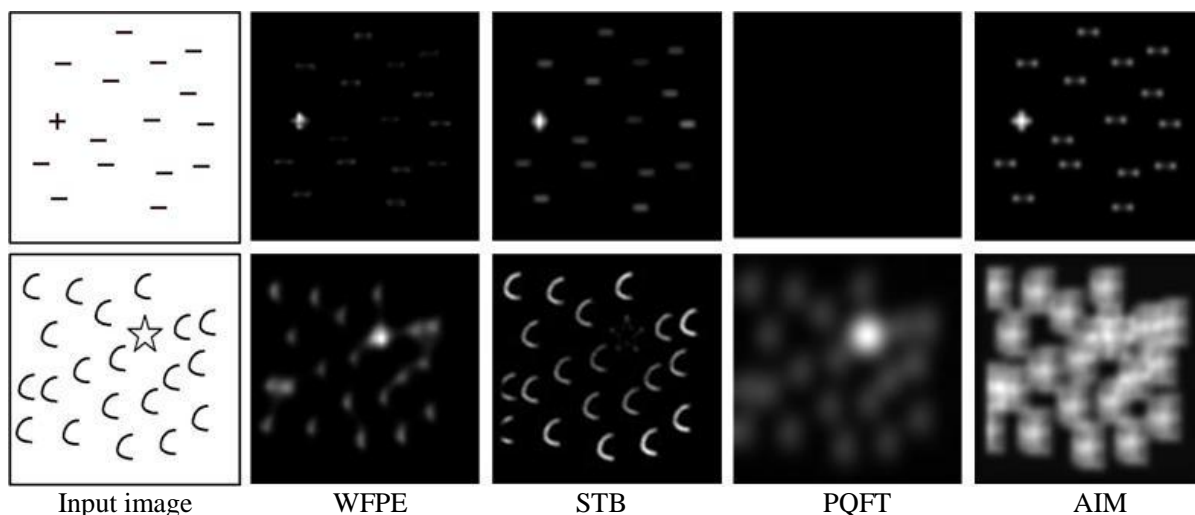


图 3.5 不同形状色显著性测试结果
Fig. 3.5 The saliency maps of different shapes

图 3.6 所示的一组实验是针对不同密度的图像进行的。对于第一幅图像，中间的高密度区域的显著性最高，四种算法也全部得到了正确的检测结果。而对于第二幅图像中的低密度区域除了 WFPE 和 AIM 的显著图中对应区域的显著度最高之外，PQFT 的显著图中没有任何规律，而在 STB 的计算结果中可以发现该算法将图像中的几个高亮度的面积很小的点区域作为显著目标，这一检测结果显然不符合人类的视觉显著性机理。

图 3.7 所示的实验的测试图像为包含不同颜色与不同方向个体的心理学图像。对于第一幅含有阿拉伯数字 5 的图像，WFPE 算法的显著图中倾斜的 5，红色的 5 以及右上角的小 5 显著度最高，STB 的显著图中显著度较高的个体为分布在外围的几个数字 5，PQFT 与 AIM 的显著图结果中倾斜 5 和右上角的小 5 表现出来的显著性较高丢失了右下角的红色的 5。第二幅图像的实验结果中，WFPE 的计算结果中左下角的绿色斜线表现

出了较高的显著性，而其余三种算法的显著图中包含输入图像中所有斜线。

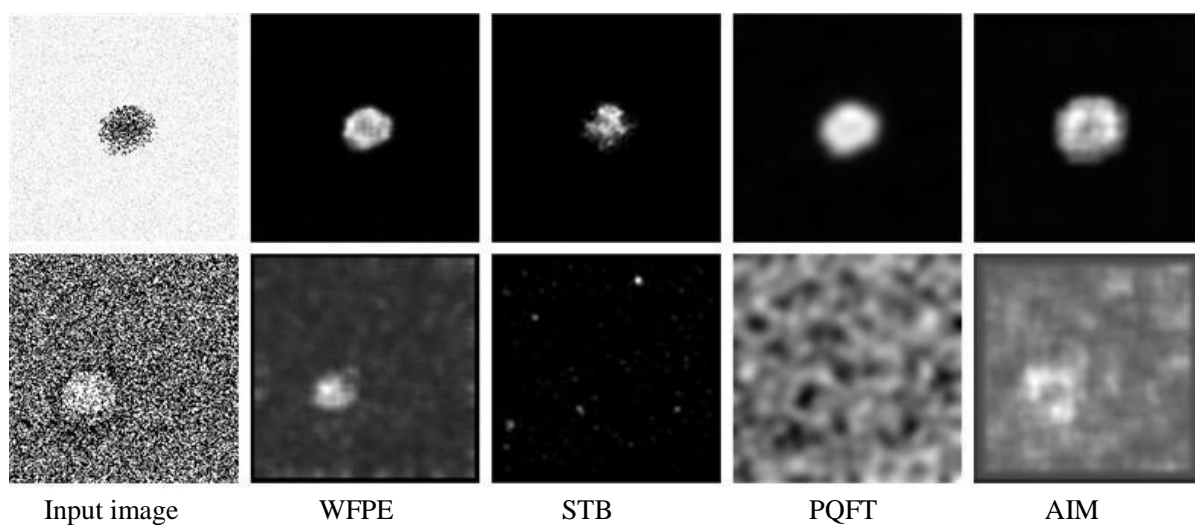


图 3.6 不同密度显著性测试结果

Fig. 3.6 The saliency maps of different densities

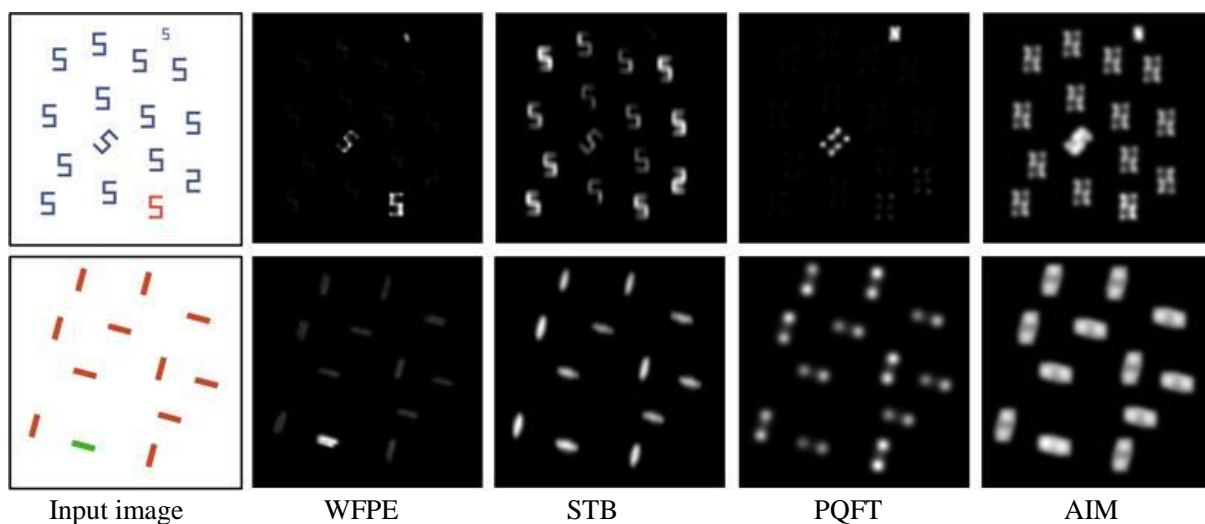


图 3.7 不同方向不同颜色显著性测试

Fig. 3.7 The saliency maps of different colors and orientation

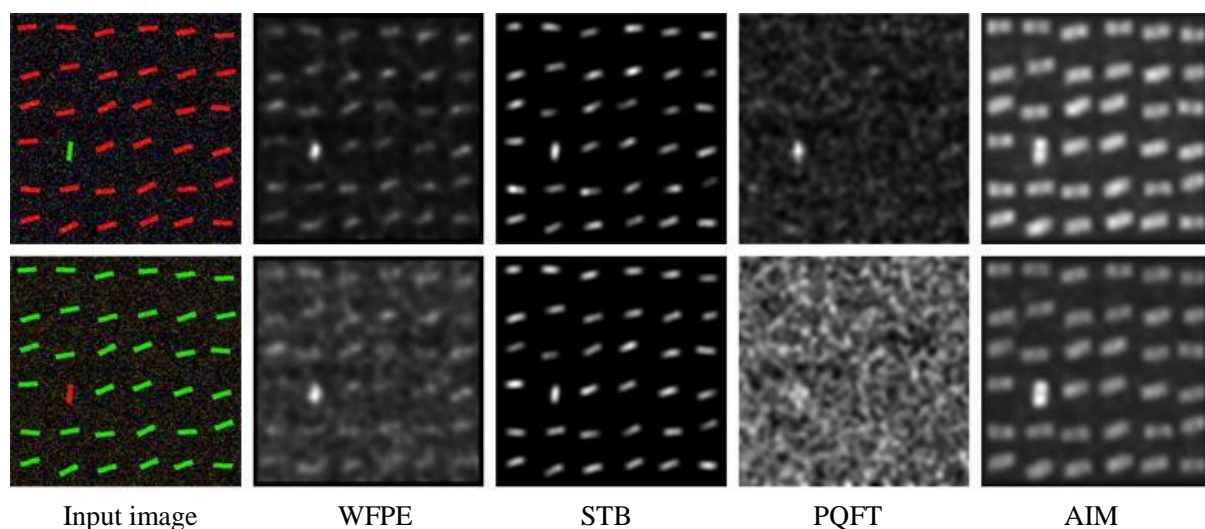


图 3.8 添加噪声图像的显著性测试结果

Fig. 3.8 The saliency maps of the noised image

最后一组图片中加入了一定的噪声干扰，如图 3.8 所示。根据视觉注意机制原理，第一幅图像中，众多红色横线中一条绿色的竖线会吸引人们的注意，所以其在显著图中应该表现出较高的显著性。而在第二幅图像中，红色的竖线在绿色的横线中表现出较高的显著性。在噪声干扰的情况下，只有 WFPE 模型对两幅图像的显著结果均符合人的注意机制。

3.6.2 自然图像实验结果

为了验证提出算法的正确性，我们采用 120 幅自然图像作为测试集进行测试。这些自然图像数据来自于 2005 年 Bruce 发表的文献[18]中所使用过的实验图像。测试集中不仅有 120 张自然图像还包括 20 个志愿者对每张图的测试所得显著图，这些显著图由志愿者人眼对图片的注意位置、注意持续时间、总注意点、平均持续时间等信息综合得来。



图3.9 不同显著性算法的实验结果

Fig. 3.9 The detection results of nature images processed by different algorithms

如图 3.9 所示为 120 幅图像中随机选出的七幅图像的测试结果对比。图中可以看出本章提出的 WFPE 的计算结果更接近测试结果。虽然 STB 也能得到显著点，但是显著

区域的形状通常是不规则的，不能表征显著目标，而且对于复杂场景算法通常会失效如图 3.9(f)。而 PQFT 的结果通常在强烈的边缘处显著度较高^[21]。而 AIM 对于一些图像的实验结果与心理测试结果偏差较大，如图(c)(d)(e)。

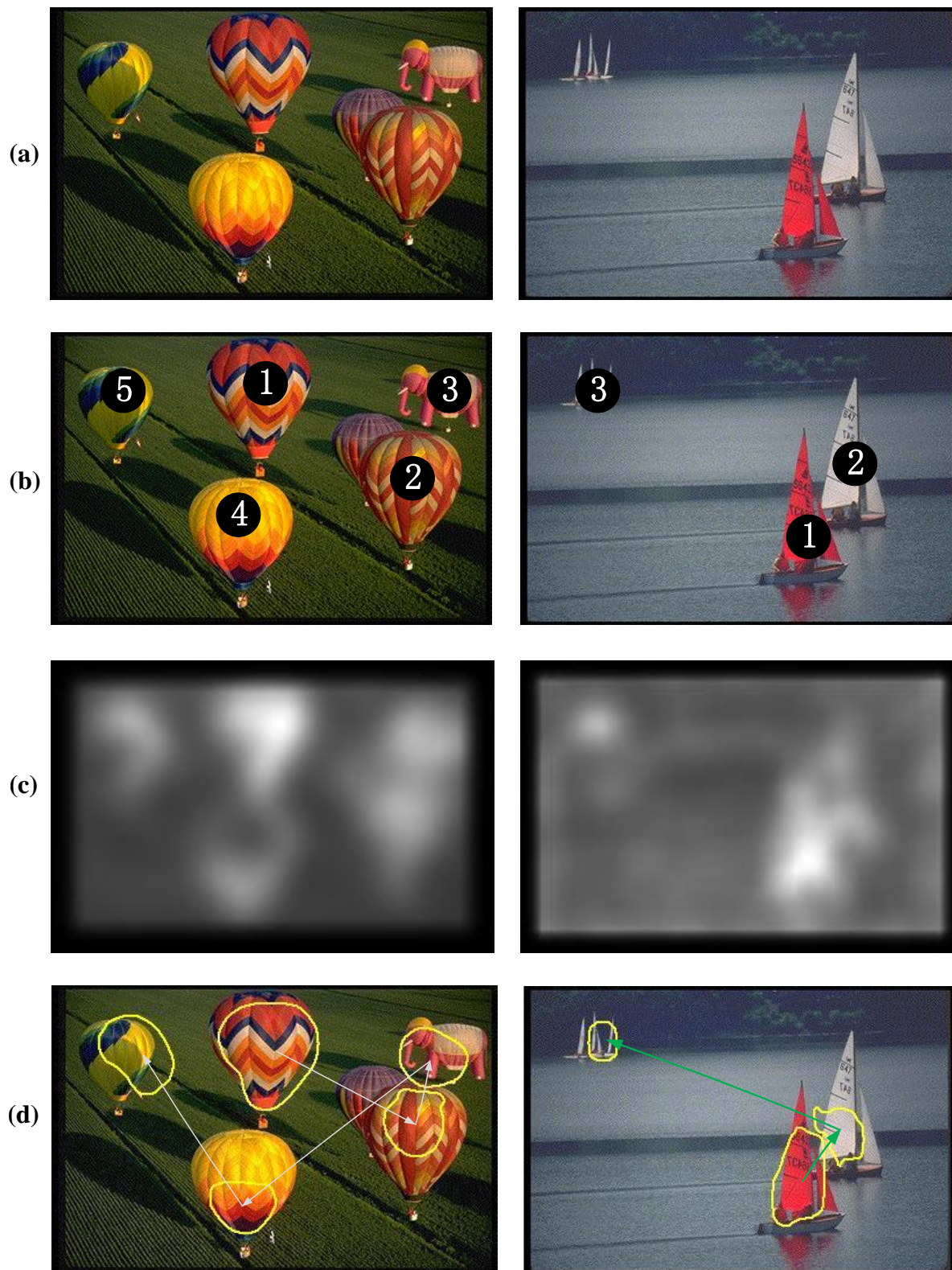
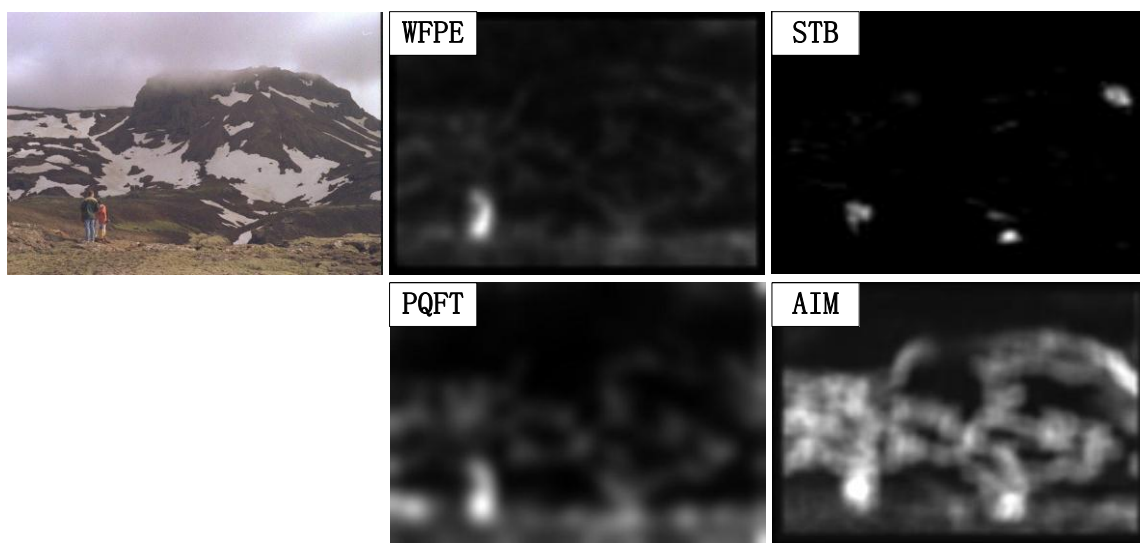


