

密级: _____



中国科学院大学
University of Chinese Academy of Sciences

硕士学位论文

图像的显著性区域检测技术研究

作者姓名: 叶天才

指导教师: 张冬明 副研究员

中国科学院计算技术研究所

学位类别: 工学硕士

学科专业: 计算机应用技术

培养单位: 中国科学院计算技术研究所

2015年05月

Research on Salient Region Detection

**By
Ye Tiancai**

**A Dissertation Submitted to
Graduate University of Chinese Academy of Sciences
In partial fulfillment of the requirement
For the degree of
Master of Computer Science**

**Institute of Computing Technology
Chinese Academy of Sciences
May,2015**

声 明

我声明本论文是我本人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢的地方外，本论文中不包含其他人已经发表或撰写过的研究成果。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了谢意。

作者签名：

日期：

论文版权使用授权书

本人授权中国科学院计算技术研究所可以保留并向国家有关部门或机构送交本论文的复印件和电子文档，允许本论文被查阅和借阅，可以将本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、汇编本论文。

(保密论文在解密后适用本授权书。)

作者签名：

导师签名：

日期：

摘要

随着数码相机以及移动设备的普及，图像已经逐渐成为人们记录日常生活的主要方式之一。对图像内容进行自动处理和分析，从而提取出特定的、有价值的信息，也成为一个亟待解决的问题。然而，对图像中一般物体的定位与检测一直是一个难点。

近年来受到广大学者关注的显著区域检测算法为解决上述问题提供了一个方向，通过对图像进行显著性分析，可以检测并定位人们感兴趣的物体，从而为一般物体的检测定位提供了极强的参考价值。本文围绕图像的显著性分析及其应用展开了研究，主要成果包括：

1. 深度调研了显著区域检测的各类算法。我们深入研究了显著区域检测的各类典型算法，包括基于局部对比度、基于全局对比度、基于频域分析，以及基于机器学习的算法，同时对比了各类算法的优缺点，为开发性能更好的新算法提供了努力方向。
2. 基于条件随机场的显著区域检测算法。我们首先提出了全局颜色对比度、全局颜色紧致度、全局颜色中心度以及局部颜色对比度四种显著性特征，随后通过条件随机场，我们有机互补的结合这几种特征，计算图像的显著图。实验表明，我们的方案能够充分的利用各类特征挖掘图像的显著性，在ASD数据集上取得了良好的效果。
3. 基于蒙特卡洛采样的显著区域检测算法。我们观察到，图像的显著区域在空间上的约束条件能够非常有效的优化检测效果，然而，这些空间约束（紧密性、连通性、包络性）都是二值化的概念，在现有的算法框架下无法实现。为此，我们提出了基于蒙特卡洛采样的显著区域检测框架，并加入了这些空间约束条件。我们在ASD,ECSSD两个公共数据集上进行了实验，实验表明，我们的算法在准确率，召回率上均能达到与经典算法相匹敌的程度，同时拥有较小的时间复杂度。
4. 显著区域检测在图像检索上的应用。脱离实际应用的算法将失去其价值，为此我们选取了图像检索这个应用场景，验证显著区域检测算法的实际应用价值。我们实现了一个基于经典的BOF模型的图像检索系统，并在该系统上结合了显著区域检测算法。实验表明，显著区域检测算法能有效提升图像检索的性能。

关键词：计算机视觉， 图像处理， 显著性区域检测， 目标检测， 图像检索

Research on Salient Region Detection

Ye Tiancai (Computer Application)

Directed by Associate Professor Zhang Dongming

As the population of digital camera and mobile devices, images have already become the frequently used method for people to record their lives. Automatically processing and analysing the images have also been the most wanted problems to be solved. However, general object detection which is many tasks' foundation is something hard to solve until now.

Recently, salient region detection has attracted many researchers' attention, and it provides an solution for general object detection. We made a deep research on this topic, and our mainly achievements are listed as follows:

1. A survey about salient region detection has been made. We investigate the state-of-the-art method on salient region detection until now, including local contrast based method, global contrast based method, frequency domain based method and machine learning based method.
2. Salient region detection based on conditional random field(CRF). We firstly propose some salient cues such as color contrast, color centering, color compactness etc. Then a CRF model is proposed to integrate these salient cues. Experiments show that our method is efficient and practical.
3. Salient region detection based on Monte Carlo sampling. Based on the observation that salient regions tend to be compact, connected and surrounded, However, concepts of spatial structure only have definite meanings in binary images. Thus, a Monte Carlo Sampling based Saliency model is proposed to utilize these features. Experimental results on two datasets show that, compared with eleven state-of-the-art methods, our approach has a competitive performance and also runs very fast.
4. Salient region detection applied to image retrieval. We evaluate the performance of salient region detection when applied to image retrieval. A BOF based image retrieval system is realized. Then we integrate salient region detction into this system. Experiments show that, salient region detection indeed raise the performance of image retrieval system.

Keywords: Computer Vision, Image Processing, Salient Region Detection, Object Detection, Image Retrieval

目 录

摘要	I
目录	V
图目录	IX
表目录	XI
第一章 绪论	1
1.1 课题研究的背景与意义	1
1.2 国内外研究现状	3
1.3 本文所做的工作	4
1.4 论文组织结构	4
第二章 显著性区域检测基础	7
2.1 基于局部对比度的方法	7
2.2 基于全局对比度的方法	8
2.3 基于频域分析的方法	10
2.4 基于学习的方法	10
2.5 最新工作与研究趋势	12
2.6 本文研究框架与思路	13
2.7 本章小结	14
第三章 基于条件随机场的显著性区域检测	15
3.1 引言	15
3.1.1 研究背景	15
3.1.2 研究动机	15
3.1.3 解决方法概要	16
3.2 显著性特征	17
3.2.1 全局颜色对比度	17
3.2.2 全局颜色紧致度	18

3.2.3 全局颜色中心度	18
3.2.4 局部颜色对比度	19
3.2.5 全局特征计算加速	19
3.3 条件随机场结合多特征方法	20
3.3.1 条件随机场简介	20
3.3.2 建模目标	21
3.3.3 我们的模型	22
3.4 实验结果	23
3.4.1 实验方法	23
3.4.2 评价标准	23
3.4.3 实验结果	24
第四章 基于蒙特卡罗采样的显著性区域检测	27
4.1 引言	27
4.1.1 研究背景	27
4.1.2 研究动机	27
4.1.3 解决方法概要	28
4.2 采样方法	28
4.3 生成显著图	30
4.3.1 生成距离图	30
4.3.2 生成初始二值标记图	30
4.3.3 生成最终的二值标记图	31
4.3.4 生成显著图	32
4.4 效率问题	32
4.5 实验结果	33
4.5.1 在ASD数据集上的结果	33
4.5.2 在ECSSD数据集上的结果	33
第五章 显著区域检测在图像检索上的应用	39
5.1 引言	39
5.1.1 基于内容的图像检索	39
5.1.2 研究动机与方法	39
5.2 基于词袋模型的图像检索系统	40

目 录

5.2.1	SIFT特征提取	40
5.2.2	建立码书	41
5.2.3	倒排索引	42
5.2.4	整体系统框架	42
5.3	利用显著区域检测优化结果	43
5.3.1	问题来源	43
5.3.2	优化方案	43
5.4	实验结果	44
5.4.1	数据集和评价指标	44
5.4.2	实验结果	44
第六章	总结与展望	45
6.1	研究工作总结	45
6.2	未来工作的展望	45
参考文献		47
致谢		i
作者简历		iii

图 目 录

1.1 计算机与人类对图像处理的差异	2
1.2 显著性区域检测示例	2
2.1 对比度在视觉感知中的作用	7
2.2 使用少量高频颜色量化图像	9
2.3 不同图像的log spectrum呈现为相似的曲线	10
2.4 通过求得log spectrum的残差得到显著区域	11
2.5 结合多个显著图	11
2.6 研究框架图	13
3.1 通过结合多个特征产生显著图	16
3.2 a.欧式距离度量颜色的不足 b.使用感知距离度量颜色的效果	17
3.3 颜色紧致度示例	18
3.4 图模型示例	21
3.5 二维晶格状模型	22
3.6 与其他10种方法的对比：准确率-召回率曲线	24
3.7 与其他10种方法的对比：F-score曲线	25
3.8 与其他10种方法的主观视觉对比	25
4.1 与其余10种方法的对比示例	28
4.2 系统框架图	29
4.3 包络性的作用	32
4.4 ASD数据集上与其他10种方法的对比：准确率-召回率曲线	33
4.5 ASD数据集上与其他10种方法的对比：F-score曲线	34
4.6 ASD数据集上与其他10种方法的主观视觉对比	34
4.7 ECSSD数据集上与其他10种方法的对比：准确率-召回率曲线	35
4.8 ECSSD数据集上与其他10种方法的对比：F-score曲线	35
4.9 ECSSD数据集上与其他10种方法的主观视觉对比	36
5.1 BOW模型	40

5.2 倒排索引示例	42
5.3 BOW模型框架图	43
5.4 检索示例	44

表 目 录

5.1 实验结果	44
--------------------	----

第一章 緒論

许多计算机视觉相关系统中，目标的检测与定位都是不可或缺的关键步骤，目标检测的精度将直接影响整个视觉信息处理的效果。对于人类而言，在一个视觉场景中迅速找到并定位自己所感兴趣的物体，是自然且非常简单的事情，然而应用计算机进行目标的自动检测仍旧面临许多难题，导致实际检测性能不高。为了简化问题，提高检测性能，满足实际应用需求，许多对象检测与定位方法，一般都是针对特定目标（比如人脸、车辆、行人等等）进行检测。近年来获得研究人员的广泛关注的显著性区域检测算法另辟蹊径，为类别无关的对象检测与定位提供更为有效的思路。本文针对显著性区域检测算法进行研究，采用概率统计方法，深入分析了复杂自然图像中影响显著区域检测的多方面因素，提出了更高效的特征提取、模型构建等技术手段与算法。

1.1 课题研究的背景与意义

图像作为记录信息、传递思想和表达情感的重要媒介，在现代人的日常生活中扮演了重要角色。另外随着智能手机和数码相机等硬件设备的普及，人们创造、获取图像的手段也日益方便与灵活。再加上近年来社交网络、微博、网络相册等共享平台快速兴起，以及图像本身所具有的内容直观、获取容易、传播方便、表现力丰富等优势，数字图像迅速成为日常生活中最受欢迎的一种知识传播和信息共享媒介[12]。对互联网上海量的数字图像进行处理与应用，挖掘出人们所需的信息，不仅会给人们带来许多便利，同时也蕴含了极大的商业价值。

然而，计算机对图像的处理方式与人类有着极大的不同。计算机以像素为基本单位，通过记录像素的颜色以及像素间的排列关系来存储、处理图像，这是一种非常低阶的处理手段；而人类则是通过高阶概念，如场景、物体、及物体间相互关系为基本处理单位，对图像进行认知和记忆。这种本质上的差异，导致目前计算机对图像的理解与处理远远不能与人类相比。因此，从某种程度上来讲，要想有效减小计算机与人类在图像处理方面的差距，必须首先解决计算机从离散的像素到更高层次的认知单位的映射问题。譬如传统的图像分割，则是希望将一群离散的像素映射组织为若干有共同属性、含义的基本语义单元。

更为重要的是，在当前大数据时代的背景下，对海量图像进行快速有效的处理更加成为了一个亟待解决的问题。然而，正如前面所讲到的，如果继续以像素为基本处理单位，很难实时有效的处理这些图像数据。假如我们能够将图像中人类感兴趣的区域提取出来，不仅减小了后续处理的工作量，同时也从语义级别对像素进行了一次映

