

**Visual Saliency
Detection Methods and
Applications**

**视觉显著性
检测方法及应用**

钱晓亮 王 慰 王延峰 曾 黎 著



中国工信出版集团



电子工业出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
<http://www.phei.com.cn>

视觉显著性 检测方法及应用

钱晓亮 王 慰 王延峰 曾 黎 著



電子工業出版社
Publishing House of Electronics Industry
北京 · BEIJING

内 容 简 介

视觉显著性检测是计算机视觉领域近年来的一个研究热点,具有广泛的应用前景。本书介绍了视觉显著性检测的基本知识和现有方法,并系统总结了作者近几年在视觉显著性检测方法、视觉显著性用于红外目标预检测和太阳能电池片表面缺陷检测等方面的研究成果。全书共7章,分为4部分:第1部分(第1章)详细介绍了视觉显著性检测的定义、分类、应用等相关基础知识;第2部分(第2章)对视觉显著性检测领域的研究现状进行了分类介绍,并分析了现有工作存在的问题;第3部分(第3~5章)详细介绍了本书提出的3种视觉显著性检测方法;第4部分(第6、7章)详细介绍了视觉显著性在红外目标预检测和太阳能电池片表面缺陷检测中的应用。另外,第3~7章均包含相关的实验案例,以便有兴趣的读者进一步钻研探索。

本书可为高等院校电气工程、控制科学与工程、计算机科学、信息科学、人工智能等领域的研究人员和工程技术人员提供参考,也可作为相关专业高年级本科生和研究生的教学参考书。

未经许可,不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有,侵权必究。

图书在版编目(CIP)数据

视觉显著性检测方法及应用/钱晓亮等著. —北京:电子工业出版社,2018.8

ISBN 978-7-121-34803-7

I. ①视… II. ①钱… III. ①计算机视觉—检测—研究 IV. ①TP391.41

中国版本图书馆CIP数据核字(2018)第171123号

策划编辑:米俊萍

责任编辑:米俊萍

特约编辑:丁福志

印刷:北京虎彩文化传播有限公司

装订:北京虎彩文化传播有限公司

出版发行:电子工业出版社

北京市海淀区万寿路173信箱 邮编:100036

开本:720×1000 1/16 印张:14.25 字数:256千字 彩插:10

版次:2018年8月第1版

印次:2018年8月第1次印刷

定价:68.00元

凡所购买电子工业出版社图书有缺损问题,请向购买书店调换。若书店售缺,请与本社发行部联系,联系及邮购电话:(010)88254888,88258888。

质量投诉请发邮件至zlts@phei.com.cn,盗版侵权举报请发邮件至dbqq@phei.com.cn。

本书咨询联系方式:mijp@phei.com.cn。

前言

FOREWORD

随着信息技术的发展,图像数据的规模变得越来越大,面对如此庞大的图像数据,如何能够快速准确地完成各种图像分析任务已经成为当前研究的一个重要方向。视觉显著性检测是计算机视觉领域中模拟灵长类动物的视觉注意机制而出现的一个课题,它可以引导机器视觉系统将有限的硬件资源优先分配给少数几个显著区域,为后续的视觉处理(如图像理解、场景分析、目标检测和识别等)提供极大的便利。

自 1998 年 Laurent Itti 提出第一种视觉显著性检测方法以来,视觉显著性检测技术得到了越来越多的关注。越来越多的科研人员投入到视觉显著性检测的研究中,涌现出了大量的视觉显著性检测方法、开源代码和 Benchmarks,各种基于视觉显著性的应用也层出不穷。目前,有关视觉显著性检测的论文、评论等相关资料较为丰富,但是相关的书籍不多,有限的几本著作也侧重于视觉显著性的应用,尚缺乏专门针对视觉显著性检测方法、使读者了解视觉显著性检测的来龙去脉、为以后进一步深入研究视觉显著性检测方法或者使用视觉显著性检测做相关应用奠定基础的书籍。

本书以“保证基础、突出能力培养”为根本出发点,对章节安排和内容描述方式做了精心设计。全书可分为 4 部分。第 1 部分详细介绍了视觉显著性检测的定义、分类、应用等相关基础知识,可使基础较弱的读者快速了解相关知识;第 2 部分对视觉显著性检测领域现有的研究工作按照设计思路的不同进行了分类介绍,并分析了现有工作存在的问题,使读者对本领域的研究现状有了大致的了解;第 3 部分按照“创新动机—设计思路—方法展开—实验验证”学术研究路线,详细介绍了本书提出的 3 种视觉显著性检测方法,使读者在学习视觉显著性检测方法的同时,了解学术研究的大致流程,对相关专业的高年级本科生和研究生培养起到促进作用;第 4 部分介绍了视觉显