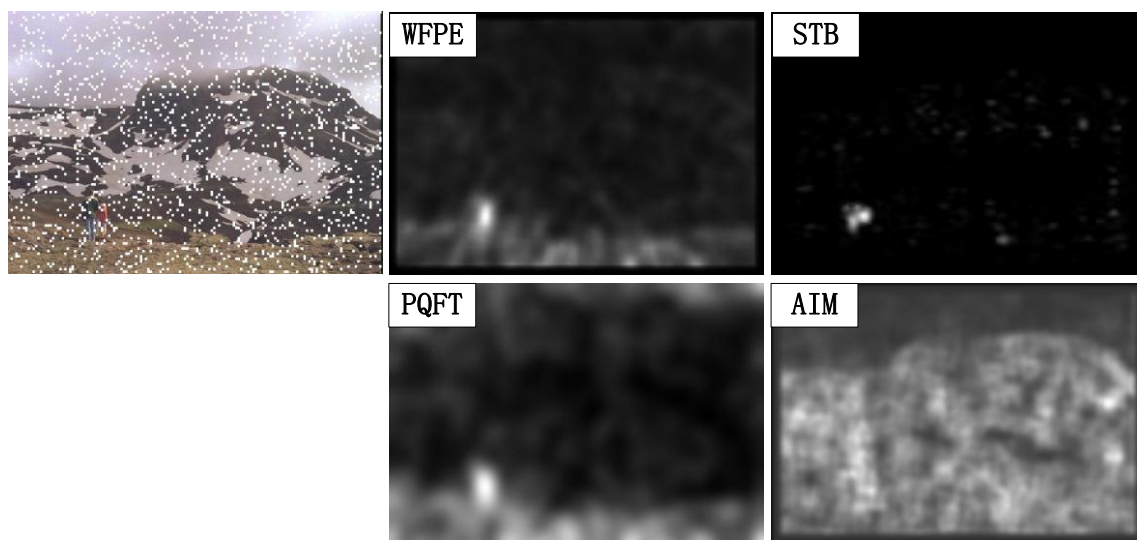
图3.10 WFPE的注意力转移结果
Fig. 3.10 The results of attention shift with WFPE

所以不论是从准确性还是显著区域形状来说，WFPE 算法都占有一定的优势。另外，我们从 iLab 实验室的图片数据库中选取了两张多目标的自然图像进行算法的注意力转移测试。图 3.10 中，图(a)为两幅测试图像；图(b)为对 15 名志愿者的测试结果，得出的对于图片的注意力转移顺序如图中物体上所标示的序号由小到大的顺序；本文的显著性算法的显著图结果如图(c)所示，图(d)中为算法的注意力转移机制的测试结果。通过结果对比可以看出，本文提出的 WFPE 算法的测试结果符合人的注意力转移规律，也就是说 WFPE 算法可以很好的模拟人的注意力选择与转移的能力。

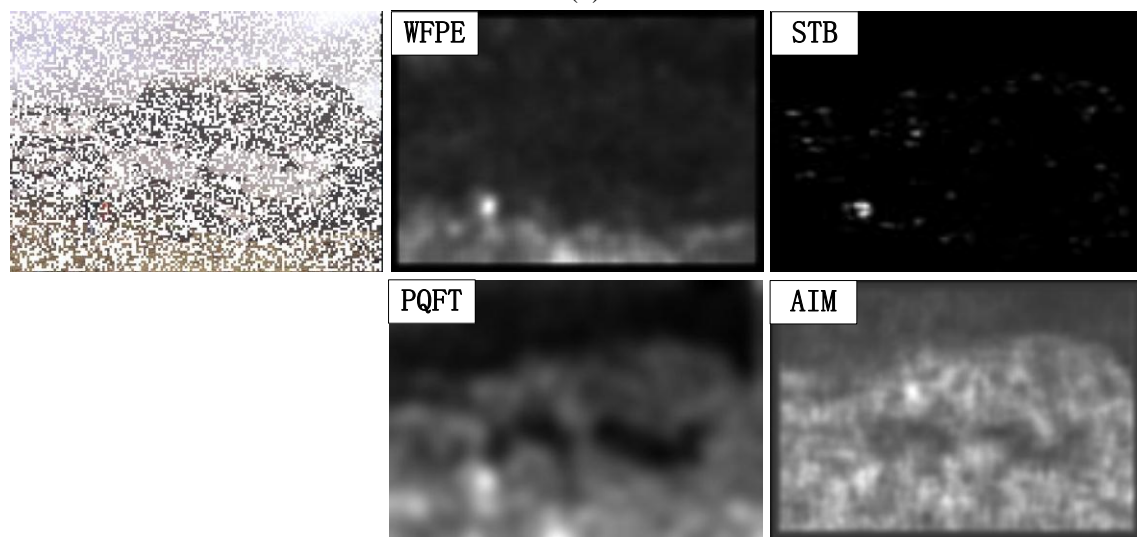
为了测试各计算模型的抗噪性能，我们又对添加过噪声的图像进行了实验。原图是文献[13]中的图像，图像中雪山脚下的两个衣着鲜艳的人为图中的显著目标。图 3.11 中图像分为四组，(a)组为原始图像（即没有添加噪声的图像）及其实验结果，(b)组为添加密度为 0.2 的白色噪声块后的图像及其实验结果；(c)组为添加密度为 0.5 的白色噪声块后的图像及其实验结果；(d)组为添加密度为 0.2 的椒盐噪声后的图像及其实验结果。每一组中的显著图结果对应分布为 WFPE（左上），STB（右上），PQFT（右下），AIM（左下）。从显著图结果可以看出 STB 在原图中检测出三个显著区域，最显著的并不是人而是一小块雪山上的区域，在添加了强度为 0.2 和 0.5 的白色噪声之后反而显著区域只有人，在添加密度为 0.2 的椒盐噪声之后又丢掉了人出现了雪山和天空两个区域，这充分说明了算法的不稳定性。而 AIM 在原图实验中检测结果中的最显著区域包含人，对于加上噪声之后的图像完全失效，可以看出其对噪声是非常敏感的。对于抗噪性稍好的 PQFT 来说可以抵抗白色噪声的干扰，对于椒盐噪声的干扰则很敏感。而本文提出的 WFPE 不论是白色噪声还是椒盐噪声都能很好的计算出显著区域。



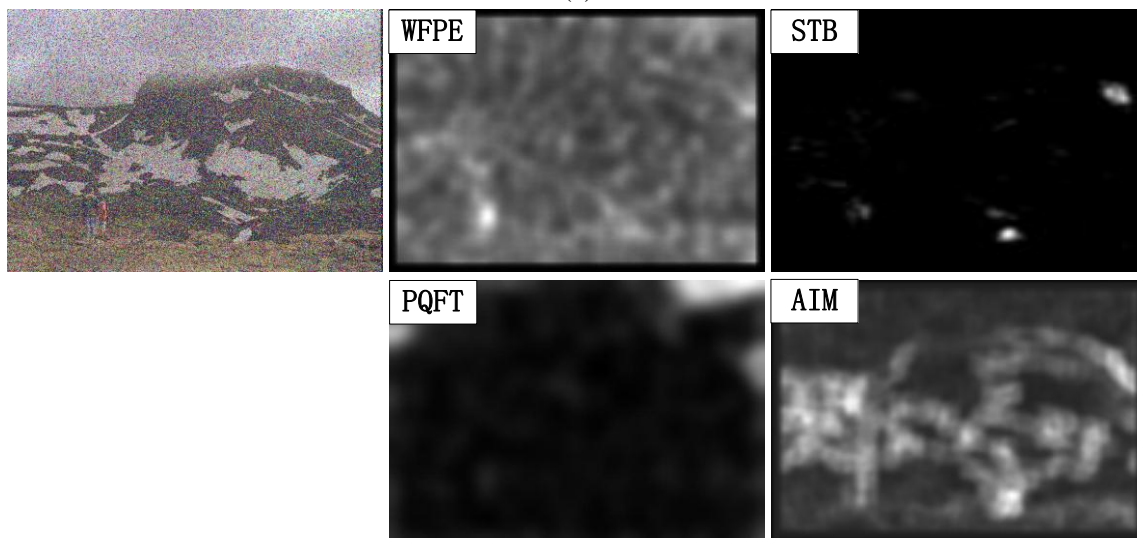
(a)



(b)



(c)



(d)

图3.11 添加噪声后的显著图结果
Fig. 3.11 The saliency maps of noised images

3.6.3 显著性评价准则

单纯依靠人眼观察显著性算法的结果图与心理测试结果图是否一致来判断不同显著性计算方法所获得的显著图的优劣是不准确的，也是不科学的，对于结果近似的计算方法更是难以评价孰优孰劣。因此，我们需要将显著性计算结果进行量化，进而定量地分析算法的各项性能和指标。

近年来，研究者开始使用 ROC(Receiver Operator Characteristic)曲线来衡量显著图预测人眼视觉系统注意力的性能^{[18][47][48]}。ROC 分析^[49]源于信号检测理论，首先是由二战中的电子工程师和雷达工程师发明的，用来检测战场中的敌军。之后很快就被引入了心理学来进行信号的知觉检测。目前 ROC 曲线分析已经在相关领域取得了很好的发展，尤其是在医学，无线电领域中，最近又在数据挖掘和机器学习领域取得了很好的发展。

ROC 曲线是一种比较两个分类模型的有用的可视工具。分类模型是将一个实例映射到一个特定类的过程。如果输出的预测是 p 而真实的结果也是 p ，那么这就叫做真阳性(TP)；然而如果真实的结果是 n ，则这就叫做假阳性(FP)。以此类推，一个真阴性发生在预测和实际结果都为 n 的时候，而假阴性发生在预测输出是 n 而实际值是 p 的时候。ROC 曲线显示了给定模型的真阳性率($TPR=TP/(TP+FN)$)与假阳性率($FPR=FP/(FP+TN)$)之间的比较评定。ROC 空间中，FPR 和 TPR 分别为 x 和 y 轴，空间中的每一个点代表一个预测输出结果。点(0,0)表示没有输出真阳性的分类，当然也没有假阳性的错误。与此相反，右上角的(1,1)点表示分类模型无条件地输出阳性分类。点(0,1)代表性能优越的分类器，真阳性率很高而假阳性率较低。对角线 $y=x$ 代表一个完全随机的预测。

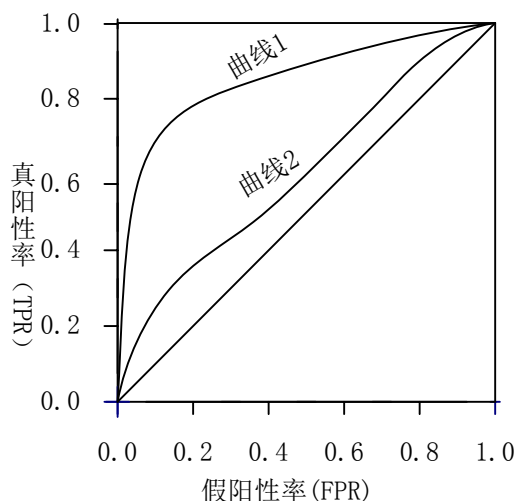


图3.12 两个分类模型的ROC曲线
Fig. 3.12 The ROC curves of two classification models

为了绘制给定分类模型 M 的 ROC 曲线，需要对样本分类预测结果正确的概率递增

排列，以分类模型认为最可能分类正确的样本为列表顶部。分类模型 M 的 ROC 曲线按以下方法绘制。从左下角开始（这里，真阳性率和假阳性率都为 0），从上至下依次检查列表中样本的实际所属类别。如果它分类正确（真阳性或真阴性），则在 ROC 曲线上向上移动并绘制一个点；反之，如果分类错误，则向右移动并绘制一个点。

图 3.12 显示了两个分类模型的 ROC 曲线。如果模型准确率很高，则随着样本在样本列表向下移动，开始出现的样本分类多半为正确分类，曲线将陡峭地从 0 开始上升。后来，我们遇到的正确分类越来越少，错误分类越来越多，曲线平缓并变得更加水平，如曲线 1 所示。曲线 2 表示样本中多数分类不正确所以曲线走势一直缓慢向右。为了评估模型的准确率，可以测量曲线下方的面积。对于分类模型来说，面积越接近 0.5，对应的模型的准确率越低，完全准确的模型的面积为 1^[50]。

然而在本文中作为评判依据的人眼实测得到的显著性分布图只是基于人眼扫视数据生成的模拟图，从图中并不能准确判断哪些像素点为正样本，哪些为负样本。因此无法按照传统方法绘制 ROC 曲线。因此，本文使用 ROC 曲线的第二种绘制方法作为评判计算模型优劣的手段。

在第二种方法中，ROC 曲线用来定量的描述两个模式类之间的可分性。已知两个模式类的概率密度函数为 w_1 和 w_2 ，如图 3.13 所示。为方便比较， w_2 画在横轴下方。

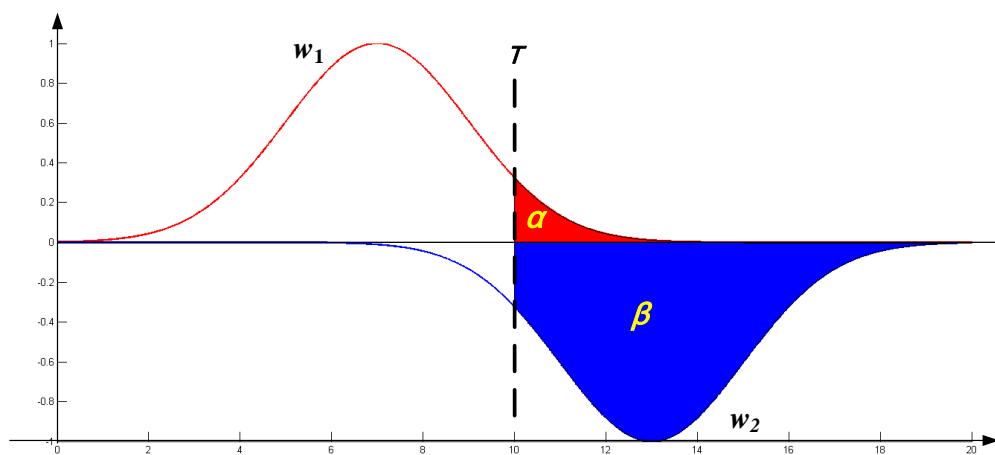


图 3.13 两种类别的可分性描述图
Fig. 3.13 The description graph of separability in two classes

令 w_1 代表阴性(negative)， w_2 代表阳性(positive)，阈值为 T ，则真阳率为图中 β 区域的面积，而假阳率为图中 α 区域面积。据此画出的 ROC 曲线可用来衡量 w_1 和 w_2 两个模式类的可分性。两个类别完全可分，则 ROC 曲线经过点(0,1)；当两个类别完全重合，则有 $\alpha = \beta, \forall T$ ，则 ROC 曲线为连接点(0,0)到点(1,1)的斜线。对比本文中的问题，令 w_1 为计算模型得到的归一化显著图（图中显著度积分为 1）， w_2 为人眼实测得到的归一化显著图。当 w_1 与 w_2 完全重合时，两图完全相同，表示计算模型得到的结果与人眼实测