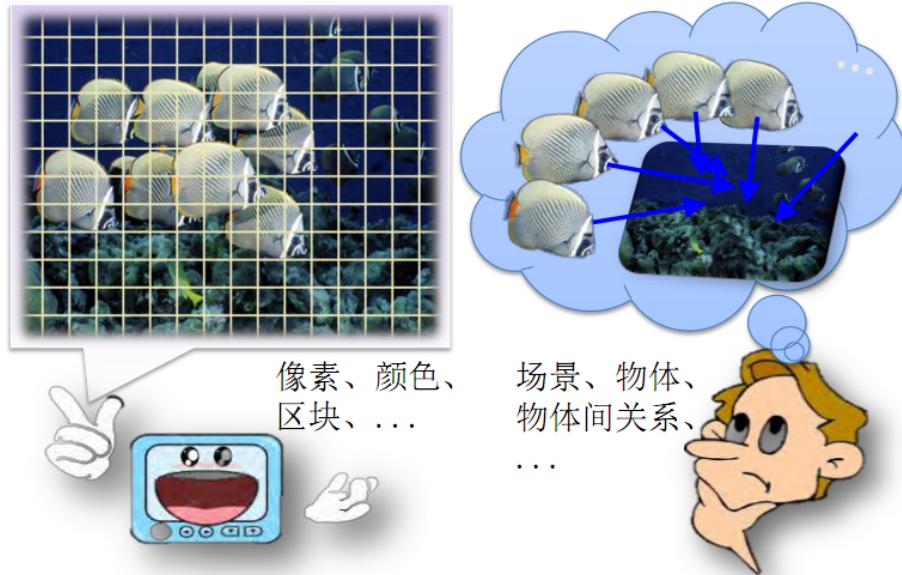


图 1.1: 计算机与人类对图像处理的差异

射，这将是大有裨益的。

在这样的需求推动下，显著性区域检测成为了一个研究热点。所谓显著性区域，即一幅图像中最能吸引人类注意力的区域，通常为一幅图像的前景物体。而显著性区域检测的目标，即通过一个与原图像大小相同的二值图像，来标明哪些区域是显著的。如图1.2所示，(a)是源图像，(b)为一个二值标准图像(Ground Truth)，用来标示哪些像素点是显著的。当然，一般的显著性区域检测方法不可能做到精确的标示哪些像素显著与否。因此，显著性区域检测的方法大多产生一个灰度图像，其中灰度值越高，表明这个像素的显著程度越高，如图1.2.(c)(d)(e)所示，分别为FT[4],HC[13],RC[13]三种方法所产生的显著图。

作为一种通用物体的定位与检测手段，显著区域检测可以很好的与其他应用相结合以提升效果，比如图像检索[43][16]，图像的自动裁剪[40][15]，自适应图像压缩[14]，以及图像分割[23][18]，等等。正是由于其广泛的应用，显著区域检测得到了来自许多不同领域学者的关注，对显著性区域的准确快速的定位，将会对计算机视觉、多媒体内容分析等领域产生十分积极的影响。

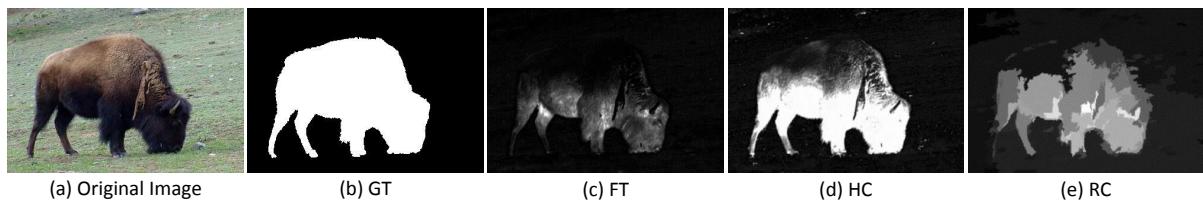


图 1.2: 显著性区域检测示例

1.2 国内外研究现状

对于人类视觉注意力的研究，可以追溯到Koch和Ullman[27]的基于生物视觉特性的计算模型。随后，显著区域检测的开创者Itti[22]在此模型的基础上进行了改进，通过结合多尺度的图像特征，在快速场景识别上取得了很好的效果。自此以后，显著区域检测就吸引了大批学者的注意。

在最初的阶段，人们受到生物视觉特性的启发，即与周围环境差异较大，有较强对比度的区域通常更能引起人们的注意，因此人们大多采用基于局部对比度的方法。在Ma和Zhang的工作中[30]，基于局部对比度的分析方法被首次正式提出，通过进一步结合模糊集理论（Fuzzy Theory），就可以得到一幅图像的显著区域。在Harel等人的工作中[19]，根据像素点之间的相似性，首先构建了一个图（Graph），随后通过马尔科夫过程的收敛性，挖掘图像区域的独特性，并对差异较大、较为独特的区域赋予较大的显著值。其他的一些学者，比如Liu[28]和Mai[31]等人，同样也将局部对比度作为一个重要的特征来发掘显著性区域。然而，基于局部对比度的方法有一个很大的缺陷，他们倾向于给物体边缘赋予较大的显著值，而并非高亮整个显著区域。

最近的几年，基于全局对比度的方法越来越流行，他们能够高亮整个区域，而并非边缘。Zhai等人[48]提出了第一个基于全局对比度的模型，一个像素的显著性被定义为与图像中所有其他像素的差异度之和。考虑到计算性能问题，他们仅仅采用了像素的亮度信息来度量像素之间的差异，而忽略了颜色信息。Cheng等人[13]注意到了这一点，充分利用了一个像素的三个颜色通道来度量像素差异。为了使计算足够高效同时又保持良好的效果，他们提出了两个解决方法，即基于颜色直方图的加速和色彩空间平滑。在Cheng等人的工作发表后，学者们注意到了基于全局对比度方法的良好效果与计算可行性，自此以后，大多最近的工作都将全局对比度作为重要的特征用来提取显著区域。

除了上述基于对比度的模型（即挖掘像素点或区域的独特性），还有另外一些基于频域分析的模型[4][21][20]，这些模型通过将图像映射到频域进行分析，并结合信号处理与数据压缩理论，能提取出一幅图像中的独特部分，有着良好的理论支持。然而，[20]指出，这些基于频域分析的模型在某种程度上等同于一个局部梯度算子并叠加一个高斯模糊，因此在检测较大的显著区域时表现较差。

另外，近年来随着机器学习的流行，也有一些学者尝试基于学习的模型。Kienzel等人[25]利用支持向量机(SVM)从已有眼动数据学习得到了一个显著区域检测模型。Mai等人[31]使用条件随机场(CRF)结合多个显著图，并充分利用了相邻像素之间的关联性来建立模型。尽管这些基于学习的方法在现有的数据集下能取得不错的效果，但是他们通常都十分耗时，同时由于机器学习对于学习数据的强依赖性，导致这类方法在不同数据集下的表现也存在较大的差异。

可以看到，国际上对于显著性区域检测的研究正呈现百花齐放的状态，各个学者

对显著性区域检测都有不同的理解与思考，同时也产生了完全不同的算法与思路，这也恰好说明了人们对于显著性区域检测这一课题还处于初级探索阶段，还未形成统一的见解与共识。

1.3 本文所做的工作

作为一种通用的物体检测手段，显著性区域有着极其广阔的应用空间[43][16][15][23]。在此如此巨大的需求推动下，本文围绕显著性区域检测的算法以及相关应用进行了研究，主要工作与成果如下：

1. 通过对不同特征的分析，提出了基于多特征结合的显著性区域检测算法。该算法首先提出了四个基本的显著性区域特征，随后通过条件随机场（CRF）建模，学习多个特征在多个尺度下的相互关联与互补关系。算法能够有效的检测图像中的视觉显著区域，并且在目前国际标准数据集上取得了良好的效果。为了验证该算法的性能，在两个国际标准数据集上，我们对比了国际上现有的十余种经典算法，实验结果显示，我们的算法具有很高的精确率与召回率，同时还有相对鲁棒的特性。
2. 通过观察空间约束对显著性区域检测的影响，我们提出了基于蒙特卡洛采样的显著性区域检测算法。该算法提出了三种有效空间约束关系：紧致性、连通性、包络性。为了有效的利用这三种空间约束关系，我们创造性的提出了利用蒙特卡洛采样，将空间约束加入现有特征中的方法。同时，这一方案有着非常高的时间效率，而采样的高度可并行化，使得实时计算显著性区域成为可能，大大提高了算法的实用价值。
3. 显著性区域检测的真正价值在于它的应用，我们尝试将显著性区域检测应用在人脸识别和图像检索中，并探讨了一些可行性方案，最后通过实验论证了显著性区域检测所带来的优化效果。

1.4 论文组织结构

本文分为五章，主要结构和内容如下：

第一章首先阐述了显著性区域检测的研究背景和意义，接着介绍了该方向的国内外研究现状以及存在的问题，最后概述了本文所做的工作和论文的组织结构。

第二章介绍了显著性区域检测的基础特征与算法，并分类探讨了目前国际上各类主流算法的思路与不足。在这一章的末尾，还介绍了该领域最新的研究进展与趋势。

第三章介绍本文提出的，基于条件随机场的多特征结合算法，并通过实验详细比较了该算法的优势与不足。

第一章 绪论

第四章介绍本文提出的，基于蒙特卡洛采样、结合空间约束特征的算法。

第五章探讨了显著性区域检测应用在人脸识别和图像检索上的价值，对存在的问题和局限性进行了讨论，最后对全文进行了总结，同时展望未来的研究工作。

第二章 显著性区域检测基础

在本章中，将分类介绍几种国际上主流的几类显著区域检测方法：基于局部对比度 (MZ[30])，基于全局对比度 (HC, RC[13])，基于频域分析 (SR[21])，基于学习 (CRF[31])。在本章的最后，还将介绍近年来的一些最新的工作和研究趋势。