著性在红外目标预检测和太阳能电池片表面缺陷检测中的应用,使读者了解视觉显著性在实际任务中的应用方法。

我们依托郑州轻工业学院河南省信息化电气重点实验室,成立了人工智能与模式识别研究室,致力于人工智能在计算机视觉和模式识别领域的研究与应用,本书即是我们研究工作的初步总结。

本书由郑州轻工业学院电气信息工程学院的钱晓亮负责第1~5章的撰写,曾黎负责第6章和第7章的撰写,国家“千人计划”专家王慰教授负责指导算法的设计和实验验证方案,王延峰教授负责全书的章节安排并审阅了书稿的全部内容。

本书的完成离不开郑州轻工业学院电气信息工程学院人工智能与智能系统团队的多位老师和研究生的支持和帮助,感谢团队中杨存祥教授、吴青娥教授、姜利英教授、过金超副教授、毋媛媛博士、陈虎博士、刘玉翠博士、赵素娜博士、王芳博士、张焕龙博士、贺振东博士、刁智华博士、杨飞飞博士、张吉涛博士和李金城硕士对本书的关心、支持与辛勤付出,感谢团队中张鹤庆、李佳、白臻、林生、栗靖、李二凯、成曦等研究生在写作过程中的无私付出与辛勤努力。本书的工作也得到郑州轻工业学院电气信息工程学院领导,以及国家自然科学基金(61501407、61632002、61472372、61603350)、国家“973”计划(613237)、河南省重大科技专项(161100211600)、河南省科技创新杰出人才项目(174200510012、184200510015)、河南省科技创新杰出青年项目(164100510017)、河南省高校科技创新团队(19IRTSTHN013)、郑州轻工业学院博士基金项目(2014BSJJ016、2015BSJJ025、2015BSJJ027、2016BSJJ006)等科研项目的支持,特此感谢。同时,特别感谢电子工业出版社的大力支持和帮助,感谢董亚峰老师和米俊萍老师付出的辛勤劳动与努力。感谢书中所有被引用文献的作者。

视觉显著性检测方法及应用一直发展较快,本书的取材和安排完全是作者的偏好,由于水平有限,书中内容可能存在遗漏和不妥之处,恳请广大读者批评指正。

作 者

2018年2月

郑州轻工业学院

目 录

CONTENTS

第 1 章 绪论	1
1.1 引言	1
1.2 视觉显著性检测的研究现状	3
1.3 视觉显著性检测的应用现状	4
1.4 本书的主要内容和章节安排	7
1.4.1 主要内容	7
1.4.2 章节安排	10
第 2 章 视觉显著性检测方法综述	13
2.1 引言	13
2.2 预备知识	13
2.2.1 视觉注意与视觉显著性	14
2.2.2 视觉显著性模型的分类	15
2.3 方法评价	18
2.3.1 基准测试库	18
2.3.2 定量对比方法	21
2.4 流行检测方法介绍	24
2.4.1 特征组合理论	24
2.4.2 信息论	28
2.4.3 图上随机游动理论	29

2.4.4 决策理论	30
2.4.5 贝叶斯理论	31
2.4.6 频域分析	32
2.4.7 机器学习	34
2.4.8 其他	35
2.5 现有方法存在的问题及解决方案	36
2.5.1 现有方法存在的问题	36
2.5.2 解决方案	38
2.6 发展趋势	40
2.7 本章小结	41
第3章 一种基于加权稀疏编码的频域方法	42
3.1 引言	42
3.2 过完备稀疏编码	44
3.2.1 过完备稀疏编码的神经生理学基础	45
3.2.2 图像的过完备稀疏编码	46
3.2.3 过完备词典的构造	49
3.3 加权稀疏编码	51
3.3.1 增量编码长度算法	51
3.3.2 稀疏编码的权重	52
3.4 图像标记算法	54
3.4.1 基于图像标记算法的显著性提取	54
3.4.2 相关理论证明	56
3.5 基于加权稀疏编码的图像标记算法	59
3.5.1 单通道的图像标记算法	60
3.5.2 多通道的图像标记算法	60
3.6 实验对比	61
3.6.1 主观对比	62
3.6.2 定量对比	64
3.6.3 算法复杂度评估	65