结果的相似度达到最高。

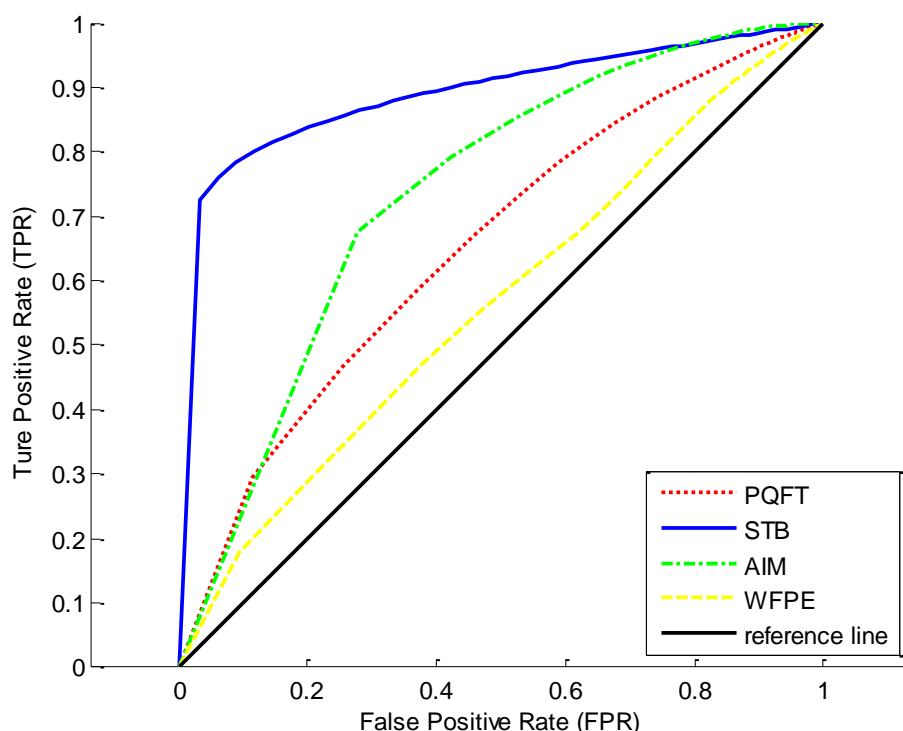


图3.14 不同显著性计算模型的ROC曲线
Fig. 3.14 ROC curves for various saliency models

实验仍然使用文献[18]中的自然图像数据库，计算模型产生的显著图可以视为 w_1 类的概率密度函数，人眼产生的显著结果视为 w_2 类的概率密度函数。首先对计算结果与实测结果的显著图进行归一化，选取阈值 T 对计算结果的显著图进行二值化。ROC 曲线用来测量两类之间的可分离度。最理想的状态是 w_1 与 w_2 的概率密度函数相同，即二者的显著图的形状相同，对应的 ROC 曲线为从(0,0)到(1,1)的对角线。同样以曲线下的面积(AUC)代表模型的准确度，但曲线下面积越小表明该显著性计算模型预测人眼注意力的性能很好。如图 3.14 所示，不同的 ROC 曲线对应不同的视觉显著性计算模型，黑色实线对应最理想的状况。从图中可以看出本文提出的 WFPE 模型在性能上优于其他现有模型。

3.7 本章小结

在分析了现有模型的优劣之后我们提出了基于小波特征的显著性算法，算法结合独立元分析计算每一像素点的概率密度，并用其表征显著性的高低。通过一系列的实验对比来证明 WFPE 算法的正确性和有效性。不仅在心理学图像测试中 WFPE 算法体现出了一定的优越性，同样在自然图像测试中与现有算法相比也有着更高的准确度。为了使实验结果更有说服力，在得到各种显著性算法的自然图像的显著图之后，又绘制了各种算法的 ROC 曲线。在 ROC 曲线的结果中 WFPE 的表现仍然优于其余算法。

第4章 基于注意力选择的间盘区域提取

随着医疗器械市场的崛起, 医疗图像在临床的应用越来越多, 医疗图像处理与分析的地位也变得愈加重要。随着图形、图像技术的发展, 医疗图像的显示方法和质量也得到了极大的改善。这不仅极大地提高了医学临床诊断水平, 也为医学的研究与发展打下了坚实的基础。但与普通图像不同, 医学图像本质上有着模糊性和不均匀的特点。正是医学图像的这些特点给医学图像处理带来了一定的困难。

对于腰椎 CT 图像来说, 虽然有骨骼成像清晰的优势, 同时也存在由噪声等各种外界因素的影响而产生的不同组织间的混叠和模糊的问题。这些问题会给腰椎 CT 图像的处理与分析带来一定的困难。本章将上一章提出的视觉显著性计算模型引入到腰椎 CT 图像的处理中, 计算其显著图并通过显著度的引导提取出图像中的间盘区域。

4.1 预备知识

4.1.1 CT 图像及其窗口技术

CT是对人体的某一部分按一定厚度的层面用X线束进行扫描, 扫描人体组织的X射线有一部分被组织吸收, 还有一部分穿过人体被检测器接收后产生信号。X线的穿透能力会随着人体组织的疏密程度不同而不同, 因此检测器会接收到的有差异的射线信号。这些有差异的射线信号转变为数字信号后由计算机进行处理, 输出即可得到CT图像。

CT图像同普通图像一样也是由不同灰度表示。图像的灰度表示人体组织对X线的吸收程度, 实质上也是对组织密度的一种反映。其中, 白色表示该区域对X线的吸收能力较强或密度较高的区域, 如骨骼; 黑色表示该区域对X线的吸收能力较弱或者密度较低的区域, 如肺部。

CT图像(DICOM格式^[51]) 中存放的是CT每个断层图像像素的CT值的编码形式, 灰度级为2048或4096。但受显示原理的限制, 只能对灰度图像做到256灰度级的精确显示, 这就需要定义像素的CT值到屏幕像素对应的灰度值(0~255)的映射。目前比较常用的是窗口技术来进行映射。

窗口技术引入窗宽和窗位的概念, 充分利用图片中的信息来观察不同的组织^[52]。窗宽是CT图像上显示的CT值范围, 在此范围内的组织和病变会以不同的灰度显示在图像中。窗位为窗的中心位置。因此对于同一幅CT图像, 只需在监视器上调节出不同的窗宽和窗位, 可分别观察到不同组织或病变的变化, 变换公式如式(4.1)所示。

$$p_w(x, y) = \begin{cases} 0, & p(x, y) \leq L_w - \frac{W_w}{2}; \\ \frac{p(x, y) - (L_w - \frac{W_w}{2})}{W_w} I_{\max}, & L_w - \frac{W_w}{2} < p(x, y) \leq L_w + \frac{W_w}{2}; \\ I_{\max}, & p(x, y) > L_w + \frac{W_w}{2}; \end{cases} \quad (4.1)$$

其中, L_w 和 W_w 分别代表窗位和窗宽, I_{\max} 是显示设备的最大灰度级(对一个8位显示器, $I_{\max}=255$)。

4.1.2 腰椎知识

腰椎为脊柱中的一部分,而脊柱就是人们俗称的脊梁骨,位于背部正中,由椎骨和椎间盘连结而成。正常成年人的脊柱由33块椎骨和23块椎间盘组成。如图4.1所示,从脊柱顶部开始颈椎7块(C1-C7),胸椎12块(T1-T12),腰椎5块(L1-L5),骶椎5块(S1-S5)和尾椎4块。从颈部到腰部,椎骨和椎间盘的大小逐渐增大,且所有椎骨中腰椎的椎体最大。脊椎骨的主体不仅负责负重,还有保护脊髓与神经末梢的作用,它们之间通过椎间盘紧紧连接。椎间盘不仅能支持在脊柱上强大的压力,也能用来减少椎骨的磨损和撕裂。正是有了椎间盘的存在才使得相邻的椎骨可以有有限的相对运动。

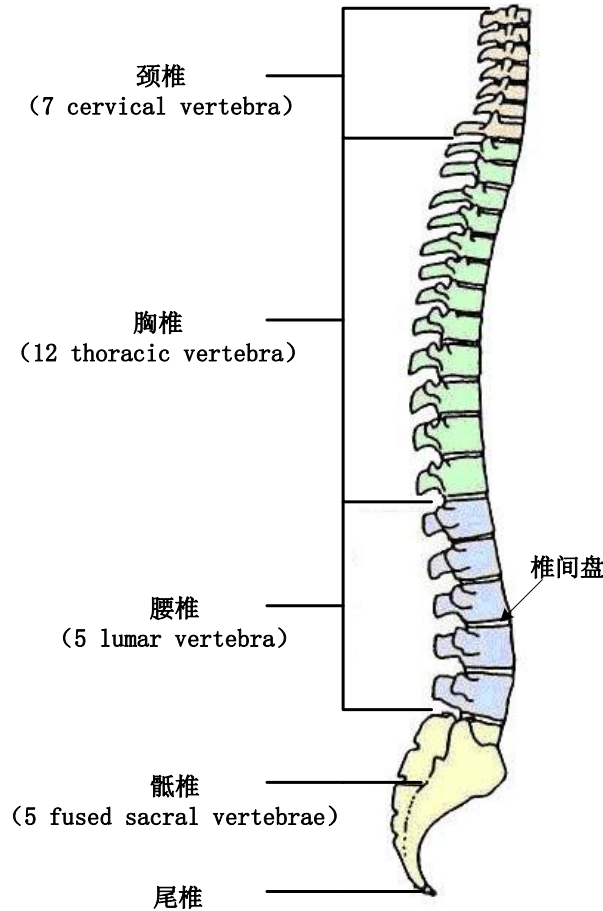


图4.1 人体脊柱左侧图
Fig. 4.1 Left lateral view of the human spine

4.2 腰椎 CT 图像显著性计算流程

受腰椎 CT 图像成像效果的限制，在对腰椎 CT 图像进行显著性计算之前要对其进行一定的预处理，整体流程如图 4.2 所示。首先对图像进行预处理，其中包括无关边缘剪切，图像增强，边缘提取以及滤波处理。对处理后的图像进行显著性计算得到其显著性图，显著图中腰椎的间盘部分会表现出较高的显著性。根据显著度就可以提取出整个图像中的间盘区域，以便进行后续的分割处理。

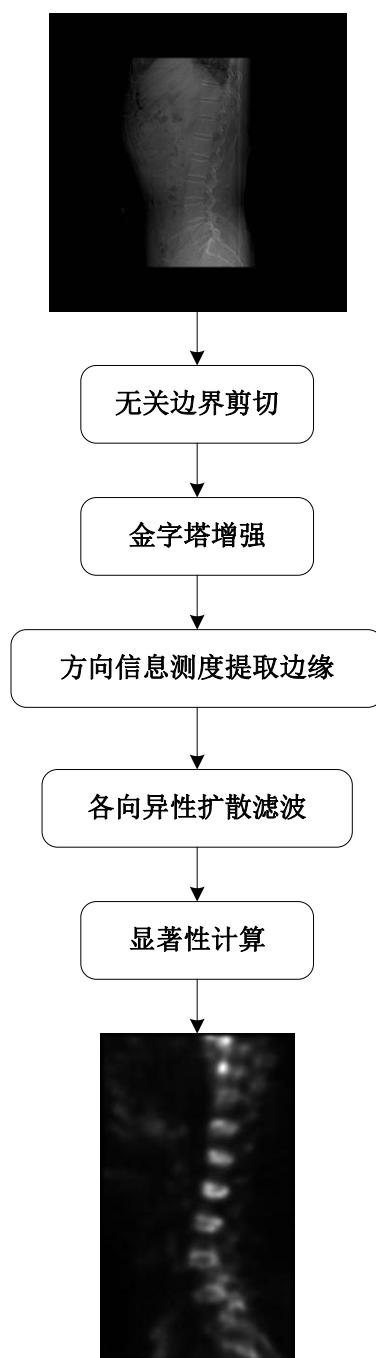


图4.2 腰椎图像显著图计算流程

Fig. 4.2 Computation flow of lumbar image's saliency map

4.3 图像预处理

每个人的每一节腰椎间盘的形态都不尽相同，加之CT图像成像过程的影响使得到的CT腰椎图像的对比度比较低，这些因素都会对图像的后续处理产生不利的影响。所以要检测出每一节间盘的位置，必须先要做好腰椎CT图像的预处理工作。原始腰椎CT图像为 512×512 的DICOM格式图像，即按照DICOM标准而存储的医学图像。DICOM文件包括一个文件头和一个DICOM数据集。文件头中包含了标识数据集的相关信息，如生成该文件的应用程序，文件传输格式等等。DICOM数据集是DICOM文件的主要组成部分，集合中除了医学图像外，还包括一些诸如图像大小，患者姓名等和医学图像有关的信息。由于受显示器显示原理的限制，对于16位的DICOM格式图像在显示时只能显示为8位的灰度级，所以这里将图像转为8位的灰度图像进行处理，转换前对DICOM图像进行窗宽窗位的调节以突出感兴趣的脊骨部分图像。如图4.3所示，为了满足 512×512 的规范，图像中包含一个黑色边界区域，这些区域与图像处理任务无关，这里将其剪掉，不仅可以提高图像的处理速度还能减少无关区域对图像处理效果的影响。接下来就对剪切后的腰椎灰度图像进行预处理，便于进行后续的间盘检测。



图4.3 调节窗口参数后的腰椎CT图像

Fig. 4.3 The lumbar CT image after the adjustment of window parameters

4.3.1 Gauss-Laplace 金字塔增强

目前，临床上常用的CT图像增强的方法有对比度拉伸和直方图均衡两种。这两种方法的局限在于只是基于直方图的变换，和像素的位置信息无关，所以只适合处理灰度级分布集中的图像。而CT图像，其细节信号（高频，例如骨纹理的细节）虽然集中在较窄的灰度范围内，但其有用的灰度范围分布较广，人眼很难分辨，所以利用传统对比度增

强和直方图均衡无法满足临床需求。综上，本文利用一种Gauss-Laplace金字塔结构多分辨率处理图像^[53]，以非线性的方式增强不同分辨率下图像的细节信息，再叠加到原尺寸就可以突出图像中的病灶细节。

基于Gauss-Laplace金字塔图像增强算法包括Gauss金字塔分解，Laplace金字塔系数调整和Gauss金字塔重建三个步骤。

在Gauss金字塔分解阶段，将原始图像作为第0层图像进行Gauss卷积核滤波。对滤波后的第0层图像进行下采样得到第1层子图像，其水平和垂直分辨率各为基准图像（即第0层图像）的1/2。再以第1层图像为基准图像继续进行Gauss滤波以及下采样，如此反复进行操作。

若共采样 n 次，则共建立 $n+1$ 层Gauss金字塔，第 n 层的子图像分辨率为原始图像的 $1/2^n$ 。接下来从Gauss金字塔的顶端即第 n 层图像 G_n 开始，对该层子图像用Gauss卷积核滤波并作上采样，形成与 G_{n-1} 具有相同分辨率的子图像 G'_{n-1} 。 G_{n-1} 与 G'_{n-1} 之间的差值就是第 $n-1$ 层的Laplace子图像 L_{n-1} 。

以此类推至金字塔的第0层图像可得到Laplace金字塔。由Gauss的滤波特性可知，Laplace子图像可以反映出图像的细节信息。

Laplace金字塔系数调整阶段通过调整Laplace金字塔系数来增强各个尺度上的结构特征，由此可形成一个目标边界。本文使用如式(4.2)所示的函数 $h(\cdot)$ 对Laplace金字塔系数进行调整。

$$h(|x|) = \begin{cases} \frac{k}{|x|^p} + 1; & |x| \leq 1, p \in [0, 0.5] \\ k + 1; & |x| > 1 \end{cases} \quad (4.2)$$

式中描述的增益调整函数通过乘以 $k+1$ 来增强绝对值大于1的灰度值，对于绝对值小于1的灰度值则将调整函数用抑制因子 $1/|x|^p$ 进行调整，调整公式如式(4.3)所示。

$$L'_j(x) = h(x) \cdot L_j(x) \quad (4.3)$$

Gauss金字塔重建阶段，把增强的Laplace子图像 L'_j 与同层的Gauss子图像相加进行图像重建。由于每个层次上都通过调节系数增强幅度，最后加至Gauss图像底层即原始图像时，图像的细节信息已经充分的显示出来。图4.4所示为腰椎CT图像增强前后对比图，其中图(a)为由DICOM图像调整窗宽窗位后剪去无关边界后的灰度图像，其中L1~L5为5节腰椎骨，椎体之间的密封体即为椎间盘，椎间盘在图中表示为上下两条不闭合的边缘。图(b)为金字塔多尺度增强后的图像。通过对比发现原图中不明显的椎体边缘细节在处理有了明显的增强。

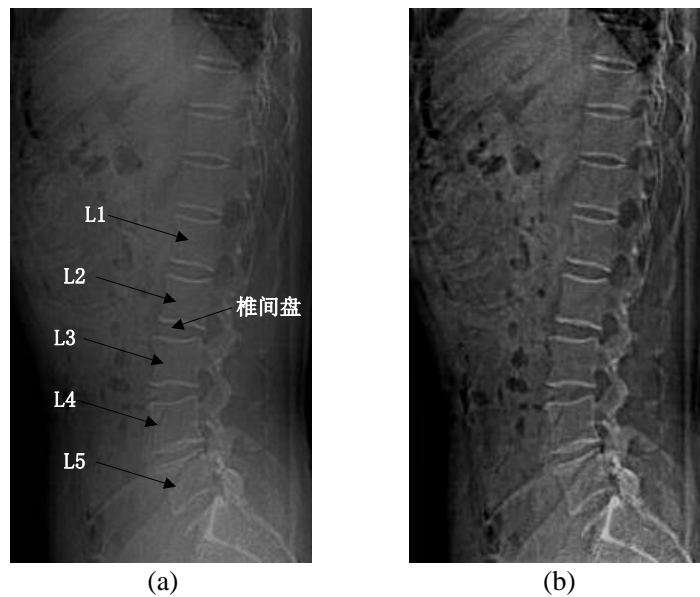


图 4.4 腰椎图像增强前后对比图

Fig. 4.4 Comparison between the original image and the enhanced image

4.3.2 基于方向信息测度的边缘提取

经过金字塔多尺度增强后虽然图像中的细节得到了一定的增强，但是并没有从背景图像中分离出边缘或者目标图像。图像中的边缘是图像中的细节部分（高频信息），表现为灰度突变部分，对视觉有很重要的作用。图像中的边缘信息不仅具有局部灰度突变的极大值，同时还具有很强的方向性，而噪声却是随机的。基于这种思想，本文采用抗噪性能较强的图像方向信息测度(orientation information measure, OIM)来提取腰椎 CT 图像中间盘的边缘^[54]，该方法具有很强的抗噪性能。