2.1 基于局部对比度的方法

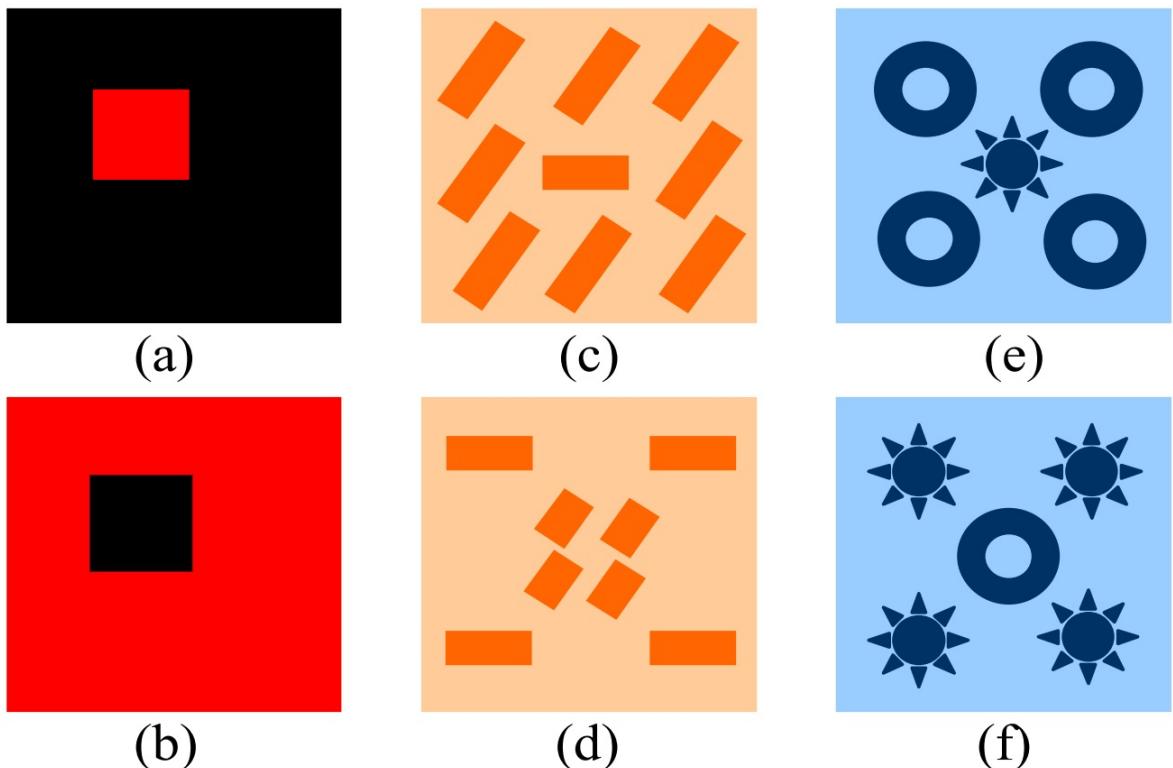


图 2.1: 对比度在视觉感知中的作用

传统的图像处理技术通常只考虑了颜色、纹理和形状这三个基本特征，然而，对比度在视觉感知中的重要作用却常常被人们所忽略。以下这个例子可以说明对比度的重要性[30]，如图2.1所示的三组合成图像，在图2.1(a)中，有一个红色的方形盒子被黑色的背景所围绕。很显然，红色盒子所在区域更容易吸引人们的注意力，对这个现象的一个解释是，因为红色是一种更容易引起人们视觉注意力的鲜艳色彩。然而，图2.1(b)却无法支持这个假设，显然在这幅图像中，黑色的区域却成为了吸引人们注意力的区域。这个现象表明，尽管颜色能够影响人们的视觉感知，但是却并非影响视觉注意力最重要的因素。图2.1(c)(d) 展示了两幅有一定纹理方向的图像，类似(a)(b)两

图, 说明了纹理并非视觉感知中最重要的因素。图2.1(e)(f)则说明了形状在视觉感知中的作用。然而, 以上三组图像都有一个共同的特征, 那就是, 显著区域通常被背景区域围绕并呈现高对比度(无论是颜色、纹理还是形状上的高对比度), 这在一定程度上说明了高对比度的物体通常能吸引人眼的注意力。基于局部对比度的方法利用了图像区域相对于(一个小的)局部邻域的稀缺度来检测显著性区域。在此定义下, 检测子的形式通常表现为“中心-周围差异”, 不同方法的变化在于: 1) 尺度问题(即“中心-周围差异”中, “周围”的大小); 2) 差异的度量问题(在哪种特征空间下比较、周围像素点的权值等)。本节以MZ[30]方法为例, 介绍基于局部对比度的方法。在MZ中, 局部对比度被定义为:

$$C_{i,j} = \sum_{q \in \Theta} d(p_{i,j}, q) \quad (2.1)$$

其中 $c_{i,j}$ 代表在 (i, j) 位置的像素点的对比度, $p_{i,j}$ 代表该像素点的某种特征向量(例如颜色向量), q 代表周围像素点的特征向量, $d(x, y)$ 表示 $p_{i,j}$ 与 q 的差异, 通常使用欧式距离度量, Θ 即代表周围像素点的集合, 通过控制 Θ 的大小, 即可控制“感知域”的大小。最后, 将所有图像的对比度的值normalize到0,1之间, 即得到像素的显著值。可以看出, 这个定义下, 一个像素点的显著值实际上就是它与周围像素点的差异和。

实际上, 对这个公式做一些变形, 就可以衍生出其他的方法: 1) 调整 Θ 的范围, 就可以调整检测子的尺度。近期的一些工作, 已经不满足于通常的矩形或圆形邻域, 开始寻找一些特性更好的邻域(比如邻域边界与图像边缘一致的邻域)。2) 改换特征空间, 比如在梯度空间下, 或者其他自定义的特征空间。3) 加入权值, 公式2.1认为周围像素点对中心像素点的贡献是一样的, 权值都为1, 但是实际上, 与中心像素点距离近的像素可能对其对比度的贡献更高, 所以可以对距离中心近的像素点赋更大的权值。

以上就是基于局部对比度的显著区域检测方法的核心。然而, 这种方法有一个很严重的问题, 它倾向于给物体边缘赋予较大的显著值, 因为物体边缘的“中心-周围差异度”通常较大。在尺度选的较小时, 检测子实际上已经近似退化为一个边缘检测子。

2.2 基于全局对比度的方法

在上一节提到, 基于局部对比度的方法, 倾向于给物体边缘赋予较大的显著值, 而无法均匀的高亮整个显著区域。为了解决这个问题, 研究者想到采用全局对比度计算显著图。相比基于局部对比度的方法, 基于全局对比度的方法就不存在只高亮边缘的问题, 因为在全局环境下(即将像素点的邻域扩展到全图), 具有相同特征的像素(如颜色)必然赋予相同的显著值。实际上, 早在2006年, Zhai[48]等人就已经提出了全局对比度这个概念, 只是当时他们考虑到计算复杂度的问题, 仅仅采用了像素的亮

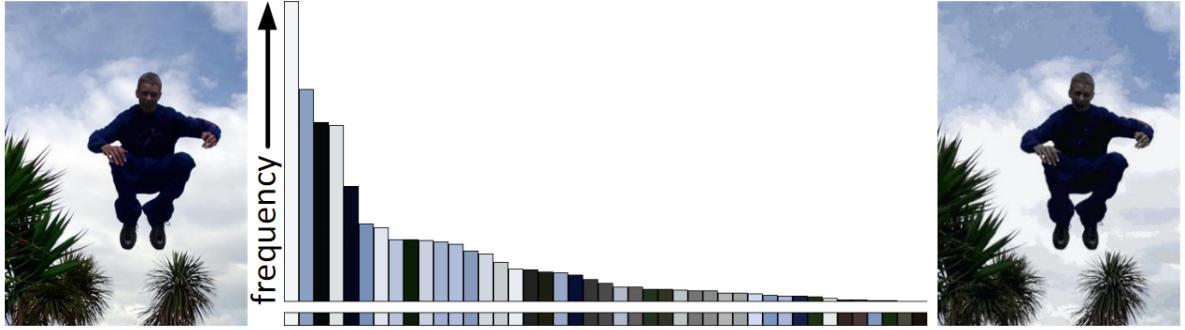


图 2.2: 使用少量高频颜色量化图像

度值来计算像素点之间的差异。显然，用一个像素点的亮度值来度量像素点之间的相似性是非常粗糙的，比如纯红和纯蓝的亮度值是一样的，但是这两个颜色显然有着很大的差别。因此，在Cheng的工作中[13]，使用了颜色的三个通道进行计算。在HC方法中，一个像素点 I_k 的显著值定义为：

$$S(I_k) = \sum_{\forall I_i \in I} D(I_k, I_i) \quad (2.2)$$

其中 I 代表整个图像像素点的集合， I_i 代表第*i*个像素点在Lab颜色空间的颜色向量， $D(*, *)$ 为欧式距离算子。很明显，在这个公式中，颜色相同的像素肯定会得到相同的显著值，为了进一步减少计算开销，上式可以进一步简化为：

$$S(I_k) = S(c_l) = \sum_{j=1}^n f_j D(c_l, c_j) \quad (2.3)$$

其中 c_l 代表像素 I_k 的颜色向量， c_j 代表第*j*个颜色的颜色向量， f_j 代表第*j*种颜色在整个图像中出现的频率。经分析可知，计算上式的时间复杂度为 $O(N) + O(n^2)$ ， N 为像素的个数， n 为整个图像中不同颜色的数量。如果不经过任何优化，原颜色空间的颜色个数为 $n = 256^3$ ，显然这样的时间复杂度是不可忍受的。因此，Cheng首先将每个颜色通道量化为12个值，这样颜色的数量下降为 $n = 12^3 = 1728$ 。进一步的，如图2.2所示，忽略占像素数量较少的颜色，使用出现频率高的颜色替代，使得剩余的颜色占到像素数量的95%，颜色的数量可以进一步下降到85左右（忽略的颜色使用剩余颜色中最相近的代替），同时基本不影响图像的主观视觉质量[13]，这样就大大降低了时间复杂度。然而，如此量化后，会给图像带来很明显的量化痕迹，比如原本非常相近的两个颜色，可能会被量化到不同的值。因此，为了优化效果，Cheng最后还使用了一个颜色空间的平滑，即对相近颜色的显著值进行了加权平均。具体做法如下：

$$S'(c) = \frac{1}{(m-1)T} \sum_{i=1}^m (T - D(c, c_i)) S(c_i) \quad (2.4)$$

其中 $T = \sum_{i=1}^m D(c, c_i)$ 是颜色 c 与最相近的 m 种颜色的距离和。文章中还介绍了RC方法，即Region Based Contrast，该方法首先使用一般的分割方法将图像分割为许多区域，然后使用该区域的平均颜色作为区域的颜色值，最后使用相似的方法，以区域为基本计算单元，计算区域的全局对比度。

2.3 基于频域分析的方法

基于频域分析的方法主要将图像转化到频域处理，并结合信息压缩理论，将图像中“新颖”的频率部分分离出来，最后再转回空间域。这里以SR[21]为例子进行分析。在信息压缩理论中，一幅图像的信息量可以被分解为两部分：

$$H(\text{Image}) = H(\text{Innovation}) + H(\text{Prior Knowledge}) \quad (2.5)$$

对于 $H(\text{Prior knowledge})$ 是我们已知的一些先验知识，所以可以不用编码，只需对 $H(\text{Innovation})$ 这部分进行编码传输即可。解码时，由于先验知识已知，我们通过解码 $H(\text{Innovation})$ 这部分，就可以恢复原图像，从而达到信息压缩的目的。而显著性区域即对应 $H(\text{Innovation})$ 这部分内容，这也是基于频域处理方法的理论基础。

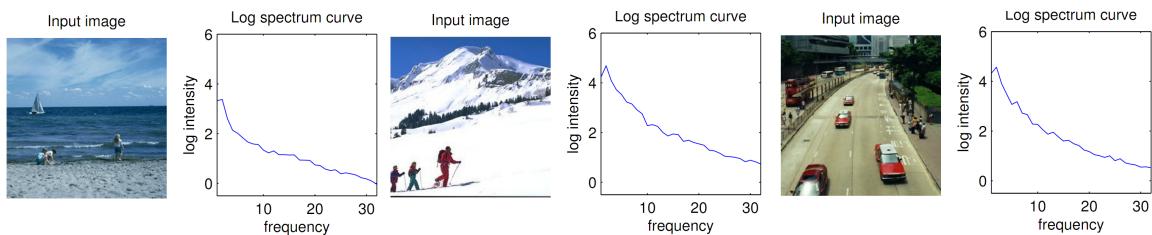


图 2.3: 不同图像的log spectrum呈现为相似的曲线

文章指出，图像的log spectrum具有相似的变化曲线，如图2.3所示，这也就是对应的先验知识，因此通过求得一大批图像log spectrum曲线的均值，我们就可以得到图像在频率域的“先验知识”。如图2.4，将图像的log spectrum减去先验，得到的残差，即为显著区域对应的频域的表达，再通过逆变化，变换到时域，即可得到显著图。