3.7 本章小结	66
第 4 章 基于最优对比度的视觉显著性检测方法	68
4.1 引言	68
4.2 总体思想	69
4.2.1 最优对比度引入的动机	70
4.2.2 实现方案	75
4.3 候选中心外围对比度	75
4.3.1 计算原理	75
4.3.2 实现细则	80
4.4 最优对比度	82
4.4.1 单尺度下的最优中心外围对比度筛选	82
4.4.2 多尺度增强算法	84
4.5 实验对比	86
4.5.1 主观对比	87
4.5.2 定量对比	89
4.6 本章小结	93
第 5 章 融合长期特征和短期特征的贝叶斯模型	94
5.1 引言	94
5.2 总体思想	95
5.2.1 先验知识的作用和使用方式	95
5.2.2 当前观测信息的作用和使用方式	99
5.2.3 先验知识和当前观测信息的融合	101
5.2.4 实现方案	102
5.3 长期特征和短期特征	103
5.3.1 长期词典和短期词典	103
5.3.2 特征提取	105
5.4 基于贝叶斯模型的视觉显著性检测	105
5.4.1 贝叶斯模型	105

5.4.2 特征概率分布的估计	108
5.4.3 视觉显著性随空间位置变化的条件概率分布估计	111
5.5 实验对比	112
5.5.1 主观对比	113
5.5.2 定量对比	115
5.6 本章小结	116
第 6 章 基于视觉显著性的红外目标预检测	117
6.1 引言	117
6.1.1 研究背景与意义	117
6.1.2 相关工作	118
6.1.3 总体解决方案	121
6.2 基于小数目标尺度的红外图像混合滤波算法	124
6.2.1 红外图像噪声分析	124
6.2.2 降噪算法的总体设计思路	125
6.2.3 小数目标尺度	126
6.2.4 基于小数目标尺度的自适应高斯滤波器	130
6.2.5 基于小数目标尺度的自适应中值滤波器	130
6.3 视觉显著性检测方法的选择	132
6.3.1 综合定量对比的实验准备	133
6.3.2 综合定量对比	134
6.3.3 原理分析	138
6.4 红外目标预检测	139
6.4.1 基于视觉显著性检测的窗口特征算子	139
6.4.2 基于 SLIC 超像素分割的窗口特征算子	140
6.4.3 窗口特征算子参数的估计	142
6.4.4 窗口特征的贝叶斯融合	143
6.4.5 目标窗口的确定	144
6.5 实验对比	145
6.5.1 红外图像降噪实验对比	145
6.5.2 红外目标预检测实验对比	147
6.6 本章小结	150

第 7 章 基于视觉显著性的太阳能电池片表面缺陷检测	151
7.1 引言	151
7.1.1 研究的背景与意义	151
7.1.2 相关工作	152
7.1.3 存在问题及解决方案	158
7.2 太阳能电池片表面图像预处理	159
7.2.1 图像采集	159
7.2.2 图像降噪	162
7.2.3 栅线删除	163
7.2.4 栅线填充	164
7.3 基于视觉显著性的缺陷初始检测	165
7.3.1 自学习特征提取	165
7.3.2 低秩矩阵复原	167
7.3.3 获取视觉显著图	169
7.4 基于视觉显著性和图像分割的缺陷精确定位	171
7.4.1 图像分割	171
7.4.2 基于视觉显著性的缺陷定位	174
7.5 基于形态学的检测结果优化	175
7.5.1 形态学理论	175
7.5.2 形态学优化	176
7.6 实验设计	178
7.6.1 主观对比	180
7.6.2 客观对比	184
7.7 软件设计与使用	187
7.7.1 检测软件的设计	187
7.7.2 软件功能及操作介绍	189
7.8 本章小结	191
参考文献	192

第1章

绪论

1.1 引言

在我们每天所接收的大量信息中，大约有 80%是来自视觉的，然而，人脑的神经资源是有限的，在人的视觉通路中存在一个信息瓶颈^[1]，因此，尽管在我们周围充满了各种各样的物体，但在任何时刻仅仅只有少量的刺激能够进入视觉系统得到进一步的加工处理，进而影响我们的行为反应。在这个过程中，视觉注意机制扮演着重要角色。