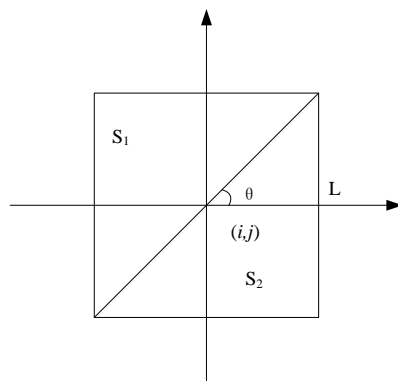


图 4.5 方向信息测度示意图

Fig. 4.5 The schematic diagram of OIM

设当前像素点坐标为 (i, j) ，其邻域为 $R = \{(m, n) | |m - i| \leq L, |n - j| \leq L\}$ ， L 表示邻域长度的一半，邻域内像素点的灰度值记为 $x_{mn}, (m, n) \in R$ ，示意图如图 4.5 所示。 l_θ 是角度为 θ 且过点 (i, j) 的一条直线，直线将邻域分为 S_1 和 S_2 。方向信息测度可由式(4.4)-(4.7)得到。

$$f_{s_1} = \sum_{(m,n) \in s_1} x_{mn}, f_{s_2} = \sum_{(m,n) \in s_2} x_{mn} \quad (4.4)$$

$$d_\theta = |f_{s_1} - f_{s_2}| \quad (4.5)$$

$$d_{\theta_{\min}} = \min_{0 \leq \theta < 180^\circ} (d_\theta), d_{\theta_{\max}} = \max_{0 \leq \theta < 180^\circ} (d_\theta) \quad (4.6)$$

$$M_{mn} = d_{\theta_{\max}} - d_{\theta_{\min}} \quad (4.7)$$

由于边缘两侧像素灰度值差异较大,因此当 l_θ 沿边缘的方向时, d_θ 最大;而当 l_θ 垂直于边缘轨迹的方向时, d_θ 最小,由式(4.7)可知,边缘处的点的 M_{ij} 的值较大。而平滑区域的像素灰度值基本相同,所以不论 l_θ 取什么方向, d_θ 数值的差异都很小,因此 M_{ij} 的值也就相对较小。对于噪声来讲,不论是平滑区的噪声还是边缘区的噪声,都是随机的且不具有方向性,所以无论 l_θ 取什么方向,两边的噪声分布都没有太大差异,因此,噪声对 M_{ij} 值的结果产生的影响很小。综上可知,方向信息测度不仅能够有效地表征图像中的方向信息,而且具有较好的抗噪能力。

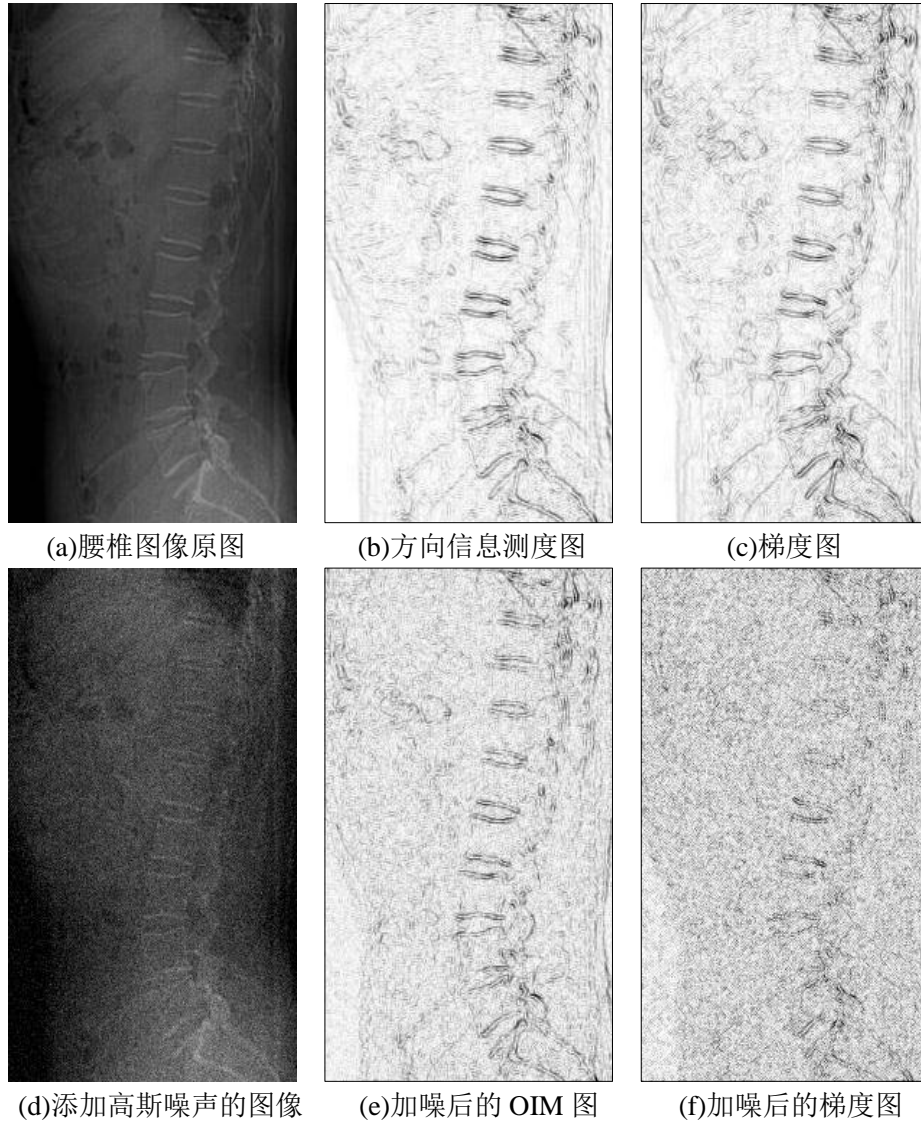


图 4.6 梯度模值与方向信息测度对比

Fig. 4.6 The comparison between gradient module and OIM

图 4.6 所示为未添加噪声图像与添加噪声图像的实验结果对比图。图(a)为腰椎图像原图，图(b)为图(a)的方向信息测度图像，图(c)为图(a)梯度模值图像，图(d)为添加均值为 0，方差为 0.005 的高斯噪声的图像，图(e)为图(d)的方向信息测度结果，图(f)为图(e)的梯度模值结果。从实验结果可以看出，添加噪声之后的图像的梯度模值结果中理论上梯度模值较高的椎骨区域已经不能从背景中分辨出来，噪声干扰很大。而方向信息测度结果虽然受到影响，但影响较小。

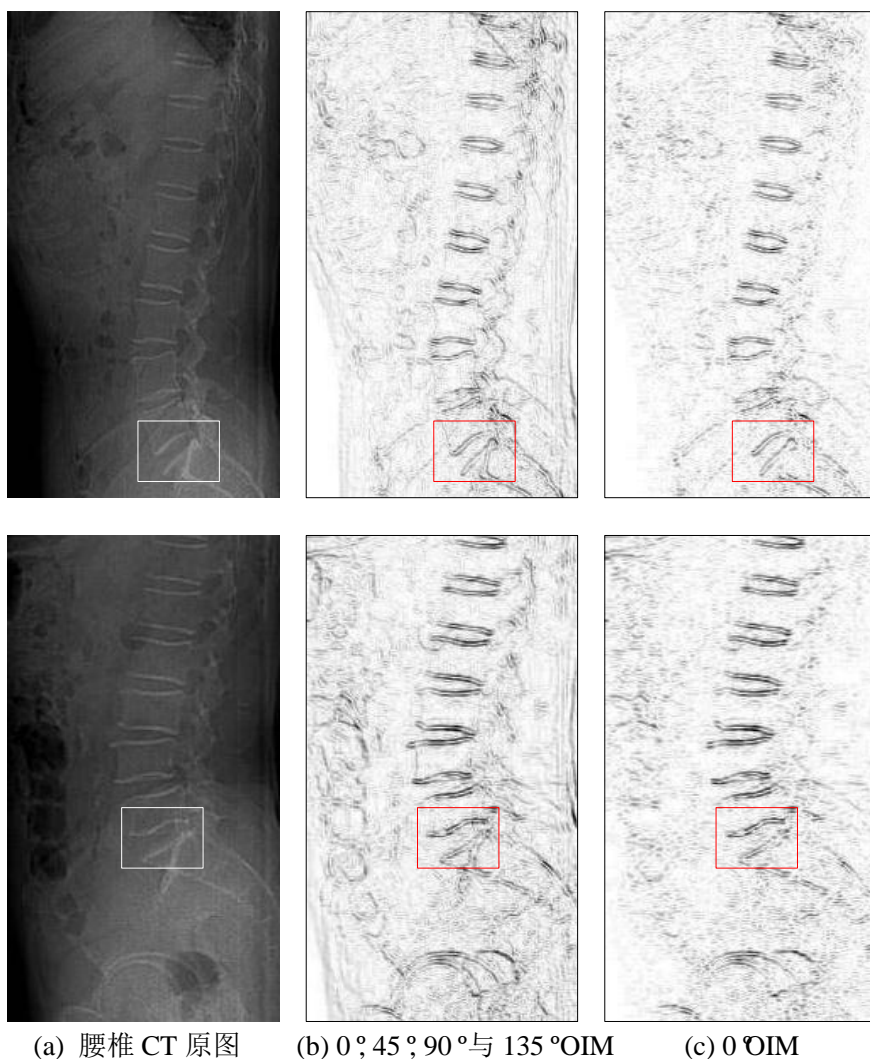


图 4.7 方向信息测度结果图
Fig. 4.7 The results of OIM

由于腰椎图像中的间盘的边缘大部分近似处于水平方向，只有最后一块间盘即最后一块椎骨与骶骨之间的椎间盘为左下倾斜，且脊柱为明显的垂直方向。所以为了准确地检测出间盘并避免脊柱对检测结果的影响，定义 0° 方向的信息测度算子，窗口大小取 3×3 。图 4.7 为两组图像的方向信息测度实验结果，其中图(a)为腰椎 CT 图像，图(b)为经过 0° 、 45° 、 90° 与 135° 四种方向信息测度计算之后得到的方向信息测度图像，图(c)为 0° 方向的信息测度图像。图中红框中所示的间盘 L5-S1 右下角有强边缘干扰，且该干扰边缘的倾

斜角度较大。若计算四种方向的信息测度图像则这种干扰会影响图像的检测以及后续的分割结果，但如果只计算 0° 方向的信息测度图像则可以有效地去除干扰。

4.3.3 基于各向异性扩散的图像滤波

虽然在进行方向信息测度计算之后提取出椎骨的边缘部分。但是由于只计算了 0° 方向的信息测度，所以软组织区域中会出现大量的噪声，这些噪声会对后续的处理产生一定的影响。所以目前要解决的是如何去除图像中产生的大量噪声，但同时又能保持图像中的边缘信息不被破坏。虽然传统的去噪方法如均值滤波、中值滤波等实现简单，但效果不理想，在平滑噪声的同时也模糊了图像细节。所以本节引入非线性扩散方程来对图像进行去噪。

4.3.3.1 PM 扩散方程

图像处理中，扩散的思想来自图像的多尺度描述。对原始图像 $I_0(x, y)$ 的多尺度滤波可由式(4.8)描述，即以时间 t 作为高斯核 G 的方差与原始图像做卷积，即可得到从高分辨率到低分辨率的图像多尺度描述。

$$I(x, y, t) = I_0(x, y) \cdot G(x, y, t) \quad (4.8)$$

1990 年, Perona 和 Malik 在文献[55]中提出了非线性扩散的偏微分方程法, 如式(4.9)所示。

$$\begin{cases} \frac{\partial I(x, y, t)}{\partial t} = \text{div}[c(\|\nabla I(x, y, t)\|)\nabla I(x, y, t)] \\ I(x, y, 0) = I_0(x, y) \end{cases} \quad (4.9)$$

其中, div 代表散度算子; ∇ 为梯度算子; $\|\cdot\|$ 表示幅度; $c(\|\nabla I\|)$ 是扩散方程, 时间算子 t 表示去噪过程与扩散持续时间相关。PM 模型的基本思想就是根据 $\|\nabla I\|$ 的大小调节扩散平滑的强度。

理想状态下, 图像的边缘区域的梯度值较大, 为保持边缘信息, 可以设置扩散方程来使模型在边缘处实现较弱的平滑; 对于梯度值较小的平坦区域则采用较大的扩散系数实现较强的平滑。根据梯度值和扩散系数的关系, Perona 和 Malik 提出了形如式(4.10)-(4.11)的两种形式的扩散方程, 其变化曲线如图 4.8 所示。

$$c_1(\|\nabla I\|) = \exp[-(\|\nabla I\|/k)^2] \quad (4.10)$$

$$c_2(\|\nabla I\|) = \frac{1}{1 + (\|\nabla I\|/k)^2} \quad (4.11)$$

式中的 $\|\nabla I\|$ 可视为边缘检测器, k 可被视为扩散门限。如果 $\|\nabla I\|$ 远远大于 k , 那么 c 的函数值趋于 0, 扩散被抑制; 如果 $\|\nabla I\|$ 远小于 k , 则 c 的函数值趋于 1, 扩散则被增强。为区别于无方向各向同性扩散方程, 称各个方向上扩散速度不同的扩散方程为各向异

性扩散方程^[64]。

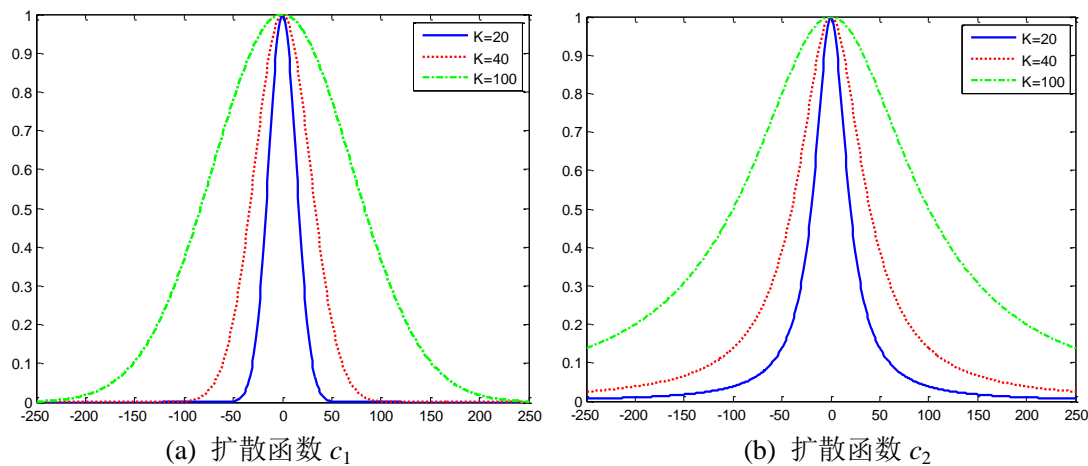


图 4.8 扩散函数随 $\|\nabla I\|$ 的变化曲线

Fig. 4.8 The curves of diffusion function

对于图像处理，则需将式(4.8)所示的偏微分方程离散化之后方可应用。Perona 和 Malik 提出的模型离散表达式如式(4.12)所示。

$$I_p^{t+1} = I_p^t + \frac{\lambda}{|\eta_p|} \sum_{q \in \eta_p} c(\nabla I_{p,q}^t) \quad (4.12)$$