然而有文章[20]指出，这些基于频域分析的方法实际上等同于基于局部对比度的方法加上一个高斯模糊，因此与基于局部对比度的方法具有相同的缺陷，即容易高亮物体的边缘。

2.4 基于学习的方法

这里介绍Mai等人[31]提出的使用条件随机场(CRF)的模型。作者基于这样的观察：既然现在有这么多的显著区域检测方法，而且不同的方法针对不同的场景都会有一定的效果，那么如果能够将这些方法结合起来，肯定能够互相补充，得到更好的结

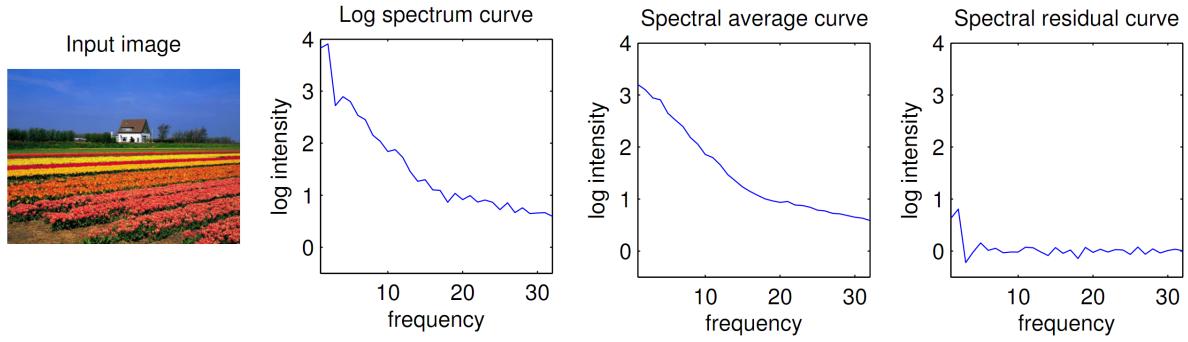


图 2.4: 通过求得log spectrum的残差得到显著区域

果。如图2.5所示，(c)(d)(e)(f)四种方法均能取得一定的效果，但是都有各自的缺陷，比如(f)只高亮了显著区域的边缘，而(e)方法对显著和非显著区域的区分不明显，而通过CRF结合多种方法相互补充，可以看到CRF-GIST方法可以取得很好的效果。实际上，这个思想在Itti的学生Borji的一篇文章中[10]就已经涉及到了，只是在Borji的文章中，结合多种特征的方式比较简单，即采用简单的加权平均，或者对应像素相乘等等。Mai等人指出[31]，简单的相加相乘并没有考虑到相邻像素之间的关联性，即如果能加入空间位置信息（相邻且相似的像素的显著性应该比较接近），应该能取得更好的效果。

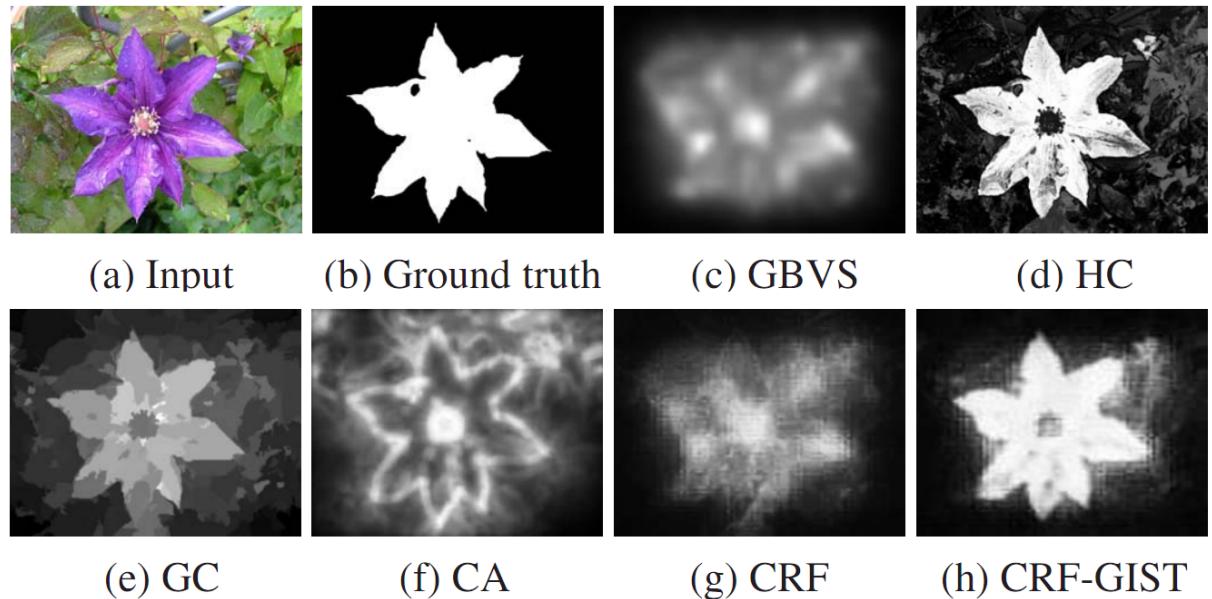


图 2.5: 结合多个显著图

在Mai的模型中，给定一幅图像 I ，我们使用一个二值掩图 $Y = \{y_p | p \in I\}$ 来标出显著物体。在CRF模型中，以每个像素点为顶点，8领域像素点之间相连，形成一个晶格

状的图。那么，在这个图下， Y 在给定图像 I 下的条件概率可以写为，

$$P(Y|I) = \frac{1}{Z} \exp\left(\sum_{p \in I} F_d(y_p, I) + \sum_{p \in I} \sum_{q \in N_p} F_s(y_p, y_q, I)\right) \quad (2.6)$$

其中 p 代表 I 中的一个像素点， y_p 是其显著与否的标记。 $F_d(y_p, I)$ 是node feature function， $F_s(y_p, y_q, I)$ 是edge feature function，描述了相邻像素点之间的联系。

Node feature function仅仅与输入的已有显著图 S_i 有关，即

$$F_d(y_p, I) = \sum_{i=1}^m \lambda_i S_i(p) + \lambda_{m+1} y_p \quad (2.7)$$

其中 λ_i 是条件随机场的一组待学习的参数。Edge feature function描述了相邻像素之间的关系：

$$F_s(y_p, y_q, I) = F_e(y_p, y_q, I) + F_c(y_p, y_q, I) \quad (2.8)$$

其中 $F_e(y_p, y_q, I)$ 考虑到了这样一个事实，如果两个像素点在同一幅显著图中的显著值差别很大，那么他们最后倾向于拥有不同的显著性标记。

$$F_e(y_p, y_q, I) = \sum_{i=1}^m \alpha_i (\mathbf{1}(y_p = 1, y_q = 0) - \mathbf{1}(y_p = 0, y_q = 1))(S_i(p) - S_i(q)) \quad (2.9)$$

其中 α_i 为CRF的待学习参数， $\mathbf{1}(\cdot)$ 为指示函数（括号内真为1，假为0）。 F_s 的第二项可以看做一个惩罚项，当像素点的颜色相似却被标上不同的标记时要进行惩罚：

$$F_c(y_p, y_q, I) = \mathbf{1}(y_p \neq y_q) \exp(-\beta \|I(p) - I(q)\|) \quad (2.10)$$

其中 $\|I(p) - I(q)\|$ 代表像素 p 和 q 的颜色差（Lab颜色空间）， β 被设置为 $(2 < \|I(p) - I(q)\| >^2)^{-1}$ ， $< \cdot >$ 代表计算期望。

最后，通过训练这个模型，得到所有参数的最优值。当计算新的显著图时，我们取每个顶点（像素点）被标记为1的概率作为该像素点的显著值。

可以看出，基于学习的方法原理较为复杂，同时学习与训练比较耗时。另外，[28]指出，基于学习的方法比较依赖训练数据集，因此在不同的数据集合上，表现差异较大。

2.5 最新工作与研究趋势

除了以上一些经典的方法，近年来在该领域也出现了许多新颖且实用的工作[10][46][41][45]，大家对显著性区域检测的研究方向有了如下一些共识和趋势：

- 多尺度，多特征的融合。在小尺度下，一些很小的，对比度又很大的区域很容易影响到整个图像的显著性检测，而在大尺度下，则可以屏蔽这些区域，但是检测

的粒度又会变得粗糙。因此，通过结合多个尺度，可以大大提高检测的精度和鲁棒度。对于多特征，前面已经说过，单一的特征都有其适用范围，不可能对于所有情况都有效，因此结合多个特征（颜色，纹理等等）可以使之互补。

- 趋向于基于区域的特征。基于点的特征比较敏感，很容易受到噪声点的影响。最近的一些工作，基本都是基于区域的特征。尤其值得注意的是，大多使用了SLIC[5]这个超像素分割算法，将空间上近似的像素点分割为一个小的区域进行计算。
- 加入一些更高层的特征。底层的特征在简单场景下有良好的效果，但是在复杂场景下往往失效，加入高层特征（如人脸识别子，汽车识别子等等）有助于改善结果。
- 将时间复杂度作为一项重要的考察指标。显著性区域检测的价值在于将其与其他应用相结合，单独的使用显著性区域检测算法没有任何意义。作为一项预处理技术，其时间复杂度决定了其实用性。

2.6 本文研究框架与思路

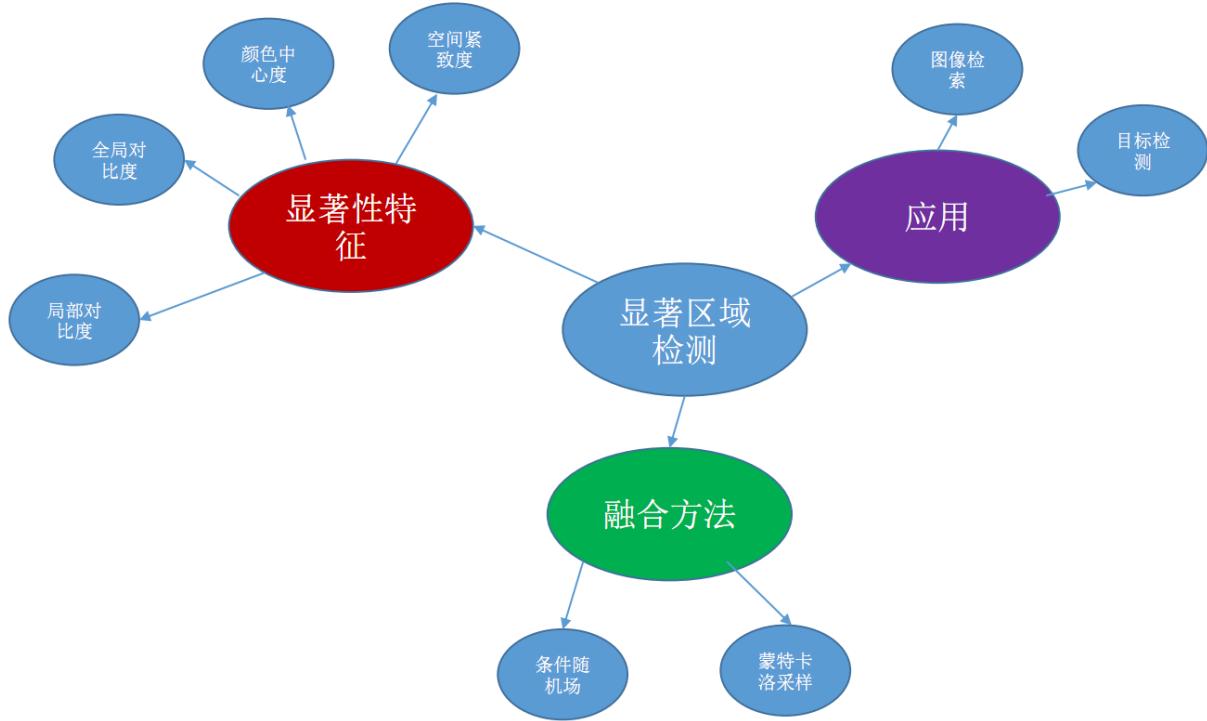


图 2.6: 研究框架图

本文研究框架如图2.6所示，围绕显著区域检测，主要研究了显著性特征、融合方法及其应用。我们按照如下思路开展我们的研究工作：

1. 首先解决人眼视觉注意力的机制，即什么样的图像区域能引起人眼的视觉注意（显著性特征）。我们提出了多个符合人眼视觉注意力机制的，且可计算的显著性特征，包括局部对比度、全局对比度、颜色中心度、空间紧致度等等。
2. 接着解决如何将多个特征融合起来的问题。各个特征都具有特定场景下的特定意义，因此将这些特征结合起来，产生互补效应以提高检测效果，是非常重要的。针对不同的应用场景和需求，我们提出了条件随机场和蒙特卡洛采样两种特征融合方法。
3. 最后，我们解决显著区域检测的应用问题。显著区域检测作为一项计算机视觉的预处理技术，如果无法应用于实际场景中，将失去其存在的价值。我们重点研究了显著区域检测在图像分割、图像检索中的应用。