视觉注意 (Visual Attention) 机制是人类视觉系统 (Human Visual System, HVS) 内在的一个重要机制^[2]。HVS 在面对一个复杂场景时，总会迅速地选择少数几个与当前行为或视觉任务有关的感兴趣的区域进行优先处理，该过程称为视觉注意，被选中的区域称为注意焦点 (Focus of Attention, FoA) 或注意的单元^[3,4]。HVS 能够大大地减少输入的视觉数据量，以不同的次序和力度对各个场景区域进行选择性地加工，从而避免了计算浪费，降低了分析难度。图 1-1 所示的计算机游戏“找茬”可以说明视觉注意机制的存在。游戏在计算机屏幕上并排放了两张非常相似的图像，但其中有几个地方有差别，游戏任务是找出所有这些存在差别的地方。虽然我们的视野覆盖了整个计算机屏幕，但是我们通常需要花费较长的时间才能找出所有存在差别的地方。造成这个结果的原因就是 HVS 每次只能对 FoA 进行仔细观察和分析，却不能感知其他的细节信息。此外，“变化无感知”^[5,6]或“忽略无感知”^[7]等实验现象也是视觉注意机制的反映。



图 1-1 计算机游戏“找茬”

视觉注意机制与感知、学习、记忆等其他认知活动紧密相关，对于它的研究涵盖了认知心理学、认知神经科学、认知科学和计算机科学等多门学科。在计算机视觉领域，视觉注意机制的研究主要集中在图像（视频）的视觉显著性检测（Saliency Detection）。所谓视觉显著性检测，通俗地说，就是计算图像中各个部分吸引人们视觉注意的程度，该程度称为显著性（Saliency），计算的结果称为显著图（Saliency Map, SM），以显著图作为视觉显著性检测的输出。

在计算机视觉领域，视觉显著性检测的研究具有很高的应用价值。随着信息技术的发展，图像数据的规模越来越大，面对海量的图像数据，视觉注意机制中的显著性计算可以为选择少数几个视觉显著区域和排除冗余提供快速的计算机制^[8]，帮助计算机快速而准确地完成各种图像分析任务。例如，在机器视觉领域，机器人导航及目标检测、跟踪和识别非常重要，引入视觉显著性检测方法可以使机器人更快、更有效地进行视觉搜索。在图像压缩领域，将人容易注意到的图像区域压缩得轻些，而将人不容易注意到的区域压缩得重些，这样在保持总压缩率不变的情况下，可以提高图像的视觉感知质量。在视频监控、自动驾驶、数字媒体库中的信息检索及医学图像分析领域，采用视觉显著性检测做预处理，可以在海量数据中高效地找到有用的信息。

本书致力于静态图像视觉显著性检测方法的研究，并拓展了视觉显著性检测在红外目标预检测及太阳能电池片表面缺陷检测方面的应用。本章剩余部分的安排如下：1.2 节和 1.3 节分别介绍视觉显著性检测的研究现状和应用现状，1.4 节介绍本书的主要内容和章节安排。

1.2 视觉显著性检测的研究现状

按照计算机视觉中对视觉信息的处理方式划分,现有的视觉显著性模型可分为两大类,分别为自下而上(Bottom-up)和自上而下(Top-down)的视觉显著性模型。自下而上型的视觉显著性模型由数据驱动,独立于具体任务;而自上而下型的视觉显著性模型受意识支配,依赖具体任务。从现实生活的角度来看,自上而下的视觉显著性模型是多数,因为人们往往都是一带而过地在视觉场景中进行搜索,然而,目前研究较多的是自下而上的视觉显著性模型,因为自上而下的视觉显著性模型依赖具体任务,不便于进行通用的研究。研究自下而上的视觉显著性模型时,默认观察者处于自由观测(Free-viewing)的状态,即观察者不带任何目的。本书研究的是自下而上的视觉显著性模型。

从研究对象来说,视觉显著性模型可分为基于空间的和基于空间-时间的视觉显著性模型,具体来说,就是可分为静态图像和动态视频的视觉显著性模型。目前,研究静态图像视觉显著性模型的居多。事实上,动态视频的视觉显著性模型可看作是单帧图像的视觉显著性模型加上运动信息。在静态图像上表现较好的视觉显著性模型,通常在动态视频上也表现较好^[9],也就是说,静态图像的视觉显著性建模是动态视频的视觉显著性建模的基础。本书研究的是静态图像的视觉显著性模型。