其中， λ 是控制扩散总体强度的常数， I_p^t 为当前图像的离散采样， p 为像素的二维坐标， η_p 表示像素 p 的邻域， $|\eta_p|$ 为邻域空间的大小。除图像边缘外，通常情况下邻域空间的大小为 4。PM 模型中，梯度值线性近似为式(4.13)。

$$\nabla I_{p,q}^t = I_p^t - I_q^t, \forall q \in \eta_p \quad (4.13)$$

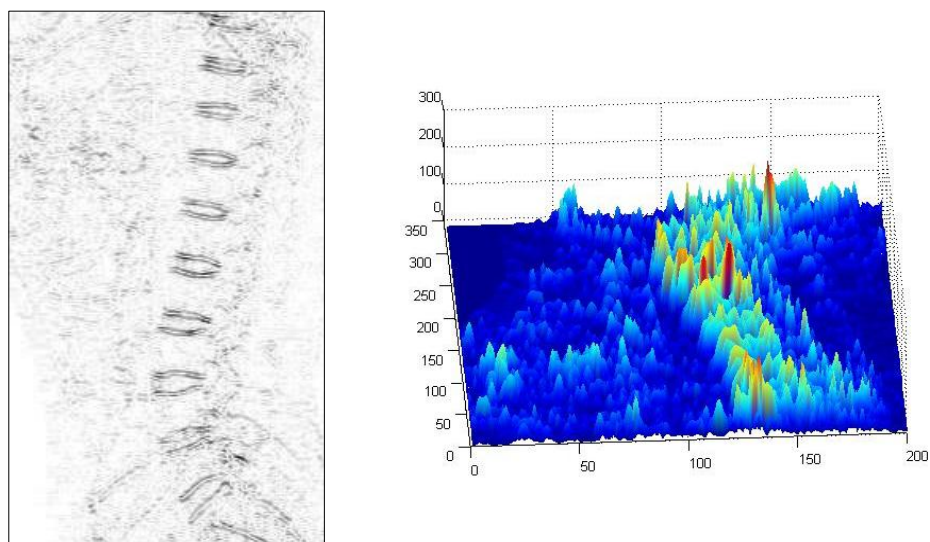
4.3.3.2 实验结果对比

在 0° 方向信息测度的腰椎图像结果中，除了有明显的间盘边缘存在之外还有软组织区域的大量非边缘区域被提取出来，这些细小的类似高斯噪声的成分会对接下来的处理产生一定的影响。所以为了使待提取的间盘边缘区域不被平滑，下面引入 PM 扩散方程对 0° 方向信息测度的腰椎结果图像进行去噪。

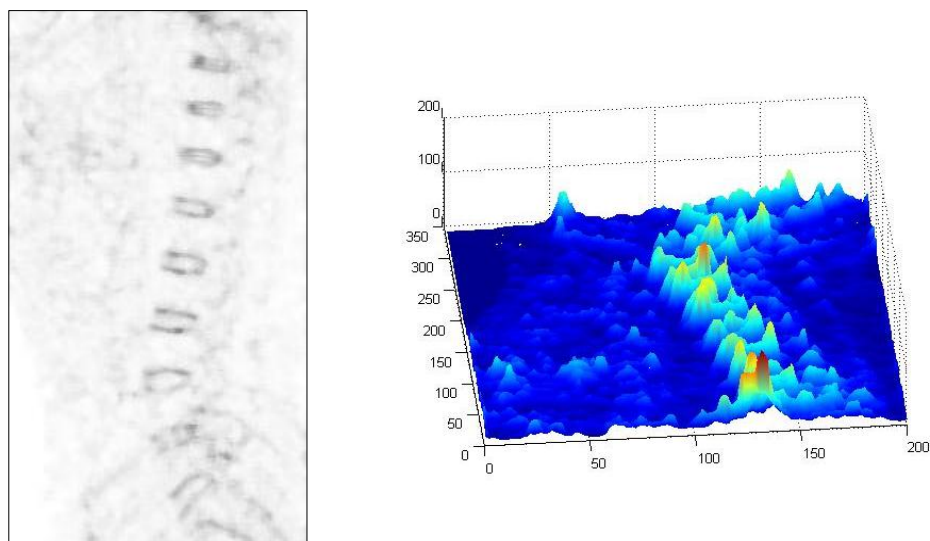
图 4.9 所示为方向信息测度结果图及其中值滤波与各向异性扩散滤波结果对比。从图(a)中可以看出， 0° 方向的方向信息测度结果图中间盘区域表现出明显的边缘，但是软组织区域中存在大量的类噪声的细小目标干扰。

图(b)为对方向信息测度图进行中值滤波的结果，图中可以看出虽然滤波对大部分干扰进行了很好的平滑，但是同时也破坏了间盘区域的边缘信息。

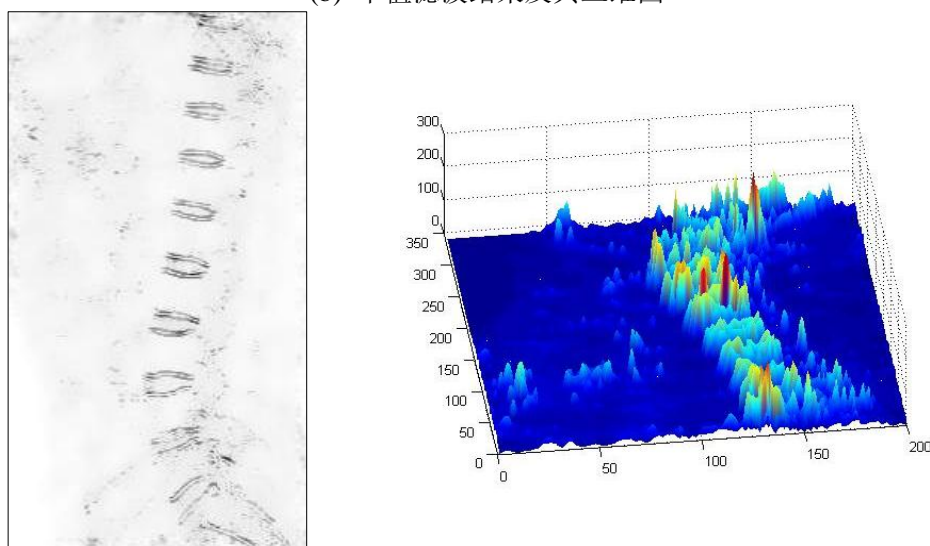
而图(c)为对方向信息测度图进行各向异性扩散滤波的结果，该图中的软组织干扰得到了较好的抑制，同时也很好地保持了边缘信息。



(a) 0°方向的方向信息测度图及其三维图



(b) 中值滤波结果及其三维图



(c) 各向异性扩散滤波结果及其三维图

图 4.9 各向异性扩散与中值滤波结果比较

Fig. 4.9 The results of anisotropic diffusion compare with median filter

4.4 基于显著性检测的间盘粗提取

经过图像增强与边缘提取之后的图像已经提取出边缘部分，虽然经过滤波后已经抑制了大部分细小目标的干扰，但通过各向异性扩散的原理可知，利用各向异性扩散滤波只能去除细小目标的干扰，而对于稍大些的目标由于其边缘感明显则会将其保留。这些保留下来的稍大些的目标仍然会给间盘分割带来一定的困难。这些干扰大多分布在除椎骨以外的其他组织区域，如果分割时能忽略这些干扰区域而只针对椎间盘区域进行分割则分割效果就不会受到这些干扰区域的影响。基于此，本文采用第三章提出的基于小波变换的视觉显著性模型来检测在随机噪声中的较为显著的边缘信息。椎体边缘即椎间盘具有较高的亮度且在复杂背景中有特殊的方向性，这些特性使其在显著图中表现为显著度较高的显著区域，这样就可以通过注意力选择的显著性算法对间盘区域进行粗提取。

本小节就通过利用第三章提出的显著性计算方法对腰椎 CT 图像中的椎间盘进行粗定位。由于腰椎 CT 图像均为灰度图像，所以不需要考虑颜色分量的处理，只需提取灰度分量即可。对原 CT 图像进行一系列预处理后得到 0° 方向的方向信息测度结果图，对其进行显著性检测，结果如图 4.10 所示。图中可以看出在腰椎 CT 图像中的腰椎间盘部分显著度较高，还有一些高亮的边缘也有着一定的显著性，除此之外的其他组织部分的显著性几乎为 0，这样就除去了其他非目标区域的干扰。

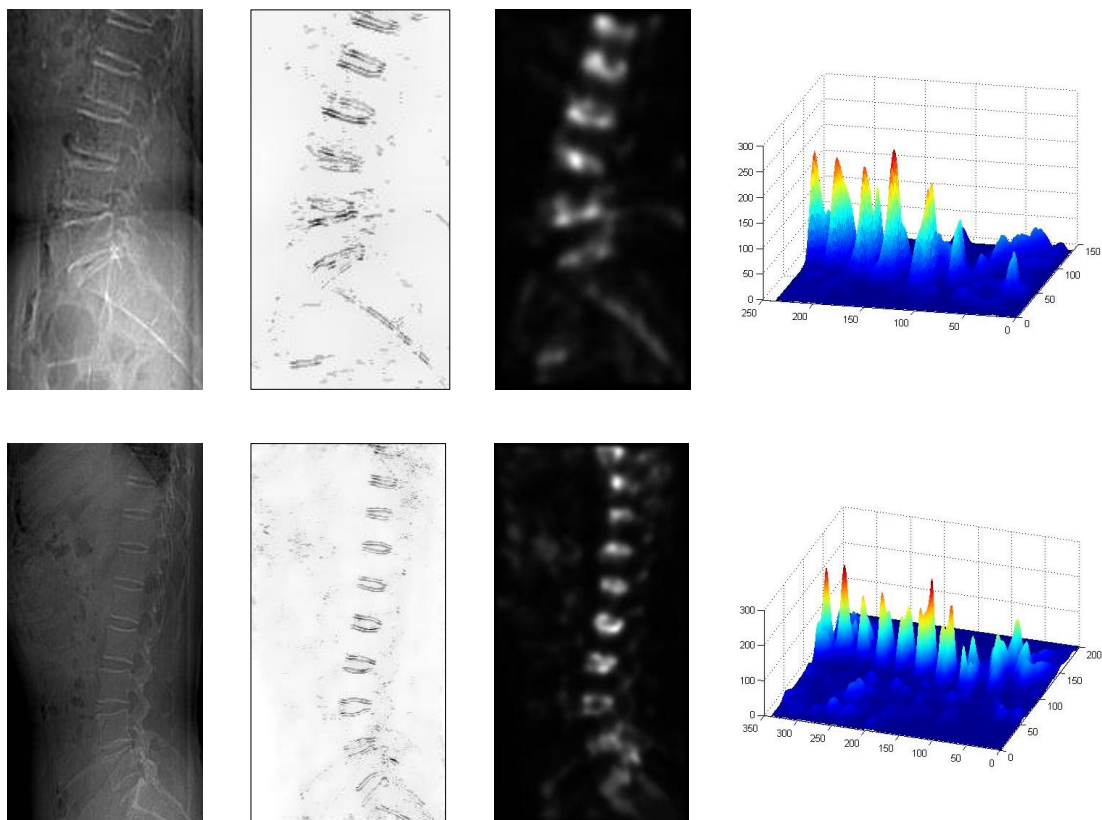


图4.10 腰椎CT图像的显著图及其三维显示

Fig. 4.10 Saliency map of lumbar CT image and its 3D display

4.5 本章小结

本章主要实现了腰椎间盘的粗提取,为后续的间盘分割做准备。首先对原始DICOM格式的腰椎CT图像进行窗口调节以便获得更清晰的椎骨区域,为方便处理将其另存为灰度图像。然后剪去图像中无关的边界部分,只保留有实际意义的人体成像部分图像。针对CT图像中对比度较小,我们采用了一种多尺度技术来对图像进行图像增强。由于腰椎图像中腰椎表现为近似的水平方向,所以对增强后的图像进行了 0° 的方向信息测度的计算,计算结果图中间盘边缘信息被很好的提取出来。但是在软组织区域会出现一些类噪声的细小目标干扰。对这些干扰采用各向异性扩散进行滤波,滤波结果在去除细小目标干扰的同时也保持了边缘信息。最后对滤波得到的结果图像进行了显著性计算,得到的显著图中间盘区域由于具有明显的方向性而表现出了较高的显著度。根据显著度的高低便可提取出间盘的大概位置,大大方便了后续的处理。

第 5 章 基于 Hough 森林的间盘分割

目前医学图像的分割领域中基于形变模型的分割方法应用较多，并且在一定程度上获得了成功。但对于腰椎 CT 图像分割椎间盘的任务来说，形变模型并不适用，因为腰椎 CT 图像中的间盘在图像中表现的并不是闭合的轮廓，而是通过上下椎骨的边界体现出来的，这就无法控制闭合曲线使其在到达椎间盘的边缘时停止演化。本章在上一章的显著性计算的基础上进行间盘区域精提取，进而进行区域分割得到间盘轮廓。

5.1 间盘分割整体流程

上一章中已经利用 WFPE 显著性计算方法得到了腰椎图像的显著性图，但是显著图只能作为腰椎间盘位置的粗估计，并不能据此获得间盘的准确位置，这对于间盘分割来说这是远远不够的。针对这一问题，本章提出了基于显著性计算结果与 Hough 森林的腰椎间盘检测分割算法。算法中采用先利用显著图对腰椎图像中的间盘区域进行粗提取，然后利用 Hough 森林对间盘区域图像进行精提取，最后对间盘区域进行区域分割的思想来对腰椎 CT 图像进行分割。

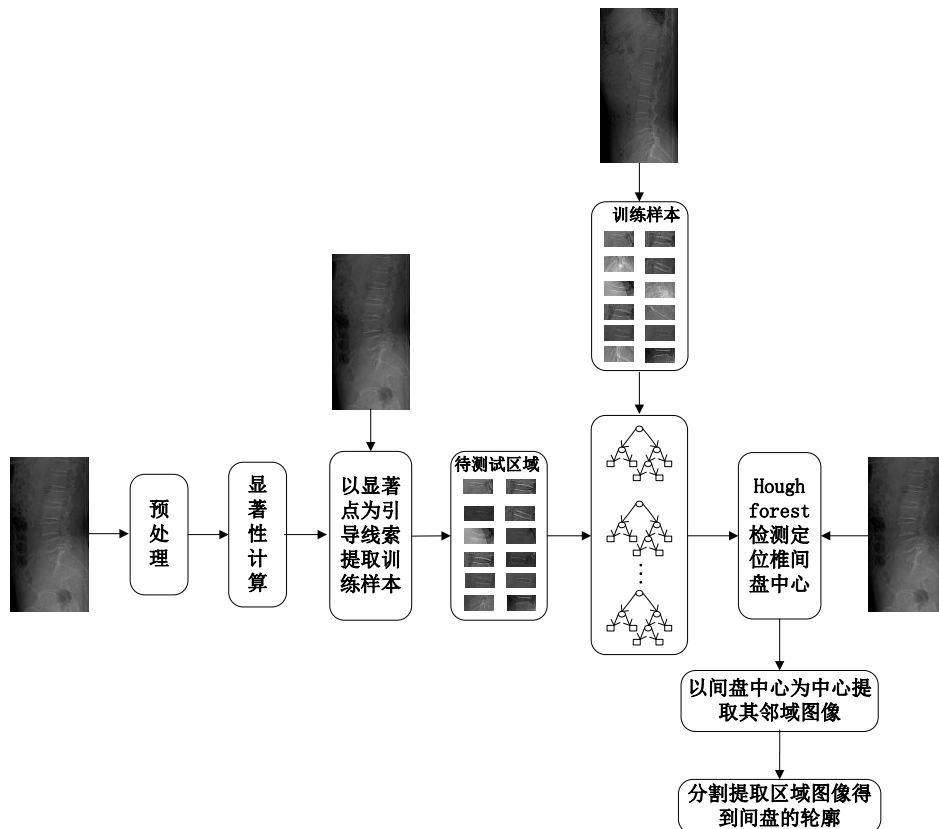


图 5.1 间盘分割的整体流程

Fig. 5.1 The whole flow of lumbar disk segmentation

如图 5.1 为间盘分割的整体流程图。首先提取腰椎图像中的间盘区域图像块作为正训练样本,提取非间盘区域图像块作为负训练样本,由正负训练样本训练生成随机森林。然后以腰椎图像显著图中各点的显著度为引导提取图像中显著度较高区域即间盘区域作为待检测间盘区域图像进行间盘中心检测,经由随机森林对间盘中心投票的 Hough 图像结果中的最大值点为间盘中心位置。最后利用中心位置坐标并结合先验知识提取更精确的间盘区域图像进行区域分割得到间盘轮廓。

5.2 基于 Hough 森林的间盘检测

本节要解决的腰椎间盘检测问题属于基于图像的目标检测这一研究领域。作为计算机视觉领域的重要研究内容,目标检测的研究一直处于很活跃的状态,因此关于目标检测的算法也是多种多样的。其中,隐式形状模型(ISM)就是流行的方法之一,还有许多基于隐式形状模型的改进方法。Gall 于 2009 年提出的基于 Hough 森林的目标检测就是隐式形状模型(ISM)的一种改进^[56]。在此基础之上,本文引入 Hough 森林来进行间盘的检测定位。