2.7 本章小结

本章介绍了显著性区域检测的四类经典算法：基于局部对比度、基于全局对比度、基于频域分析和基于学习的方法。基于局部对比度的方法符合人类对生物视觉的认知，然而却容易高亮区域边缘；基于全局对比度的方法则改良了上述缺点，能较好的高亮整个显著区域；基于频域分析的方法拥有良好的理论基础，但是其处理效果同基于局部对比度的方法一样，容易强调边缘的显著度；最后基于学习的方法能有效的结合各类特征，得到良好的效果，但是其训练时间过长，且与数据相关。另外，还介绍了国际上关于该课题最新的一些思路和研究趋势，这为我们未来的工作指明了方向。在本章的最后，我们提出了自己的研究框架与思路，在接下来的章节中，我们将按照该框架图开展我们的研究工作。

第三章 基于条件随机场的显著性区域检测

3.1 引言

3.1.1 研究背景

视觉显著性是由于物体、人或者像素相较于其近邻在某种特征上更为突出，从而吸引人眼注意力的一种特性。视觉显著性的信息可以有益于其他很多应用，比如图像检索[43][16]，图像的自动裁剪[40][15]，自适应图像压缩[14]，以及图像分割[23][18]，等等。

正是由于其广泛的应用场景，显著性区域检测吸引了各个领域的众多学者。目前绝大多数算法都是自底向上的计算模型，通过计算局部或者全局对比度来生成显著图。

对于基于全局对比度的方法，一个区域的显著度是通过计算其与整个图像其他区域的差异度来得到的。在Zhai和Shah的工作中[48]，一个像素的显著性是通过计算其与整幅图像中其它像素的对比度来得到的，由于时间复杂度的问题，他们仅仅采用了亮度信息而忽略了其他颜色通道。显然，仅仅通过亮度信息来计算像素之间的差异性是一个非常粗略的抽象。Cheng等人[13]完善了这个工作，他们尝试使用完整的颜色信息去评估像素之间的差异性，为了降低计算复杂度的同时尽量减少失真，他们引入了直方图加速和彩色空间平滑两个工具。他们的方法在性能上远远超越了Zhai的方法。

基于局部对比度的方法则通过计算图像区域的独特性来得到其显著度，一般是计算区域与周围一个小的邻域的差异性。这个计算模型非常符合生物视觉特性，由Koch和Ullman等人最先提出[27]，之后由显著性区域检测领域的先驱者Itti等人加以完善[22]，将模型改为计算在多尺度下的中心-周围差异，并取得了良好的效果。在Ma和Zhang等人的工作中[30]，使用模糊集理论(fuzzy growing)改良了该算法，取得了更好的结果。其他很多学者，包括Harel等人[19]，以及Liu等人[28]，都将局部对比度作为一个很重要的特征。

3.1.2 研究动机

尽管基于局部对比度和全局对比度的方法都能取得一定的效果，但这两类方法都存在先天的缺陷。基于全局对比的方法由于是在全局域计算对比度，因此相同颜色的像素必然会被赋予相同的显著值，不管他们是处于前景还是背景中，这显然没有考虑到各个像素所处的周围环境；基于局部对比度的方法虽然考虑了像素的周边环境，但

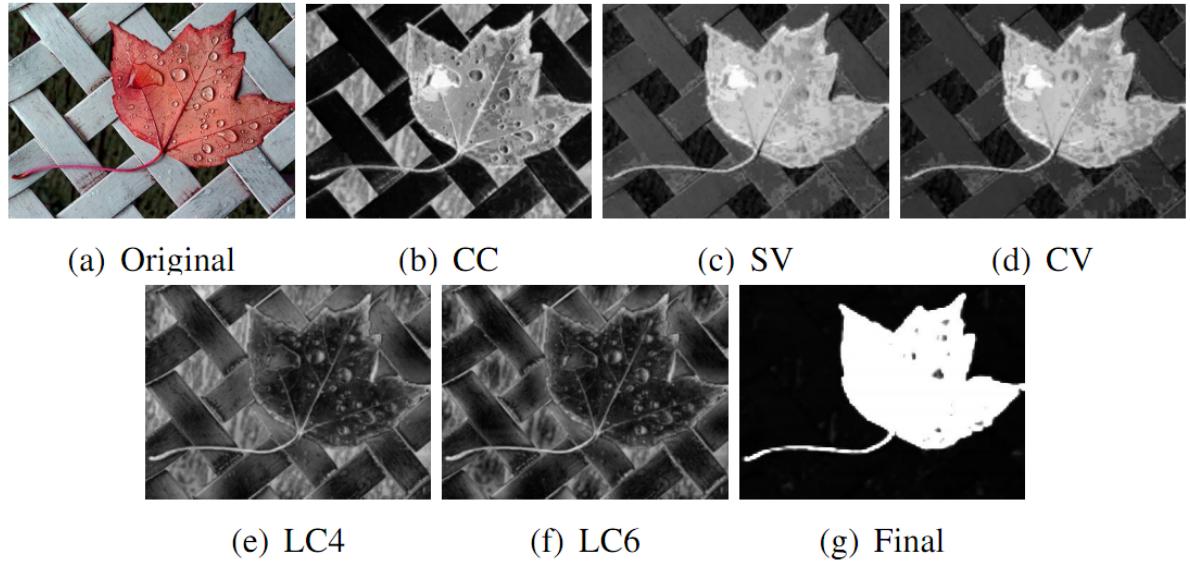


图 3.1: 通过结合多个特征产生显著图

是却倾向于赋予区域边缘更高的显著度，在邻域缩小到一定程度时，其退化为边缘检测算子。在我们的工作中，我们希望有一种显著性区域检测算法能达到如下效果：

1. 能够均匀的高亮整个显著区域，从而有利于后续处理，完整的分离出显著区域。
2. 能根据区域在不同背景下，赋予不同的显著值，从而区分前景与背景中具有相同颜色的像素。
3. 算法尽量鲁棒，能针对各种不同场景的图像均具有一定效果。

在我们的研究中发现，局部和全局的显著性区域检测算法在某些方面是互补的，如果我们能够有机的结合不同的特征，应该可以得到更好，更鲁棒的结果。

3.1.3 解决方法概要

通过对以往经典方法的观察，我们意识到，单独依靠某一类方法，将会存在先天的不足与缺陷。在我们的方法中，首先提出了多种显著性特征产生不同的显著图（包括全局颜色对比度，全局颜色紧致度，全局颜色中心度，两种尺度下的局部颜色对比度），然后我们通过条件随机场(CRF)有机的结合这些特征，并结合相邻像素之间的关联性（相邻且相似的像素应该具有相近的显著度），达到了提高检测准确度的目的，如图3.1所示，由原图像(a)，在不同的特征下，我们产生了5副显著图，通过训练好的条件随机场模型，我们得到了最终的显著图(g)，可以看出，我们的方法能够产生高质量，区域连续的显著图。

3.2 显著性特征

3.2.1 全局颜色对比度

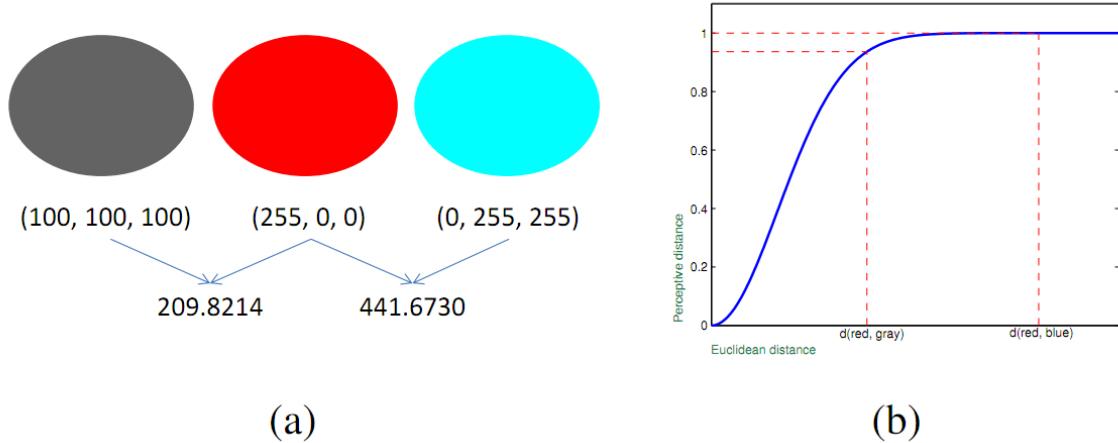


图 3.2: a. 欧式距离度量颜色的不足 b. 使用感知距离度量颜色的效果

与以往基于全局对比度的方法不一样的是，我们对颜色差异的度量进行了更加深入的研究。如图3.2.a所示，三种颜色，从人眼看来，差别都非常大，大多数人都只能说这三种颜色差异很大，却无法区分哪两种颜色之间的差异更大。然而，如果使用欧式距离进行度量，红色与蓝色的差异是红色与灰色差异的两倍，这显然是不符合人眼对色彩的感知的，使用这种不符合人眼感知的度量方式进行对比度的计算，效果上必然也会存在一定差异。

因此，我们提出了基于人眼感知的色彩差异度量方法。如下所示，我们将两种颜色的差异度定义为：

$$D(c_l, c_j) = 1 - \exp\left(-\frac{d(c_l, c_j)^2}{2\sigma}\right) \quad (3.1)$$

其中 $d(c_l, c_j)$ 代表颜色 c_l 与 c_j 之间的欧式距离， σ 表示该图像中颜色的方差。如图3.2.b所示，使用我们提出的感知距离度量后，这种差异将更加符合人眼的认知。