现有的视觉显著性检测方法根据采用数学模型的不同大致可分为7类,下面分别进行简要介绍。

(1) 基于特征组合理论。Treisman 和 Gelade 于 1980 年提出的特征组合理论 (Feature Integration Theory, FIT)^[10]在视觉注意研究领域内被广泛接受。FIT 认为,在视觉注意初期,视觉系统提取底层特征(如亮度、方向、颜色、大小等),并分别进行平行地加工,在这一过程中并不存在视觉注意机制,视网膜平行地处理各种特征。在此之后,各种特征将会被逐步整合,这个整合过程需要视觉注意机制的参与。存储在人脑中的常识也会影响特征整合。视觉注意逐步连接各种特征,最终可形成显著图。

(2) 基于信息论。基于信息论进行视觉显著性检测的基本思路是:根据图像中各个位置所包含的信息量的多少来定义显著性的大小,认为包含的信息越多越能吸引人的注意。信息量一般用香农熵计算,然而, Bruce 等人^[11]指出,相比香农熵,香农自信息 (Shannon Self-information) 更适合定义显

著性的大小。香农自信息计算的实质是统计底层视觉特征的稀少性，即对应特征出现概率较小的位置具有较大的显著性。

(3) 基于图上随机游走理论。以图论为基础，在图上建立一个马尔可夫链，以马尔可夫链上的节点为计算单位，根据随机游走者在节点上覆盖的时间或到达的频率等来衡量节点的显著性。

(4) 基于决策理论。基于决策理论的方法认为人类的视觉感受系统会对周围环境的状态做出最优的决策（所谓最优是指概率意义上的误差最小），人眼的视觉注意机制由这种决策驱动。该方法实现的大致思路：将图像中的内容分为显著区域和非显著区域两类，提取图像的底层视觉特征，计算已知特征时各位置属于显著区域的后验概率，以此作为该位置的显著性。

(5) 基于贝叶斯理论。贝叶斯理论的优势在于可以融合先验知识，在现有的基于贝叶斯理论的方法中，一般以场景内容（语义信息等）、要点（Gist）、目标特征等作为先验知识。基于贝叶斯理论的方法设计非常灵活，可以根据不同的假设或任务定义各种形式的先验知识。

(6) 基于频域分析。与其他方法在空域中进行显著性提取不同的是，基于频域分析的方法通过傅里叶变换或 DCT 变换等将图像转换到频域中，在频域中，通过对幅频谱或相位谱进行分析和计算，提取显著信息，再进行反变换得到显著图。

(7) 基于机器学习。基于机器学习的方法的核心思想是根据已知测试图像及其对应的真值训练分类器，如 SVM，通过该分类器将提取的底层视觉特征进行融合，以融合后的特征定义视觉显著性的大小。

1.3 视觉显著性检测的应用现状

在计算视觉领域中基于视觉显著性检测的应用已经越来越多，本书在这里选择几个主要的应用方向进行介绍。

1. 显著物体分割

这是目前视觉显著性检测应用较多的一个方向。其基本思路是对图像进行视觉显著性检测，得到显著图。显著物体本身具有显著性，能在显著性检测过程中检测出来，但是这种检测一般是不精确的（不能准确刻画物体的边界），不能通过简单设定阈值分割出完整精确的物体，因此，通常的做法是以

显著图为指导,结合传统的图像分割算法,如 K-Means^[12]、Meanshift^[13]、Grabcut^[14]等,进行显著物体的准确分割。图 1-2 所示为文献[14]给出的四个显著物体分割的示例。

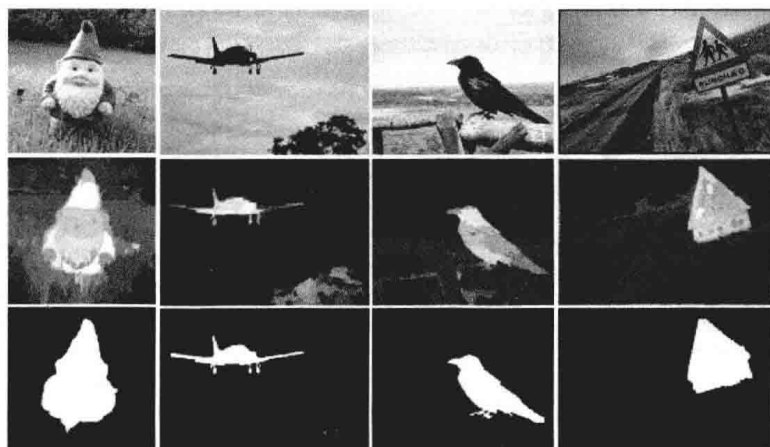


图 1-2 文献[14]中四个显著物体分割的示例

(自上而下的三行依次是原始图像、显著图、显著物体分割结果图)