基于 Hough 森林的间盘检测就是借助随机森林来完成广义 Hough 变换进而检测出间盘具体位置。随机森林(RF)是一种统计学习理论,它采用 bootstrap 重抽样方法抽取原始样本中的多个样本,对每个 bootstrap 样本建立决策树,然后将多棵决策树的预测进行组合,最后由决策树的投票得出预测结果。随机森林的预测准确率很高,不容易出现过拟合现象,而且对异常值与噪声有着很好的容忍度^[57-58]。而广义霍夫变换是一种用于区域边界形状描述的方法,它可以用来检测任意形状^[59]。

本节将介绍基于 Hough 森林的腰椎间盘检测方法,该方法建立随机森林在图像块的外观与 Hough 投票之间建立映射关系。首先通过学习间盘图像块的显著性外观和位置来构建随机森林,然后用测试间盘图像块向待检测的间盘中心进行概率投票,最后由所有的关于间盘中心的投票结果构成二维 Hough 图像,该 Hough 图像中的局部极大值位置即为待检测间盘的中心位置^{[56][60]}。得到间盘区域的中心位置之后可以根据先验知识提取间盘区域进行分割。

5.2.1 构建 Hough 森林

随机森林(RF, Random Forests)是由 Leo Breiman^[61]于 2001 年提出的,它的基本思想是将很多弱分类器集成为一个强分类器^[62]。构成随机森林的基础分类器是决策树,决策树是一个由节点和有向边组成的树状预测模型,如图 5.2 所示。树中包含根节点、内部节点和叶子节点三种节点。其中根节点只有一个,是全体训练数据的集合;每个内部节

点即非叶子节点都是一个弱分类器，它将到达该节点的样本按某个特定的属性进行分类效果最好的分类；每个叶子节点是带有分类标签的数据集合。从决策树的根节点到叶子节点的每一条路径都是一个分类形成的过程。多个决策树分类器构成一个随机森林分类器，随机森林的决策结果就是所有决策树分类结果的平均。以一个 M 类的分类问题为例，一个样本 p 通过每个决策树分类器 $T_n (n=1, \dots, N)$ 进行分类，输出结果为 M 个置信度，每个置信度 $p_{(n,p)}(f(p)=c)$ 代表该样本 p 属于第 c 类的概率，其中 $c \in \{1, 2, \dots, M\}$ ，最后随机森林的决策结果就是所有决策树结果的平均，如式(5.1)所示。

$$F(p) = \arg \max_c p_c(p) = \arg \max_c \frac{1}{N} \sum_n p_{(n,p)}(f(p)=c) \quad (5.1)$$

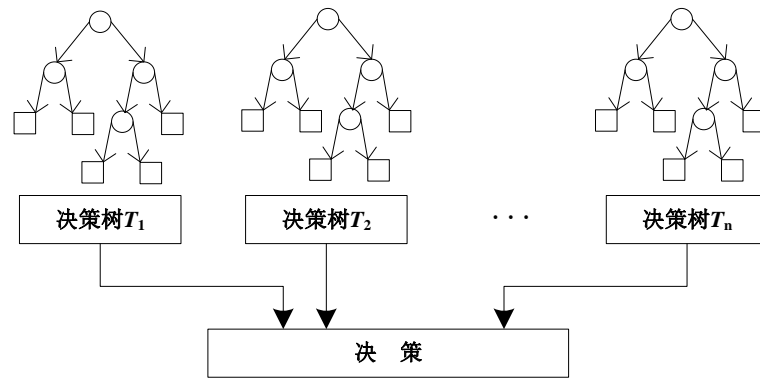


图5.2 随机森林示意图

Fig. 5.2 The schematic diagram of random forest

对于Hough森林来说，每一棵随机树 T 由样本图像块集合 $\{P_i = (I_i, c_i, \vec{d}_i)\}$ 构成，其中 i 为图像块的索引。样本图像块有正负样本之分，对于腰椎图像来说，间盘区域图像块为正样本，非间盘区域图像块为负样本。其中， $I_i = (I_i^1, I_i^2, \dots, I_i^M)$ 为图片块的外观， I_i^j 是第 j 个特征通道的特征向量， M 是通道的数量； $c_i = \{0, 1\}$ 为图像块的类别标记，若图像块为负样本，则标记为0；否则为正样本，标记为1； \vec{d}_i 为正样本图像块的偏移量，即该样本在原图像上对应区域的中心点到达该样本从属的间盘目标的中心点的位移，对于负样本来说， \vec{d}_i 未定义。

Hough森林中的每一棵随机树的构建都是在随机森林的框架下完成的，且都是从根节点开始递归构建的。根节点由全体训练图像块集合组成，针对根节点中的数据集合选取一个最优的弱分类器对其进行分类，然后根据对图像块的外观的测试结果把图像块划分到对应的子集合中，进而得到内部子节点。反复执行该过程，直至随机树达到最大深度（设为15）或所属节点的图像块数量小于最低阈值（设为20），此时该节点就变为叶子节点不再继续分裂。训练过程中，非叶子节点对应的关于图像块外观 I_i 的弱分类器有多种定义方式。训练与测试的过程中，从样本中取得的图像块的大小都是固定的（设为 16×16 ）。这个针对图像块外观的弱分类器可以由一个特征通道 $j \in \{1, 2, \dots, M\}$ ， 16×16 的

图像块中的两个位置 (p, q) 和 (r, s) ，以及一个阈值 τ 来定义。具体公式如式(5.2)所示，这个弱分类器是通过比较一对像素点在同一特征通道上的值得大小来实现分类的。

$$t_{j,p,q,r,s,\tau}(I) = \begin{cases} 0, & I^j(p, q) < I^j(r, s) + \tau \\ 1, & \text{otherwise} \end{cases} \quad (5.2)$$

随机树训练的同时也将在叶子节点 L 中存储目标图像块（正样本）的比例 C_L （若 C_L 取值为1，说明到达该节点的图像块全部来自于目标图像）以及对应目标图像块的偏移量列表 $D_L = \{\vec{d}_i\}$ 。于是，这些叶子节点就构成了一个带有目标中心估计位置信息的码本。

对于非叶子节点中的弱分类器的性能优劣的评价方法是Hough森林的关键。因此，我们要选取一个弱分类器，使得按照该分类器分类得到的两个样本集合的类别标记和偏移量的不确定性达到最小。为定量测量某个样本集合的类别的不确定度，定义样本集合 A 的类别标记不纯度为 $U_1(A)$ ，数学定义如式(5.3)所示。

$$U_1(A) = -|A| \cdot (C_L \cdot \log C_L + (1 - C_L) \cdot \log(1 - C_L)) \quad (5.3)$$

其中， $|A|$ 表示集合 A 中的样本图像块的数量。同样，偏移量不确定度对应测量偏移量 \vec{d}_i 的不纯度，定义如式(5.4)所示，其中 \vec{d}_A 是偏移量列表 D_L 中的所有偏移量的平均值。最后，我们按照式(5.5)选择两种不确定度之和最小的二进制测试。

$$U_2(A) = \sum_{i:c_i=1} (\vec{d}_i - \vec{d}_A) \cdot (\vec{d}_i - \vec{d}_A) \quad (5.4)$$

$$\arg \min_k [U_*(A_l^k) + U_*(A_r^k)] \quad (5.5)$$

其中， A_l^k 和 A_r^k 是根据第 k 个测试结果所划分的两个集合，*表示随机选取的1或2。

5.2.2 腰椎间盘突出检测

为了在测试腰椎CT图像中定位腰椎间盘的目标，将测试图像划分为若干个互相重叠的图像块（同样设为 16×16 ）。这些 16×16 的图像块中的每一块访问随机森林之后，对测试图像中腰椎间盘的中心位置进行概率投票，累积所有的投票结果可构成Hough图像，其局部极大值为间盘的中心位置，检测流程图如图5.3所示。

假设 $P(y) = (I(y), c(y), \vec{d}(y))$ 是位于测试图像块中的中心位置为 y 的图像块， $E(x)$ 表示目标中心位于点 x 的随机事件。接下来我们感兴趣的是概率 $p(E(x)|I(y))$ 的计算。由于要完成的间盘目标检测的任务，所以这里的测试图像全部作为目标图像来对待进而检测目标，故图像块的 $c(y)$ 为1。概率 $p(E(x)|I(y))$ 的计算可由式(5.6)获得，而式中的前后两项均可由腰椎间盘突出图像的Hough森林获得。

$$\begin{aligned} p(E(x) | I(y)) &= p(E(x), c(y) = 1 | I(y)) \\ &= p(E(x) | c(y) = 1, I(y)) \cdot p(c(y) = 1 | I(y)) \\ &= p(\vec{d}(y) = y - x | c(y) = 1, I(y)) \cdot p(c(y) = 1 | I(y)) \end{aligned} \quad (5.6)$$

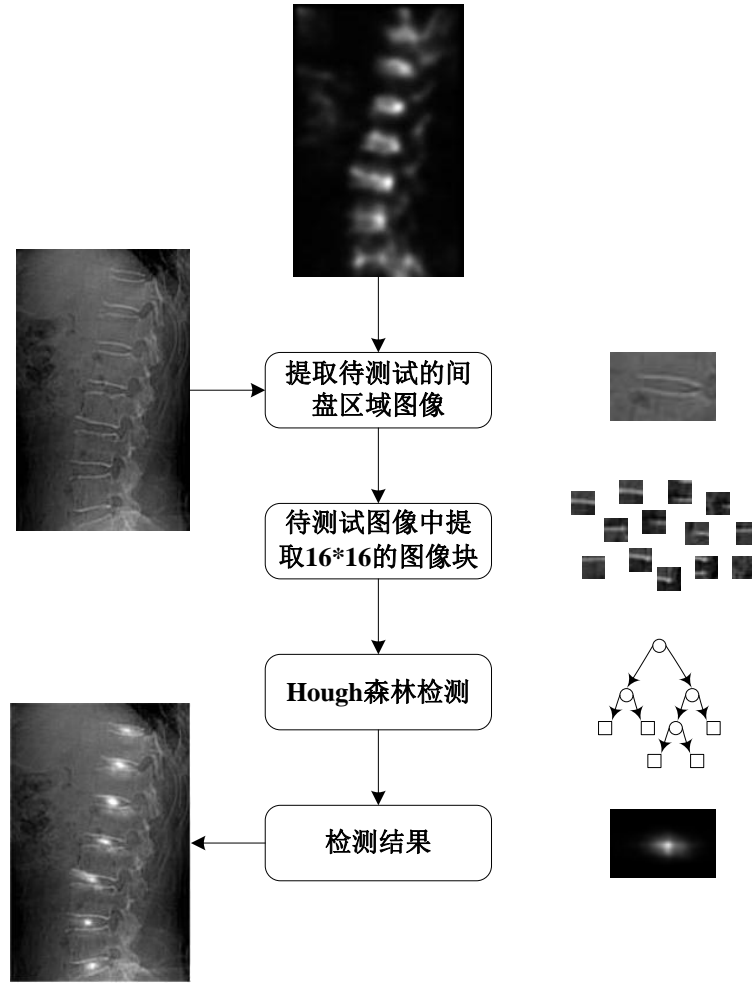


图5.3 间盘中心检测流程图

Fig. 5.3 The flow chart of disk center detection

设随机树 T 以叶子节点 L 结束，则式(5.6)中的第一项 $p(\vec{d}(y) = y - x | c(y) = 1, I(y))$ 可由基于叶子节点中偏移量集合 D_L 的Parzen窗估计来近似，而第二项 $p(c(y) = 1 | I(y))$ 直接由腰椎间盘目标图像块的比例 C_L 来估计。综上，概率估计如式(5.7)所示。

$$p(E(x) | I(y); T) = \left[\frac{1}{|D_L|} \sum_{d \in D_L} \frac{1}{2\pi\sigma^2} \exp\left(-\frac{\|(y-x) - \vec{d}\|^2}{2\sigma^2}\right) \right] \cdot C_L \quad (5.7)$$

其中， σ^2 为高斯Parzen窗的方差。该图片块访问完Hough森林中的所有树 $\{T_t\}_{t=1}^N$ (N 为森林中所有树的总棵数) 之后，整个森林中所有的树给出的概率投票的均值为该测试图像块的概率投票，如式(5.8)所示。

$$p(E(x) | I(y); \{T_t\}_{t=1}^N) = \frac{1}{N} \sum_{t=1}^N p(E(x) | I(y); T_t) \quad (5.8)$$

最后累加每一个位置 y 的概率投票并以二维Hough图像的形式输出，Hough图像中最大值点即为间盘中心点所在位置。

5.2.3 间盘检测结果

为了验证算法的有效性,我们进行了一系列的实验,实验所用的数据库来自东软飞利浦有限公司CT研发部,共包含100个患者的扫描数据。随机选取400个 40×20 的间盘图像块,每幅图像块上再取15个 16×16 的小块作为训练正样本,选取200个 60×30 的非间盘图像块并在每幅图像上取30个 16×16 的小块作为负样本。

由于间盘图像的信息并不丰富,我们选取灰度和梯度特征作为图像块的外观。训练时先随机选取正负训练样本建立5棵树,再用更难区分的正负样本建立5棵树,如此继续,直到有15棵树时停止,这样就建立了一个15棵树的Hough森林。

在通过上一章的腰椎图像的显著性计算之后,利用显著性信息我们可以自动提取腰椎图像中包含腰椎间盘的图像块来进行测试,这样避免了对非腰椎部分的检测,提高测试效率的同时也降低了误检率。

根据经验可知,像素意义下间盘区域的大小为 35×18 ,但间盘L5-S1的高度约为27,为了使取得的图像块中包含完整间盘区域,取测试图像块大小为 80×60 ,取测试样本块时仍然取 16×16 大小,检测结果如图5.4所示。

图5.4中列举了九幅有代表性的实验结果图像。从图中所示实验结果来看,算法对图(b),图(d),图(i)中的L5-S1间盘即最后一节椎间盘造成了漏检,如图中红色实线框所示。

对于图(b)中的L5-S1间盘漏检的情况与图像本身有关,图中骶骨S1的上边缘很不明显造成了图中的间盘靠肉眼也几乎不能识别,所以检测算法在此失效,对于图(b)中的点划线框中髋关节的定位点为误检情况,之所以造成误检主要是由该区域与间盘有一定的相似性,且在显著图中该区域的显著度较高导致该区域被作为待检测区域提取出来进行间盘检测。

对于图(d)中的L5-S1漏检的主要原因仍然与图像本身有关,但与图(b)不同的是,该椎间盘的上下边缘的成像效果都较好,但椎骨中间出现一个类椎骨边缘干扰,干扰造成了间盘的异样从而导致检测失败。

图(f)中的L5-S1间盘虽然成像没有问题,但是倾斜角度与其它间盘差异较大,导致该间盘漏检。而图(f)中的间盘L5,图(e)中左上角以及骶骨左下角的类椎骨边缘区域(如图中虚线框所示)虽然对识别会产生一定影响但并没有造成误检,主要是因为进行测试之前利用显著性算法先对间盘区域进行了粗提取然后再对提取出的间盘区域进行检测定位,这样就大大降低了误检类椎骨边缘干扰的几率。