在这个定义下，我们把基于全局颜色对比度的显著度定义为一个像素与图像中所有其他像素的差异：

$$S_{cc}(I_k) \propto \sum_{\forall I_i \in I} D(I_k, I_i), \quad (3.2)$$

很显然，从公式3.2中可以看到，相同颜色值的像素必然具有相同的显著值，因此，我们可以将公式变为如下形式，以便加速计算过程：

$$S_{cc}(I_k) \propto \sum_{j=1}^n f_j D(c_l, c_j), \quad (3.3)$$

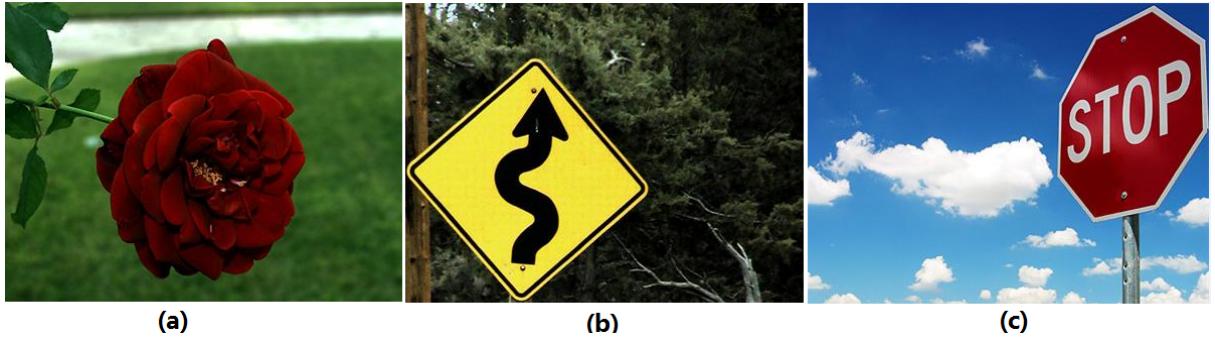


图 3.3: 颜色紧致度示例

其中 c_l 代表像素 I_k 的颜色向量, n 代表不同颜色的数量, f_j 代表颜色 c_j 在图像I中出现的频率。使用公式3.1中对颜色差异的度量, 我们得到最终的显著值计算公式:

$$S_{cc}(I_k) \propto \sum_{j=1}^n f_j D(c_l, c_j), \quad (3.4)$$

其中 $D(c_l, c_j)$ 即代表颜色 c_l, c_j 之间的感知差异, S_{cc} 代表像素 I_k 的显著度。

3.2.2 全局颜色紧致度

除了颜色对比度, 还有一些空间信息可以用来区分前景和背景像素 (即显著区域与非显著区域)。根据我们的观察, 显著区域的像素点通常在空间位置上分布比较紧致, 而背景像素则在图像中分散开来。如图3.3(a)所示, 红色花朵为显著区域, 红色相对的集中于同一区域, 表现为更加紧致集中, (b)(c)两图可以观察到类似的现象。概括来讲, 这三图的显著区域的颜色相对分布集中, 而背景像素的颜色则分布较为杂乱。在我们的全局颜色紧致度中, 我们使用某种颜色的像素在空间位置上的方差来衡量像素的显著性, 方差越小, 说明颜色越倾向于集中分布在某一块区域, 我们认为它越有可能是属于显著性区域的像素。

$$U_{sv}(I_k) \propto \frac{1}{m} \sum_{i=1}^m (x_{c_l}^{(i)} - \bar{x}_{c_l})^2 + (y_{c_l}^{(i)} - \bar{y}_{c_l})^2 \quad (3.5)$$

$$S_{sv}(I_k) = 1 - U_{sv}(I_k) \quad (3.6)$$

其中 c_l 是像素点 I_k 的颜色向量, m 是 c_l 颜色的像素数量, $\{x_{c_l}^{(i)}, y_{c_l}^{(i)}\}$ 是这些像素的空间位置坐标, $\{\bar{x}_{c_l}, \bar{y}_{c_l}\}$ 则是他们的平均坐标。 $U_{sv}(I_k)$ 代表了像素点是非显著的程度, 我们把它归一化到0和1之间, 然后通过式3.6就可以得到图像的显著图。

3.2.3 全局颜色中心度

与全局颜色紧致度类似, 我们观察到, 人眼倾向于注意位于图像中心的物体和区域, 也就是说, 空间分布上越靠近图像中心的颜色, 越有可能成为显著性区域。因此,

我们使用如下式子定义：

$$U_{cv}(I_k) \propto \frac{1}{m} \sum_{i=1}^m (x_{c_l}^{(i)} - x_c)^2 + (y_{c_l}^{(i)} - y_c)^2 \quad (3.7)$$

$$S_{cv}(I_k) = 1 - U_{cv}(I_k) \quad (3.8)$$

其中 $\{x_c, y_c\}$ 代表了图像中心的坐标，其他符号的含义与上节类似。同样的，我们把 $U_{cv}(I_k)$ 归一化到0和1之间，那么，最终显著值由式3.8得到。

3.2.4 局部颜色对比度

为了克服全局特征存在的一些固有问题，我们加入了局部颜色对比度作为补充。我们把局部颜色对比度定义为像素点与其周围局部区域内其他所有像素点的颜色对比度。定义如下：

$$c_{i,j} = D \left[v, \frac{1}{N} \sum_{q=1}^N v_q \right] \quad (3.9)$$

其中 $c_{i,j}$ 代表位于 (i, j) 坐标的像素点的颜色， N 代表了局部区域 R 中像素点的个数， v 代表与像素点相关的颜色向量，颜色差异 D 即为我们先前定义的颜色感知距离。为了检测不同scale下的显著性区域，我们将区域 R 取两个不同大小，分别计算显著值，作为两个不同的显著图的结果。

3.2.5 全局特征计算加速

由于全局特征需要计算所有颜色在全局环境下的特征，而整个颜色空间十分巨大，对于高分辨率的图像来说，计算量更是大的不可忍受，为了加速全局特征的计算，我们采用了Cheng[13]的两个方法，即色彩向量量化与色彩空间平滑。

色彩向量量化： 在我们的实现中，首先将每个彩色通道量化为16个不同的值，这样就将整个颜色空间的色彩数量减少到 $16^3 = 4096$ 。为了进一步减少颜色数量，我们选取占整个图像像素数量95%的颜色，其他颜色则采用这些高频出现的颜色替代。最终，颜色的数量能够较少到100种左右。

色彩空间平滑： 在量化之后，会产生明显的人工量化痕迹，比如两个相似的颜色可能被量化为不同的颜色值，从而对最终的显著图的准确性产生一定的影响。为了尽量较少这种影响，我们采用相似颜色的加权平均显著值来平滑显著图。具体如下：

$$S'(c_l) \propto \sum_{j=1}^n \exp\left(-\frac{d(c_l, c_j)^2}{2\sigma}\right) S(c_j) \quad (3.10)$$

其中， $\exp\left(-\frac{d(c_l, c_j)^2}{2\sigma}\right)$ 是对颜色 c_l 和 c_j 的距离度量。

3.3 条件随机场结合多特征方法

现在我们可以得到5副显著图（3副全局特征，2副在不同尺度下的局部特征），既然每个显著图都是基于不同的假设与特征，那么如果能将他们结合起来就很有可能能够提高检测的准确性[10]。但是简单的对他们进行加和平均或者对应像素点相乘并不是一个好主意，因为在这种无权重的情况下，表现较好的显著图会被表现差的显著图拉低性能[17]。而且由于相邻像素点之间存在相互作用，单独的去评估一个像素点的显著图显然是不妥当的。在这种考虑之下，我们采用了条件随机场(CRF)这一强大的工具帮助我们结合多个显著性特征。事实上，我们描述的CRF模型与Mai等人提出的模型[31]非常相似，不同的是，Mai的模型用来将其它学者提出的方法进行融合，而我们的模型用来直接将我们提出的特征进行融合。下面将对我们的模型进行描述。

3.3.1 条件随机场简介

条件随机场是一种概率图模型。不同于传统的代数解析式模型，概率图模型提供了模型的直观图形表示，并且为模型的训练求解提供了通用的方式，概率图模型具有以下优点[9]：

1. 概率图模型提供了一种简单有效的方式来可视化概率模型，同时更有利于设计新的模型。
2. 通过对图的观察，可以清楚的发现模型的一些特性，比如条件独立等等。
3. 对于复杂的模型，Inference和Learning都是非常困难的。而概率图可以将这些过程以图操作来表示，在某种程度上简化了模型的推测和学习。

如图3.4所示，这是一个复杂的概率模型，如果使用代数式表示，该模型的联合概率为：

$$P = p(x_1)p(x_2)p(x_3)p(x_4|x_1, x_2, x_3)p(x_5|x_1, x_3)p(x_6|x_4)p(x_7|x_4, x_5) \quad (3.11)$$

显然，倘若使用传统的代数式表达如此复杂的模型，很难从代数式中观察到模型的特性，更不用说去求解、学习这样一个模型了。而如果使用图3.4概率图来表达这个模型，则显得非常直观，节点之间的连线代表了条件独立性，而通过一些规则，我们可以轻易的写出其他条件概率表达式。

条件随机场属于无向图模型，是马尔科夫随机场在条件概率下的表达：

$$p(y|x, w) = \frac{1}{Z(x, w)} \prod_c \psi_c(y_c|x, w) \quad (3.12)$$

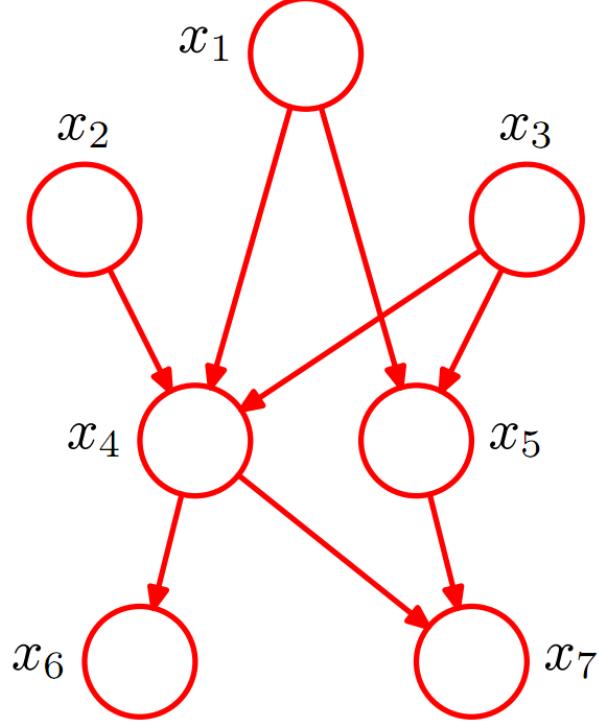


图 3.4: 图模型示例

条件随机场可以被看做是罗吉斯回归的结构化输出的扩展[32]，我们通常使用log-linear的假设来表达potentials:

$$\psi_c(y_c|x, w) = \exp(w_c^T \phi(x, y_c)) \quad (3.13)$$

其中 $\phi(x, y_c)$ 是由全局输入 x 和局部标签 y_c 产生的特征向量。条件随机场(CRF)相较于马尔科夫随机场(MRF)的优势在于它是一个判别式模型，而非生成式模型，判别式模型直接对问题进行分类，准确率通常更高。

3.3.2 建模目标

为了有效利用各个显著性特征的优势，同时又保证高质量的显著图在融合后不被低质量的显著图拉低性能，我们希望建立的模型具有以下效果：

1. 各个特征之间能够相互补充，通过类似投票的方式，决定一个像素的显著性。即若大多数特征认为该像素是显著地，那么最终结果中，该像素是显著的可能性应该更高。
2. 在一个特征中，如果两个像素点的显著值相差很大，那么他们被分配为同一标签的可能性更小。