其他引用次数较多的该方向的文献还有文献[15~17]等。

2. 图像压缩

原有的大多数图像压缩算法都是对图像中的所有区域采用相同的压缩比进行压缩,而实际上人们往往只对图像中的一部分内容感兴趣。如果在对图像进行压缩的时候能够对这些感兴趣的区域采用低的压缩比,对那些非感兴趣的区域采用高的压缩比,则既能够有效地压缩数据量,又能够保证重要的信息不丢失。图像压缩与编码是视觉注意机制当前非常活跃的一个应用领域。利用视觉注意机制,可迅速地找到图像中的关键区域。对关键区域进行无损或低损压缩,对其他区域进行有损压缩,这样既可以保证压缩后图像的质量,又可以获得较高的压缩比^[18~21]。目前,这种思想已被 JPEG2000 标准采用。

3. 目标重定位

为了在小的屏幕上有效地显示多媒体图像和视频,图像/视频基于内容的目标重定位(Retargeting)受到很大的关注^[22~28]。目标重定位的目的是在各种设备上显示图像,同时保持必要的视觉质量和保留重要的内容。为了适应小的显示器,那些不重要的像素或信息被忽略,而重要的信息则保留,同时,

保留图像中的结构信息和视觉信息。例如，一条直线在改变尺度后依然要求保持为一条直线。目标重定位的核心是计算或者确定图像的有效部分，即需要保留或显示的区域。于是，对图像进行视觉显著性检测，将显著图中显著性较大的区域作为图像的有效部分就是一个合理的选择。图 1-3 所示为一个目标重定位的示例^[28]。



图 1-3 一个目标重定位的示例

4. 图像拼接

随着图像的日益增多，人们开始关注如何有效地浏览图像，尤其是面对上千万幅图像或者搜索后返回上千万幅图像时。基于此，很多学者提出图像拼接（Collage）^[29~33]。图像拼接就是将许多图像放在一个显示器上显示。图像拼接的关键技术是确定图像中重要的感兴趣区域（Region Of Interest, ROI），也就是显示给用户的部分。而提取 ROI 的基本依据是图像的显著图。当然，要完成图像拼接还需要综合考虑各种约束因素，如显著图约束（将显著图区域显示出来）、画布约束（将所有的图显示在画布上）等，然后将这些约束转化为优化问题。图 1-4 所示为一个图像拼接的示例^[33]。



图 1-4 一个图像拼接的示例

5. 图像分类

在图像分类中，需要对图像进行特征提取。由于图像显著性较大的区域一般是图像中信息量较大的区域，所以，对于有偏见采样来说，显著性较大的区域被优先选择来提取特征^[34~36]。因此，一般先对图像进行视觉显著性检测，得到显著图，然后根据显著图依概率提取特征，将特征包含的信息量最大化，并使特征具有较强的区分能力。

6. 目标识别

面对复杂视觉场景中的目标识别问题，如何减少处理的信息量、降低计算复杂度、实现实时的目标识别一直是要解决的难题之一。视觉显著性检测的提出为这一难题提供了一种解决方法，并取得了较好的效果。在应用中，根据视觉显著性的大小，首先从视觉场景中提取可能包含目标的感兴趣区域，再根据目标模型对该区域是否包含目标进行验证。将视觉显著性检测应用于目标识别中，可将信息处理资源优先分配给那些引起注意的关键区域，从而有效地降低计算量，同时，还可以避免背景区域的干扰，获得较好的检测和识别效果^[37~41]。

除了以上列举的 6 种应用，其他基于视觉显著性检测的应用还有图像检索、主动视觉、机器人视觉导航、场景分析和视频监控等。

1.4 本书的主要内容和章节安排

1.4.1 主要内容

本书主要对视觉显著性检测方法及其应用进行了研究。对于视觉显著性检测方法，我们首先对现有的流行检测方法进行了分析，发现这些流行检测方法目前存在 3 个问题：①方法的时间效率不高，即不能在检测准确率和运算时间上达到一个较好的平衡；②方法的通用性不强，视觉显著性检测方法按任务可分为两类，一类以预测人眼关注点为主，另一类则致力于均匀地渲染图像中的显著物体，现有的流行检测方法不能在这两类任务上同时有较好的表现；③方法对信息的使用不够全面，不能充分利用人类视觉系统的先验