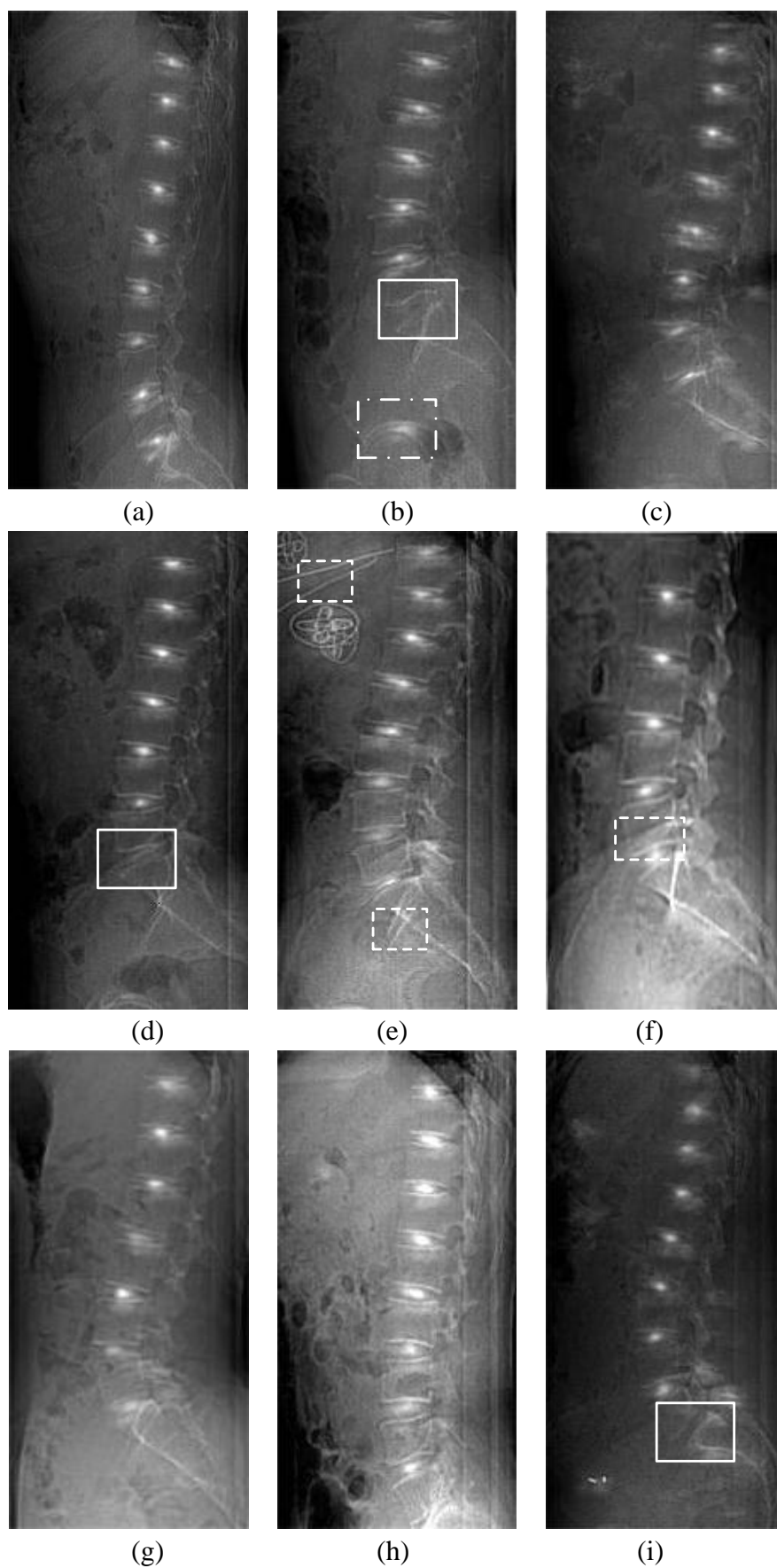


图5.4 间盘定位结果图

Fig. 5.4 The result images of disk location

5.3 基于 Otsu 的间盘区域二值分割

本小节在上一节间盘定位的基础上基于整体分割法只对腰椎 CT 图像中的间盘区域进行区域分割，这样就避免了无关的软组织区域对分割结果的影响。上一节已经利用 Hough 森林定位出间盘的中心位置，利用这个中心位置在各向异性扩散结果图像中提取出间盘区域 ROI，在 ROI 中对图像进行二值化。由于间盘是由两条边缘体现出来的，所以二值化后的 ROI 中的间盘部分仍然是两条边缘，接下来对二值化后的图像进行区域填充，之后就可以利用 canny 算子得到间盘区域的边缘。整体分割流程如图 5.5 所示。

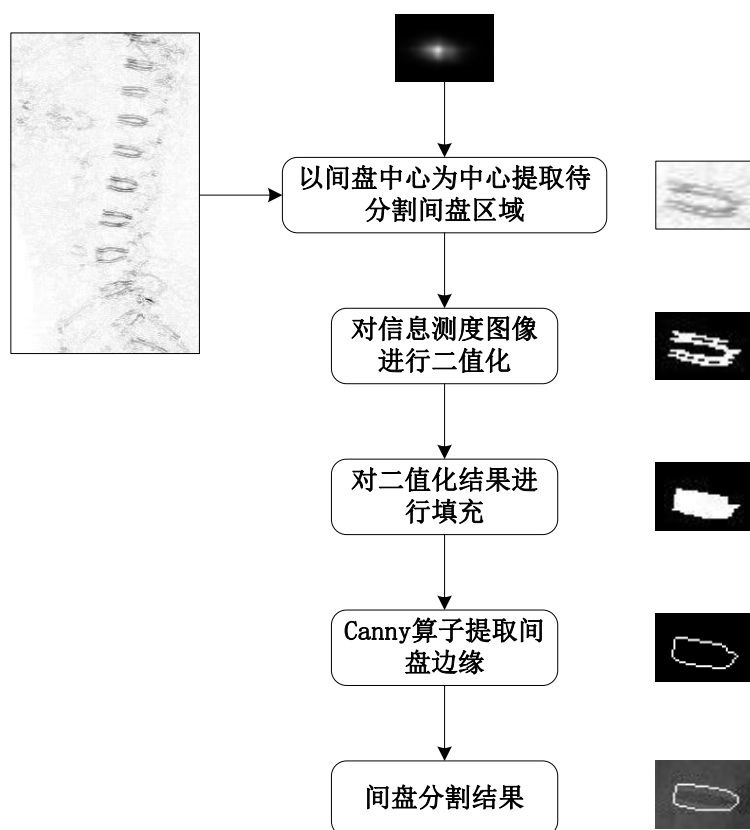


图 5.5 腰椎间盘分割流程图

Fig. 5.5 The flow chart of lumbar disk segmentation

5.3.1 提取间盘区域

上一节中已经通过 Hough 森林得到间盘的中心点位置，对于间盘区域的选取应该以定位的中心点为中心取得。计算得到 Hough 图像之后依次计算上下两个中心点 C_1 与 C_2 ，设其坐标分别为 (x_1, y_1) 与 (x_2, y_2) ，两点间的垂直距离为 $|y_2 - y_1|$ 。根据经验可知，正常情况下上下两个间盘的距离约为 40 个像素点，算法中将这个距离阈值设为 50 个像素点。若 $|y_2 - y_1|$ 小于 50 则判断为无漏检情况存在，若 $|y_2 - y_1|$ 大于 50 则判断两个间盘中间有漏检的间盘存在，取两点 (x_1, y_1) 与 (x_2, y_2) 的中心点为 (x_3, y_3) ，设 (x_3, y_3) 为漏检间盘的中心点。

由医学知识可知，由于每一节椎骨要承受着该椎体上部的躯干的重量，所以间盘的横截面大小从颈部到腰部逐渐增大，腰椎间盘的厚度为 8mm-10mm。因此我们取最后一节间盘 L5-S1 的平均大小为间盘区域大小。像素意义下，正常成人的间盘 L5-S1 的厚度为 18-23 像素宽度，宽度为 28-35 像素宽度，为了使提取的区域尽可能包含整个间盘我们取间盘区域大小为 40×30 ，并取两间盘中心点连线的垂直方向与水平方向的夹角为间盘倾斜角度。如图 5.6 所示，取中心点 O_i 与前一节间盘中心点 O_{i-1} 连线的垂直方向 l 与水平方向的夹角为此间盘的倾斜角度 θ 。

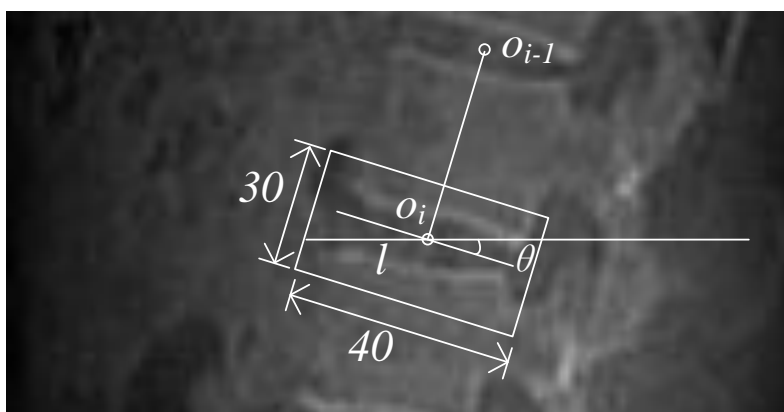


图 5.6 间盘区域提取示意图

Fig. 5.6 The schematic map of the disk region extraction

5.3.2 间盘区域分割

提取出待分割的间盘区域之后就可以通过只分割该区域进而获得腰椎间盘。由于间盘只有上下两条边缘构成且边缘不闭合，本文首先采用二值化的方法得到二值图像然后对二值图像进行填充处理后可以得到间盘的轮廓线，进而分割出腰椎间盘。由于图像二值化过程会丢失原图中许多有用信息，因此在二值化过程中，阈值的选取是否合适非常关键。尤其是对于实际应用中，如何自适应的选取阈值来二值化是完成图像处理的重要步骤。本文利用最大类间方差法进行二值化计算。

Otsu 于 1979 年提出的最大类间方差法，是一种经典的非参数、无监督自适应阈值方法，是一种直方图技术^[63]。对于图像中目标和背景差别很大的理想情况，图像的直方图中分别代表目标和背景的两个波峰之间有一个明显的波谷，谷值即为最优阈值。但实际情况中，受各种因素影响并不是每一张图片的直方图都会有明显的波谷存在，很难通过直方图来确定最佳阈值。Otsu 法就是针对这种情况提出的，它的无需其它先验知识的特点使其至今仍是最常用的二值化方法之一。算法的基本思想是：设阈值将图像分割成两组，一组灰度对应目标，另一组灰度对应背景，则使得这两组灰度值的类内方差最小，类间方差最大的阈值为最佳阈值。

设 $f(i, j)$ 为 $N \times M$ 图像 (i, j) 点处的灰度值，灰度级为 L ，即 $f(i, j)$ 取值为 $[0, L-1]$ 。记灰度为 k 的像素总数为 n_k ，则灰度值 k 出现的概率 $p(k) = n_k / (MN)$ 。

假设以灰度值 T 为阈值将像素点分为目标像素与背景像素两类：灰度值为 $[0, \dots, T]$ 的像素点记为背景像素 D_0 而灰度值为 $[T+1, \dots, L-1]$ 的像素记为目标像素 D_1 ，于是有 D_0 和 D_1 出现的概率分别为 $P_0(T)$ ， $P_1(T)$ ；两类的灰度均值分别为 $\mu_0(T)$ 和 $\mu_1(T)$ ；方差分别为 $\sigma_0^2(T)$ 与 $\sigma_1^2(T)$ 。计算方法如下：

$$P_0(T) = \sum_{k=0}^T p_k \quad (5.9)$$

$$P_1(T) = \sum_{k=T+1}^{L-1} p_k = 1 - P_0(T) \quad (5.10)$$

$$\mu_0(T) = \sum_{k=0}^T k \frac{p_k}{P_0(T)} = \frac{1}{P_0(T)} \sum_{k=0}^T k p_k \quad (5.11)$$

$$\mu_1(T) = \sum_{k=T+1}^{L-1} k \frac{p_k}{P_1(T)} = \frac{1}{P_1(T)} \sum_{k=T+1}^{L-1} k p_k \quad (5.12)$$

$$\sigma_0^2(T) = \sum_{k=0}^T (k - \mu_0(T))^2 \frac{p_k}{P_0(T)} \quad (5.13)$$

$$\sigma_1^2(T) = \sum_{k=T+1}^{L-1} (k - \mu_1(T))^2 \frac{p_k}{P_1(T)} \quad (5.14)$$

整幅图像的灰度均值为 μ ，其值可由式(5.15)计算。

$$\mu = \sum_{k=0}^T k p_k = P_0(T) \mu_0(T) + P_1(T) \mu_1(T) \quad (5.15)$$

两类的类间距离平方为 $\sigma_b^2(T)$ ：

$$\sigma_b^2(T) = P_0(T)(\mu_0(T) - \mu)^2 + P_1(T)(\mu_1(T) - \mu)^2 \quad (5.16)$$

两类的类内方差和为 $\sigma_w^2(T)$ ，计算公式如式(5.17)所示。

$$\sigma_w^2(T) = P_0(T) \sigma_0^2(T) + P_1(T) \sigma_1^2(T) \quad (5.17)$$

若最佳阈值为 T^* ，则有：

$$T^* = \arg \max_{0 \leq T < L-1} (\sigma_b^2(T)) = \arg \min_{0 \leq T < L-1} (\sigma_w^2(T)) \quad (5.18)$$

图像二值化之后得到的只是间盘的上下两个边缘，若想分割出间盘还需对二值化图像进一步处理。由于间盘为两边缘之间的区域，所以可将上下边缘中间的像素点进行填充进而获得完整的间盘区域。间盘填充可通过将每一列中第一个非零像素点与最后一个非零像素点之间的像素点置 1 来实现。最后利用 canny 算子即可得到间盘轮廓的边缘。

5.3.3 间盘分割结果

为了验证算法的有效性与可行性，在得到间盘中心位置的基础上我们继续进行分割实验，实验结果如图 5.7 所示，其中图(a)为原腰椎 CT 图像，图(b)为检测后的 Hough 图

像，图(c)为经各向异性扩散滤波的 0° 方向信息测度图像，图(d)为二值化图像，(e)分割结果图像。从分割结果图像可以看出，第一组与第二组图像的分割结果都很理想。对于第三组中的间盘 L5-S1 没有分割出来主要是由定位检测时的漏检造成的。对于第四组图像中的第一节间盘（T9-10，属于胸椎间盘）的分割效果不理想主要是由原 CT 图像在该间盘上方有强边缘干扰存在导致的。

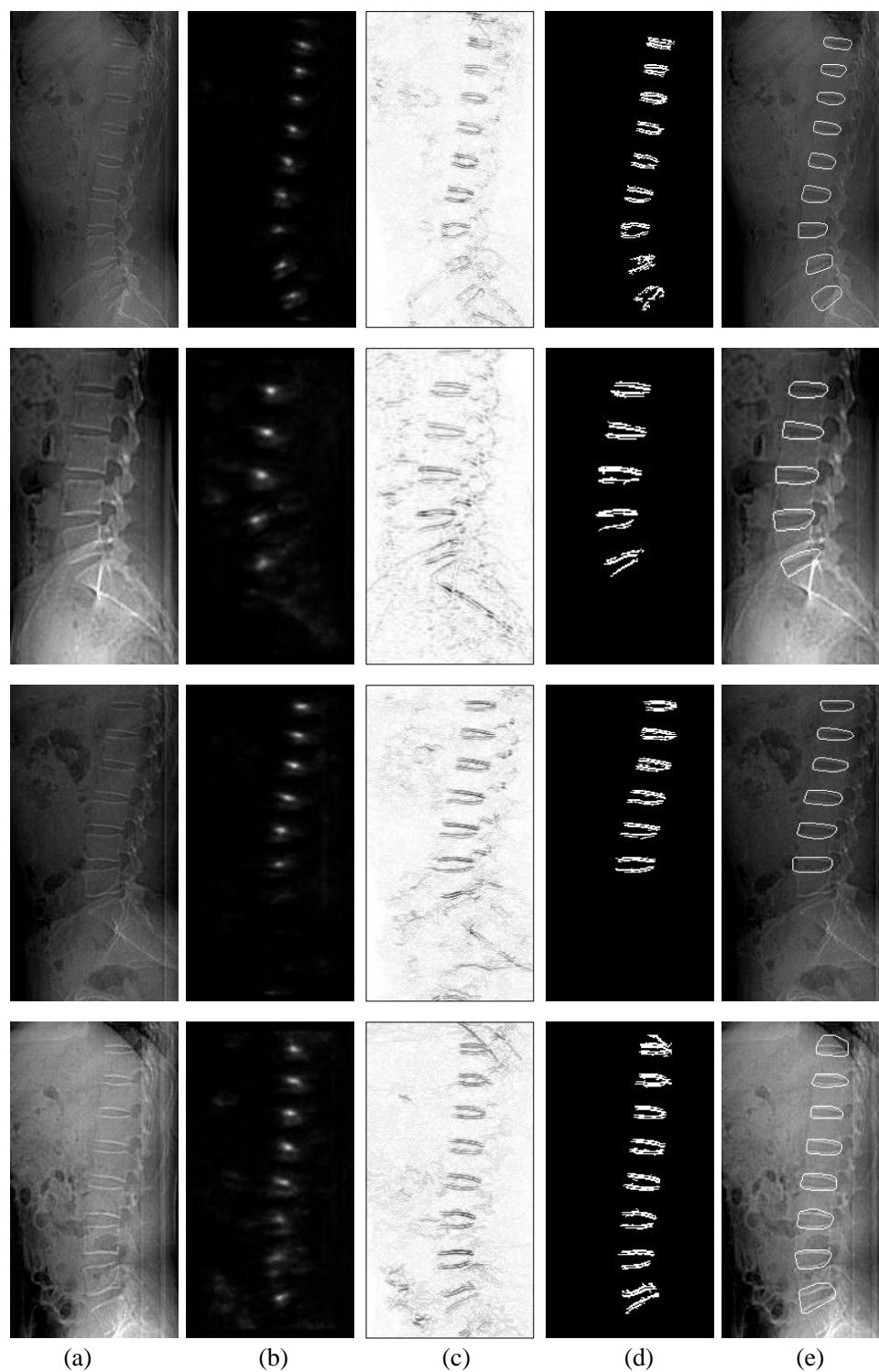


图 5.7 腰椎间盘分割结果

Fig. 5.7 The segmentation results of lumbar disc

5.4 本章小结

本章主要解决了间盘区域粗提取后的定位分割问题，椎间盘的准确分割在辅助医生临床诊断方面有着重要的参考意义。在上一章通过显著性计算得到了间盘区域的大概位置之后，提取间盘区域进行随机森林的训练以及间盘中心的检测。随后以检测得到的间盘中心位置为中心点提取其邻域图像作为间盘区域图像，对该部分图像进行基于 Otsu 的二值分割。因为有了间盘定位点的基础，分割结果受其他无关区域的影响很小，所以分割效果较为理想。但因为分割是在定位的基础上做的，所以分割结果与定位结果有直接关系。虽然在分割之前已经对定位结果进行检验，即对两个间盘中心位置间距较大的情况会取二者的中间位置作为漏检间盘的中心点，但是对于腰椎图像中第一个或者最后一个间盘漏检的情况则无法判断。所以若定位结果中出现漏检情况，则该漏检情况很可能也会出现在分割结果中。