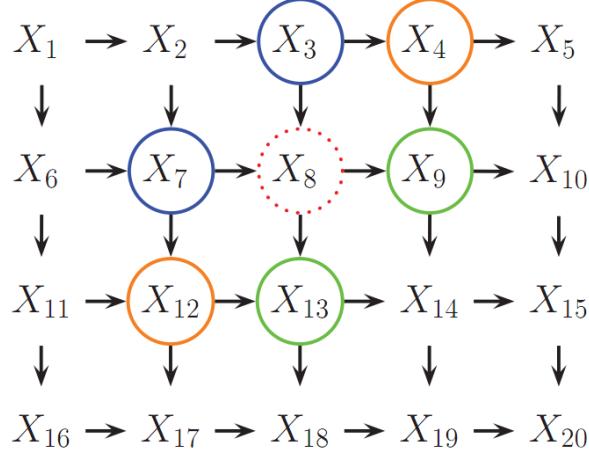


图 3.5: 二维晶格状模型

3. 如果两个相邻像素点具有相似的颜色，那么他们被分配为同一标签的可能性更大。

在我们的问题中，由于需要对图像建模，我们采用了如图3.5所示的二维晶格状模型，将每个像素点作为模型中的一个未知量（节点），同时相邻像素节点之间存在一条边(edge)，表明相邻像素点之间的条件独立关系。

3.3.3 我们的模型

在我们的模型中，给定一幅图像 I ，我们使用一个二值掩图(mask) $Y = \{y_p | p \in I\}$ 标记出显著物体。图模型以每个像素点为顶点，8领域之间的像素点之间相连，形成一个二维晶格状的概率图。那么，在这个概率图下， Y 在给定图像 I 下的条件概率可以写为，

$$P(Y|I) = \frac{1}{Z} \exp\left(\sum_{p \in I} F_d(y_p, I) + \sum_{p \in I} \sum_{q \in N_p} F_s(y_p, y_q, I)\right) \quad (3.14)$$

其中 p 代表 I 中的一个像素点， y_p 是其显著与否的标记。 $F_d(y_p, I)$ 是“节点特征函数”， $F_s(y_p, y_q, I)$ 是“邻接边特征函数”，描述了相邻像素点之间的联系。

节点特征函数仅仅与输入的已有显著图 S_i 有关，即

$$F_d(y_p, I) = \sum_{i=1}^m \lambda_i S_i(p) + \lambda_{m+1} y_p \quad (3.15)$$

其中 λ_i 是条件随机场的一组待学习的参数。而邻接边特征函数描述了相邻像素之间的关系：

$$F_s(y_p, y_q, I) = F_e(y_p, y_q, I) + F_c(y_p, y_q, I) \quad (3.16)$$

其中 $F_e(y_p, y_q, I)$ 考虑到了这样一个事实，如果两个像素点在同一幅显著图中的显著值差别很大，那么他们最后倾向于拥有不同的显著性标记。

$$F_e(y_p, y_q, I) = \sum_{i=1}^m \alpha_i (\mathbf{1}(y_p = 1, y_q = 0) - \mathbf{1}(y_p = 0, y_q = 1)) (S_i(p) - S_i(q)) \quad (3.17)$$

其中 α_i 为CRF的待学习参数， $\mathbf{1}(\cdot)$ 为指示函数（括号内真为1，假为0）。 F_s 的第二项可以看做一个惩罚项，当像素点的颜色相似却被标上不同的标记时要进行惩罚：

$$F_c(y_p, y_q, I) = \mathbf{1}(y_p \neq y_q) \exp(-\beta \|I(p) - I(q)\|) \quad (3.18)$$

其中 $\|I(p) - I(q)\|$ 代表像素 p 和 q 的颜色差（Lab颜色空间）， β 被设置为 $(2 < \|I(p) - I(q)\|^2>)^{-1}$ ， $<\cdot>$ 代表计算期望。

最后，通过训练这个模型，得到所有参数的最优值。当计算新的显著图时，我们取每个顶点（像素点）被标记为1的概率作为该像素点的显著值。CRF的训练和inference，我们采用了Mark Schmidt的开源工具包UGM[1]实现。

3.4 实验结果

3.4.1 实验方法

为了评估我们模型的性能，我们采用了Achanta等人创建的公开数据集ASD[4]。这个数据集包含1000幅图片，每一幅图像都对应有一个人工标注的显著图（二值图，用于标示显著区域）作为Ground Truth。在这个数据集上，我们评估了10种国际上的经典方法作为比较，包括IT[22], HC[13], RC[13], SR[21], AC[3], FT[4], GB[19], IG[4], MZ[30] and LC[48]。对于AC,GB,IG,IT,MZ,SR，我直接使用从[4]上随标准数据集下载的显著图，对于其他方法，我们使用作者提供的公开实现代码。

3.4.2 评价标准

显著区域检测的真正用途在于其应用。我们采用[4]中的思路，将显著图用于目标物体分割来进行评价。我们首先将显著图采用某个阈值进行二值化，则值为1的像素点为前景物体，值为0的点为背景物体。

试验中，我们采用了固定阈值来二值化显著图，为了评价显著图的质量，阈值从0到255变化，然后分别计算对应的准确率和召回率，最后绘制出准确率-召回率曲线（PR-curve）。

另外，为了进一步评价算法的性能，我们也采用了F-score的评价方式，F-score定义如下：

$$F_\beta = \frac{(1 + \beta^2) * precision * recall}{\beta^2 * precision + recall} \quad (3.19)$$

β^2 用于我们调整对准确率和召回率的重视程度，在我们的试验中被设置为0.3。

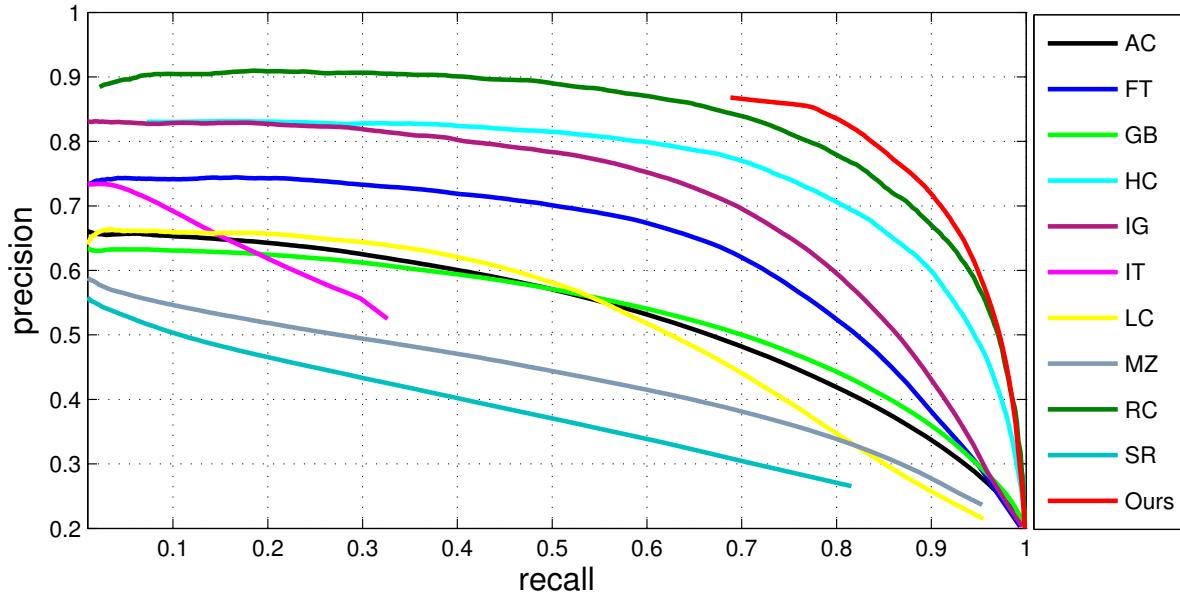


图 3.6: 与其他10种方法的对比: 准确率-召回率曲线

3.4.3 实验结果

PR曲线如图3.6所示, 值得注意的是, 我们的方法在任何阈值下都可以达到很高的召回率(≥ 0.7), 因此我们的方法是没有召回率低于0.7这部分的曲线的。如果仅仅看召回率大于0.7的部分, 我们的方法很显然优于其它方法, 拥有更高的准确率。

我们将F-score在各个阈值下的结果绘制成曲线, 如图3.7所示。可以看到, 我们的方法在各个阈值下基本都保持最高的F-score, 值得注意的是, 我们的方法的F-score在很大的阈值范围内基本保持不变, 这说明了我们的方法对阈值不敏感, 具有一定的鲁棒性。