第6章 总结与展望

6.1 论文总结

视觉选择注意机制使人类具有异常突出的数据筛选能力，面对外界时刻变化的信息，人们总能迅速觉察到那部分与自己相关的信息，并及时作出反应。随着信息技术的发展，我们所要处理的图像信息越来越多，如何迅速有效地处理日益庞大的信息库已经成为信息处理领域的一个难题。如果能将视觉选择注意机制引入到计算机信息处理中，就能够有效地解决这个难题。一般来说，图像处理通常只关注整幅图像中的部分图像，如果能直接提取出图像中感兴趣的那部分对其进行处理，不但可以降低处理过程的复杂度，还能提高图像处理的工作效率。

本文对视觉显著性计算模型及其应用进行了深入的研究，提出了新的思路，并组织了大量的实验对算法的可行性进行验证。文中提出了一种基于小波特征的概率评价显著性计算模型并将其应用到腰椎间盘的分割中。本文的主要工作内容如下：

(1) 首先介绍了当前主流的视觉注意的心理学模型与计算模型，并深入分析了各种计算模型的优劣。基于空域的计算模型虽然可以检测局部显著性，但算法复杂，计算耗时，且关于特征的选取也总是不完全。基于频域的计算模型计算速度较快，但是检测的是图像的全局显著性。基于此，本文提出了一种基于小波特征的概率评价显著性计算模型(WFPE)。在稀有特征更容易吸引注意的前提下，结合稀有特征在特征空间中出现的概率较低，我们将显著图定义为特征的概率密度图。根据熵的基本定义，若该区域的概率较低则其信息量会较高。所以，可将显著图的计算转为图像信息量的计算。在该模型中，采用更合理的 $YCbCr$ 色彩空间替代常用的 RGB 色彩空间将图像分为多通道处理。在 $YCbCr$ 色彩空间中，通过对图像进行小波多尺度分解分别得到三个颜色通道的特征向量，而特征向量的联合概率密度由独立元分析来计算。通过每个像素点在特征空间中的联合概率密度获得该点的自信息，进而得到整幅图像的显著图。

(2) 在验证了本文提出的显著性计算模型的有效性与可行性之后，将其应用到腰椎图像的椎间盘检测中。由于椎间盘在整幅图像中具有较明显的方向性，通过计算腰椎图像的显著性可以自动提取出显著度较高的间盘区域图像并将其作为下一步的待分割间盘区域图像。这样不仅缩小了检测范围，提高了处理效率，与此同时也大大降低了间盘误检情况的发生。接下来利用 Hough 森林检测出待分割区域图像中腰椎间盘的中心位置，结合先验知识再一次提取更精确的间盘区域图像来进行分割。

6.2 研究展望

本文对视觉选择注意机制进行了研究，提出了一种新的计算模型并将其与 hough 森林结合进行腰椎间盘突出分割。虽然这些工作取得了一定的成绩，但由于时间的限制加上对视觉系统中注意选择机制研究的不足，未来还有许多值得进一步探索的问题。

(1) 对于本文提出的视觉显著性计算模型，除了利用独立元分析来计算高维特征空间中点的概率密度之外是不是可以利用其它办法对高维空间进行降维。由于人类的视觉系统在注视一个物体的同时也会看到物体周围的区域，这里并不会明确的边界存在。那么如何在图像中表示这种并不明确的边界是今后研究的难点。另外，本文的显著性计算模型只针对静态图片设计，而没有加入时间信息使其可以计算动态视频的显著性，这使其缺席了关于视频处理的所有应用。

(2) 对于腰椎间盘突出检测部分，也有一些需要改进的地方。由于分割时在检测定位到间盘的中心点的基础上做的，这就意味着如果检测定位失败则分割也就不能实现，而且由于最后的分割时用二值化的方法实现的，如果间盘区域周围的图像质量较差则会影响分割效果。文中的算法完成了间盘的自动分割，接下来是否能将分割出来的间盘进行识别分类自动找出异常的间盘，这就需要继续提高分割算法的精度，这也会是今后工作的重点与难点。

参考文献

1. 宋雁斓. 视觉注意模型及其在图像分类中的应用[D], 上海: 上海交通大学, 2008.
2. Itti L. Models of Bottom-up and Top-down Visual Attention [D], Pasadena: California Institute of Technology, 2000.
3. Itti L. Visual Attention and Target Detection in Cluttered Natural Scenes [J], Optical Engineering, 2001, 40(9): 1784-1793.
4. Frintrop S. A Visual Attention System for Object Detection and Goal-directed Search [D], Germany: University of Bonn, 2005.
5. Ackerman C, Itti L. Robot Steering With Spectral Image Information [J], IEEE Transactions on Robotics, 2005, 4(21): 247-251.
6. Heidmann G, Rae R, Bekel H. Integrating Context-free and Context-dependant Attentional Mechanisms for Gestural Object Reference [C], Proceedings of the 3rd international conference on Computer vision systems, Lyon, France, 2005: 22-33.
7. 田捷, 包尚联. 医学影像处理与分析[M], 北京: 电子工业出版社, 2003: 30-44.
8. 赵燕燕. MRI脊柱图像椎间盘分割及定位算法研究[D], 北京: 北京交通大学, 2008.
9. Cave K R, Wolfe J M. Modeling the Role of Parallel Processing in Visual Search [J], Cognitive Psychology, 1990, 22: 225-271.
10. Marr D, Nishihara H K. Representation and Recognition of the Spatial Organization of Three Dimensional Structure [J], Proceedings of the Royal Society of London B, 1978, 200: 269-294.
11. Treisman A, Gelade G. A Feature-integration Theory of Attention [J], Cognitive Psychology, 1980, 12(1): 97-136.
12. Wolfe J. Guided Search 2.0: A Revised Model of Guided Search [J], Psychonomic Bulletin & Review, 1994, 1(2): 202-238.
13. Itti L, Koch C, Niebur E. A Model of Saliency-based Visual Attention for Rapid Scene Analysis [J], IEEE Transaction on Pattern Analysis and Machine Intelligence, 1998, 20(11): 1254-1259.
14. Itti L, Koch C. A Saliency-based Search Mechanism for Overt and Covert Shifts of Visual Attention [J], Vision Research, 2000, 40: 1489-1506.

15. Itti L, Baldi P. Bayesian Surprise Attracts Human Attention [C], The Neural Information Processing Systems, Los Angeles, USA, 2005: 547-554.
16. Walther D, Koch C. Modeling Attention to Salient Proto-objects [J], Neural Networks, 2006, 19: 1395-1407.
17. Bruce D B. Features that Draw Visual Attention: An Information Theoretic Perspective [J], Neurocomputing, 2005, 65-66: 125-133.
18. Hou X, Zhang L. Saliency Detection: A Spectral Residual Approach [C], IEEE Conference on Computer Vision and Pattern Recognition, Minnesota, USA, 2007: 1-8.
19. Guo C L, Ma Q, Zhang L M. Spatio-temporal Saliency Detection Using Phase Spectrum of Quaternion Fourier Transform [C], IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, USA, 2008: 417-423.
20. Chen D Y, Han P and Wu C D. Frequency Spectrum Modification: A New Model for Visual Saliency Detection [C], International Symposium on Neural Networks, Shanghai, China, 2010: 90-96.
21. 史睿琼. 脊柱MRI图像椎间盘定位及分割算法的研究[D], 北京: 北京交通大学, 2007.
22. Smyth P C, Taylor J A. Automatic Measurement of Vertebral Shape Using Active Shape Models [C], The British Machine Vision Conference, Sarasota, FL, USA, 1996: 705-714.
23. Cootes T F, Cooper D, Taylor C J, Graham J. Active Shape Models Their Training and Application [J], Computer Vision and Image Understanding, 1995, 61:38-59.
24. Cootes T F, Taylor C J. Statistical Models of Appearance for Medical Image Analysis and Computer Vision [C], Proc. SPIE Medical Imaging, San Diego, CA, USA , 2001.
25. Koch C, Ullman S. Shift in Selective Visual Attention: Towards the Underlying Neural Circuitry [J], Human Neurobiology, 1985, 4(4): 219-227.
26. 李强. 医学图像分割进展[J], 临床影像技术, 2010, 25(5): 121-124.
27. 单列. 视觉注意机制的若干关键技术及应用研究[D], 安徽: 中国科学技术大学, 2008.
28. 吴逊. 基于独立成分分析的特征提取方法研究[D], 厦门: 厦门大学, 2007.
29. Wolfe J M, Gancarz G. Guided Search 3.0: Basic and Clinical Applications of Vision Science [M], Dordrecht, Netherlands: Kluwer Academic, 1996, 189-192.
30. Wolfe J M. Guided Search 4.0: Current Progress with a Model of Visual Search [M]. New York: Oxford, In W. Gray (Ed.), Integrated Models of Cognitive Systems, 99-119.
31. Itti L, Koch C. Computational Modeling of Visual Attention [J], Nature Neurosci. Rev.,

- 2001, 21: 314-329.
32. Bruce D B, Tsotsos J K. Saliency Based on Information Maximization [J], Advances in Neural Information Processing Systems, 2006, 18(4): 155-162.
33. 过晨雷. 注意力选择机制的研究: 算法设计以及系统实现[D], 上海: 复旦大学, 2008.
34. Engel S, Zhang X and Wandell B. Colour Tuning in Human Visual Cortex Measured with Functional Magnetic Resonance Imaging [J], Nature, 1997, 388(6): 68-71.
35. 冈萨雷斯. 数字图像处理[M], 北京: 电子工业出版社, 2005: 175-182.
36. 寿天德. 神经生物学[M], 北京: 高等教育出版社, 2005: 192-210.
37. Wikipedia, <http://en.wikipedia.org/wiki/YCbCr>.
38. Lindeberg T. Scale Space Theory in Computer Vision [M], Kluwer academic publisher, 1993, 76-81.
39. Lindeberg T. Scale Space for Discrete Signals [J], IEEE Transactions Pattern analysis and machine intelligence, 1990, 12(3): 234-245.
40. Wikipedia, http://en.wikipedia.org/wiki/Independent_component_analysis.
41. Hyvärinen A, Oja E. Independent Component Analysis: Algorithms and Applications [J], Neural Networks, 2000, 13(45): 411-430.
42. Hyvärinen A. Survey on Independent Component Analysis [J], Neural Computation, 1999, 2: 94-128.
43. Hoyer P. Independent Component Analysis in Image Denoising [D], Helsinki University of Technology, 1999.
44. Himberg J, Hyvärinen A. Independent Component Analysis for Binary Data: An experimental study [R], In Proc. Int. Workshop on Independent Component Analysis and Blind Signal Separation (ICA2001), San Diego, California, 2001.
45. Rafael C G, Richard E W. Digital Image Processing [M], Publishing House of Electronics Industry, 2006, 70-78.
46. Michael I, Robert D, Lisa S C, Jonathan V. Inhibition of Return: Neural Basis and Function [J], Cognitive Neuropsychology, 1985, 2(3): 211-228.
47. Gao D, Mahadevan V. The Discriminant Centersurround Hypothesis for Bottom-up Saliency [C], The Neural Information Processing Systems, Whistler, BC, Canada, 2007.
48. Harel J, Koeh C, Perona P. Graph-based Visual Saliency [C], The Neural Information Processing Systems, Vancouver, BC, Canada, 2006.

49. <http://zh.wikipedia.org/wiki/ROC%E6%9B%B2%E7%BA%BF>.
50. Micheline K, Han J. Data Mining: Concepts and Techniques [M], Elsevier, 2010, 243-244.
51. Pianykh O S. Digital Imaging and Communications in Medicine (DICOM) [M], National Electrical Manufactures Association Press, 2000, 52-60.
52. 谢强, 张朝宗. 计算机断层成像技术: 原理、设计、伪像和进展[M], 北京: 科学出版社, 2006, 120-132.
53. Tang Z W, Zhang H, Hu G S. Multi-scale Display Processing of Radiographic Images Acquired with a Prototype Flat Panel Detector [C], IEEE Engineering in Medicine and Biology Society, Shanghai, 2006: 6595-6597.
54. 杨海军, 梁得群, 毕胜. 基于图像方向性信息测度的图像像素分类方法[J], 中国图象图形学报, 2001, 31(5): 429-433.
55. Perona P, Malik J. Scale Space and Edge Detection using Anisotropic Diffusion [J], IEEE Transactions on Pattern Analysis and Machine Intelligence, 1990, 12(7): 629-639.
56. Gall J, Lempitsky V. Class-specific Hough Forests for Object Detection [C], Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 2009: 1022-1029.
57. 刘艳丽. 随机森林综述[D], 天津: 南开大学, 2008.
58. 方匡南, 吴见彬, 朱建平, 谢邦昌. 随机森林方法研究综述[J], 统计与信息论坛, 2011, 26 (3): 32-37.
59. Ballard D H. Generalizing the Hough Transform to Detect Arbitrary Shapes [J], Pattern Recognition, 1981, 13(2): 111-122.
60. 李子龙, 刘伟铭. 基于改进Hough森林的对象检测方法[J], 计算机工程与应用, 2011, 47(26): 4-6.
61. Breiman L. Random forests [J], Machine Learning, 2001, 45: 5-32.
62. 王丽婷, 丁晓青, 方驰. 基于随机森林的人脸关键点精确定位方法[J], 清华大学学报, 2009, 49(4): 543-546.
63. Otsu N. A Threshold Selection Method from Gray-level Histograms [J], IEEE Transactions on System Man and Cybernetic, 1979, 9(1): 62-66.
64. 余锦华, 汪源源. 基于各向异性扩散的图像降噪算法综述[J], 电子仪器学报, 2011, 25(2): 105-116.

致谢

两年的研究生学习生涯转瞬即逝，值此论文即将完成之际，谨向我的导师、同学和父母致以最诚挚的感谢。

在硕士期间的研究工作中取得的成果，首先应该归功于恩师陈东岳副教授的悉心指导与耐心帮助。陈老师一丝不苟的治学态度、对研究领域新思想新观念的执著追求以及精益求精的精神都给我留下了深刻的印象。陈老师不仅在学业上给我精心的指导，使我的理论知识和编程能力都有了很大的提升，而且在生活和思想上给我悉心的关怀，为我创造了良好的学习和实践环境，使我在理论学习的同时也积累了很多的实践经验。在这两年的研究生学习过程中，我所取得的每一点进步都凝结着陈老师的心血和汗水，在此表示衷心的感谢和崇高的敬意。

感谢人机所的吴成东教授、薛定宇教授、赵姝颖教授、郝培锋教授、魏颖副教授、王晓哲副教授、闻时光老师和潘峰老师，你们渊博的学识给我留下了深刻印象，精彩的课程讲解使我开阔了视野，增长了知识。

感谢于晓升、李灵帅、卢素芬、白静等师兄师姐的谆谆教导和一起度过的美好时光；感谢实验室里共同奋战的宋超、冯琦、周婷、赵明轩、桑永嘉、陈宗文、田子恒，我们共同度过的这两年实验室生活将会是我永远难忘的回忆。

感谢一直爱护、支持我的家人好友，是你们一直在背后默默的支持和帮助我，这些都是我前进的动力。

最后，对所有关心、帮助过我的人表示深深的谢意。

硕士期间发表的论文

Dongyue Chen, Dan Zhao, Xiaosheng Yu, and Zongwen Chen. Saliency Detection Model Based on Wavelet Transform and Independent Component Analysis [C]. Third International Conference on Intelligent Control and Information Processing, 2012.(已录用待发表)

作者简介

赵丹，女，汉族，1987年4月出生，吉林省辽源市人。2006年至2010年就读于西安邮电学院测控技术与仪器专业，获得西安邮电学院工学学士学位。2010年考入东北大学模式识别与智能系统专业，师从陈东岳副教授，攻读硕士学位。

硕士研究方向为图像处理、计算机视觉。对基于生物视觉系统信息处理加工机制的注意力选择进行了深入的研究。