我们还提供了一些图片结果的主观对比, 如图3.8所示。

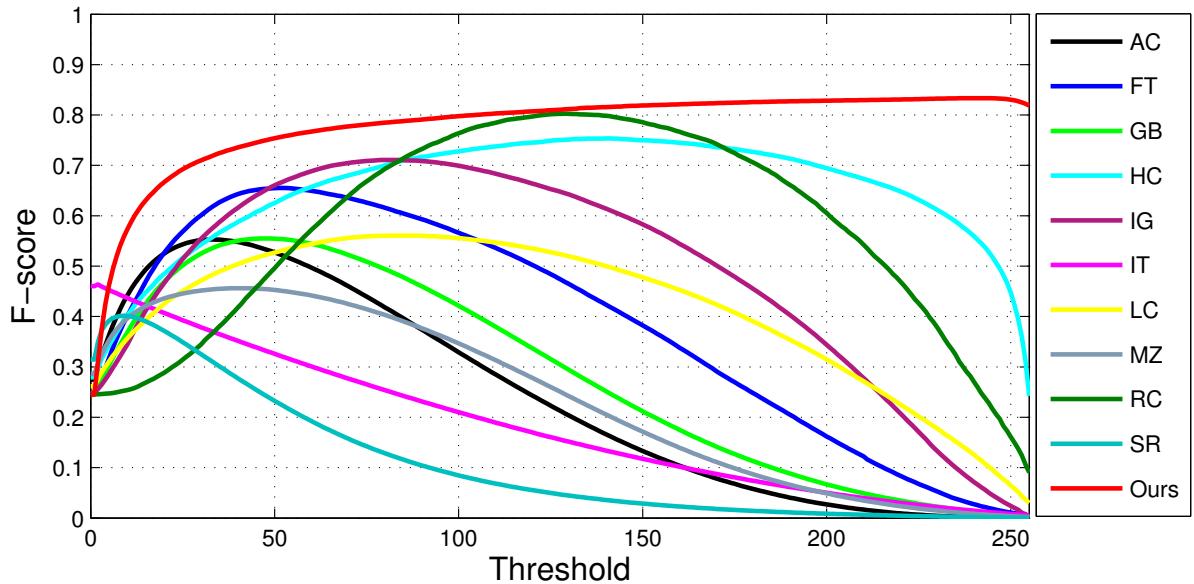


图 3.7: 与其他10种方法的对比: F-score曲线

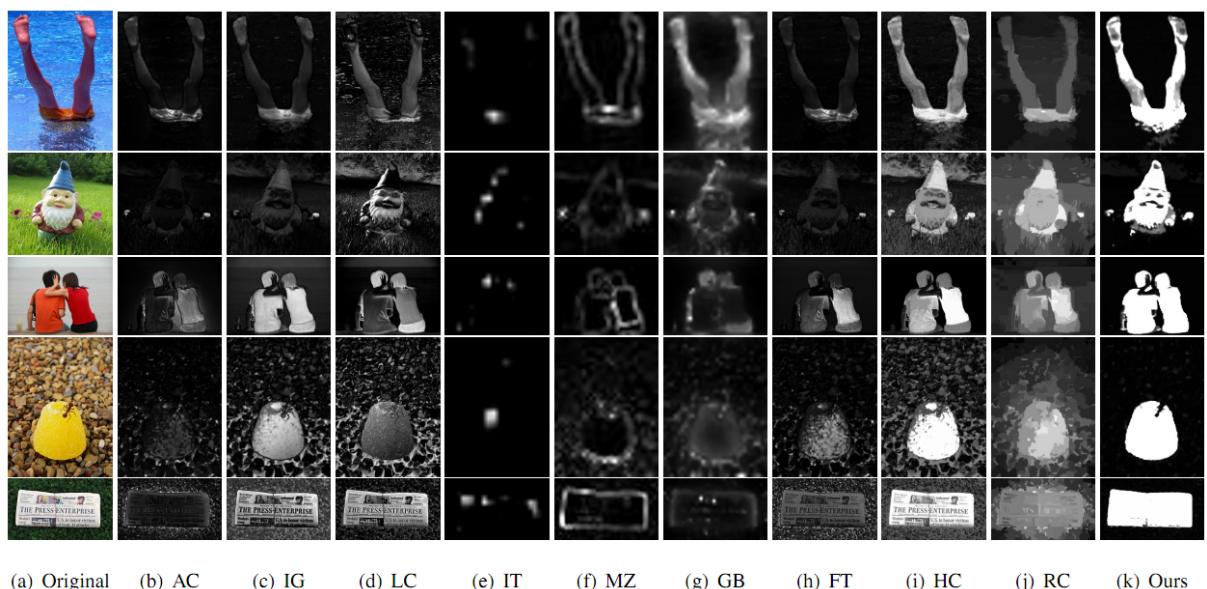


图 3.8: 与其他10种方法的主观视觉对比

第四章 基于蒙特卡罗采样的显著性区域检测

4.1 引言

4.1.1 研究背景

对视觉显著性的研究最早可以追溯到Koch和Ullman的基于生物视觉原理的模型[27]，之后Itti等人融入了多尺度图像特征进行了改进[22]。自此之后，显著性检测就吸引了来自各个领域的众多学者。

在最初的阶段，人们试图根据局部对比度挖掘图像区块的稀有性，从而去定义显著值。在Ma和Zhang的工作中[30]，首次结合了fuzzy growing与局部对比度分析用于显著区域检测。在Harel等人的工作中[19]，根据相邻像素点的相似性首先构建了一个邻接图，然后利用马尔科夫过程来挖掘像素的显著性。其他学者，比如Liu等人[28]，以及Mai等人[31]，都将局部对比度作为一个很重要的特征。

然而，基于局部对比度的方法倾向于给予边缘更高的显著值，而并非高亮整个显著区域。故近年来，越来越多的学者开始使用全局对比度。Zhai等人[48]提出了第一个基于全局对比度的显著区域检测模型，然而，考虑到直接在三通道色彩空间上计算，时间复杂度太高，他们只采用了亮度信息进行计算。Cheng等人[13]则改进了这个算法，他们应用了三通道进行计算，为了减少时间复杂度，同时引入了量化和色彩空间平滑两个工具。

随着机器学习方法的火热，机器学习也被许多学者尝试应用于该领域。Kienzel等人[48]基于眼动数据学习了一个kernel SVM，用于区分一个图像块是否显著。Ye等人[47]则采用了条件随机场，将局部和全局特征相结合，以弥补各个特征的缺陷。

4.1.2 研究动机

当前国际上的主流方法均有各自的缺陷：基于局部对比度的方法倾向于高亮边缘，而基于全局对比度的方法无法区分前景和背景中相同颜色的像素，基于学习的方法则与学习数据关系很大，在不同数据集上表现差异较大。同时，现有的方法时间复杂度都较高，实际利用价值大打折扣。

而这些基于底层特征而计算得到的显著图在许多方面都有非常明显的、可以改进的方面。如图4.1所示，基于局部对比度的方法，如MZ[30]，倾向于高亮物体的边缘，而基于全局对比度的方法如HC[13]，则会将草地也一起高亮（因为草地与前景颜色相近，相对于天空同样具有较大的对比度）。而实际上，这些方法都可以在某种程度上进行修正。首先，如果一个区域被显著的区域所包围，那么这个区域也通常是显著的，

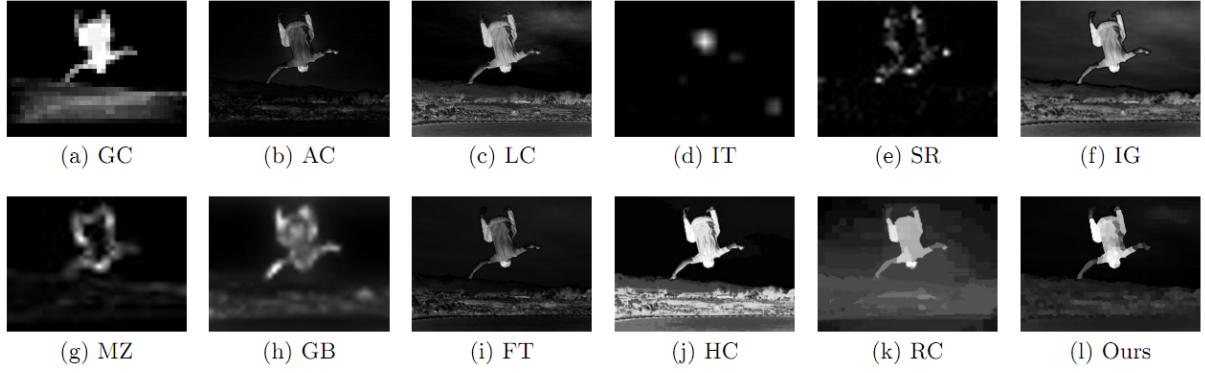


图 4.1: 与其余10种方法的对比示例

这可以解决基于局部对比度的方法的缺陷；其次，如果一个区域与图像的边框相连接，那么这个区域通常不是显著的，这可以一定程度上解决基于全局对比度的方法的缺陷。所以，如果能加入这种空间上的约束，那么检测的准确率和召回率可以大大得以提高。

然而，这三种空间特征都是二值特征，很难直接将其应用到显著图中，为此我们考虑使用蒙特卡洛采样，将显著值的度量分解为多次采样-标记的过程。

4.1.3 解决方法概要

我们为了将三种空间特征利用起来，使用了蒙特卡洛采样，将显著值的度量分解为多次采样，标记的过程，主要步骤如下：

1. 采样。对整幅图像中的像素点，以一定概率分布进行采样。
2. 对采样进行处理。对采样得到的像素点，计算整幅图像的Distance Map，继而利用紧致性和连通性计算Binary Map，最后利用包围性计算Final Binary Map。
3. 对多次采样得到的Binary Map进行加权平均，得到最终的显著图。

整个系统的框架如图4.2所示。

4.2 采样方法

显著性区域检测和图像分割实际上非常相似，区别在于，前者仅仅将图像分割为两个部分：前景和背景。因此，在显著性区域检测中，图像的像素被自然的划分为前景和背景两类，显然，属于不同类别的像素点应该具有较大的差异，而属于同一类的像素点差异较小。根据我们的假设，为了高亮前景物体，我们应该在采样的时候尽量采样背景像素，这样得到的distance map中，前景物体则会被高亮，背景物体则会呈现灰暗。

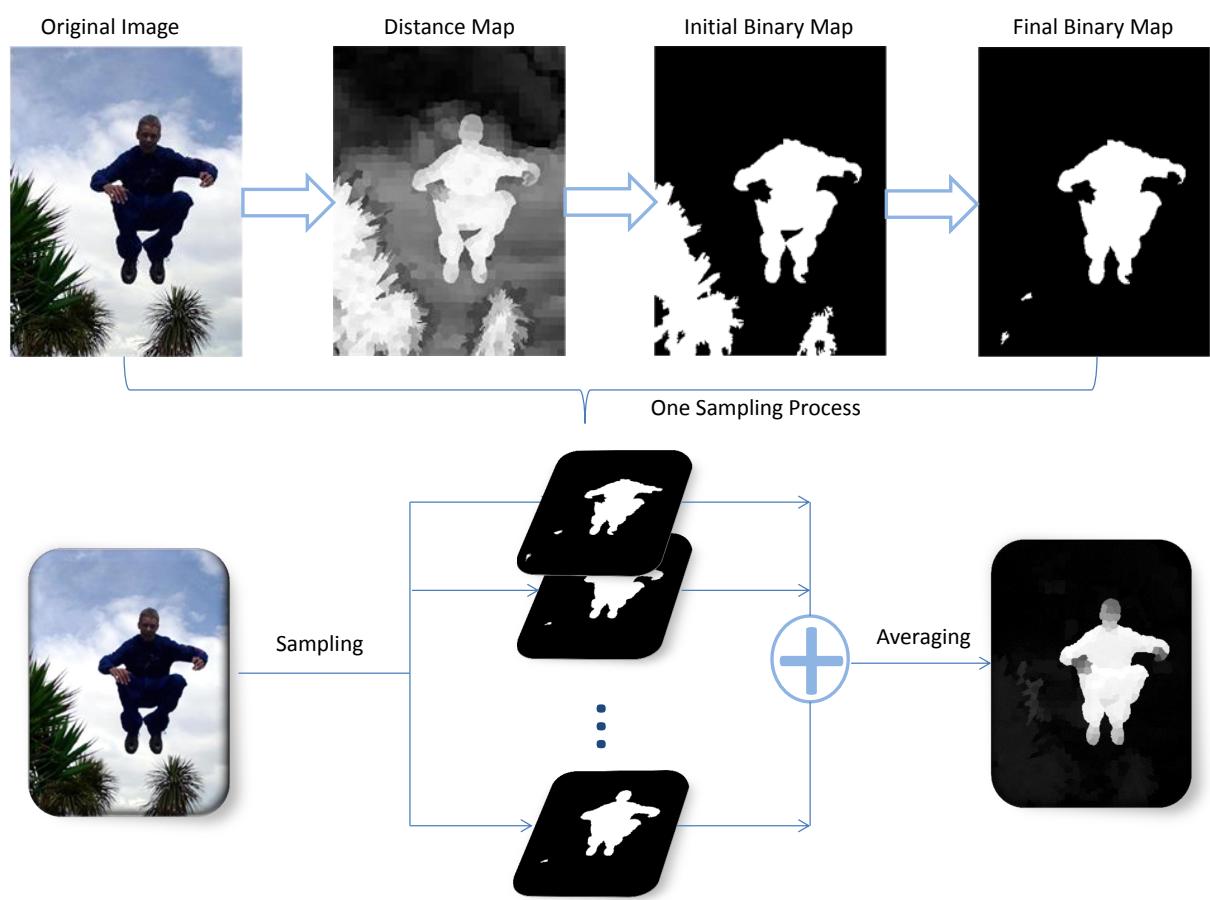


图 4.2: 系统框架图

根据心理学家的分析[42]，人类的注意力倾向于图像中心区域，同时摄影师在摄影时，也倾向于将主体部分放置在靠近中心的位置。因此，靠近图像中心的像素点，有较大概率为前景，而靠近图像边缘的像素点，则有较大概率为背景像素。因此，为了以较大概率采样得到背景像素点，我们将采样的概率分布设定如下：

$$p(I_i) = \frac{1}{Z} (1 - \exp\{-\lambda(x_i - x_c)^2\}) \quad (4.1)$$

这里， I_i 代表图像中的第*i*个像素点， $p(I_i)$ 表示该像素点被采样到的概率， x_i 是该点的坐标， x_c 是图像中心的坐标， Z 是归一化因子。

4.3 生成显著图

4.3.1 生成距离图

一旦我们采样得到一个像素点，我们就可以通过以下公式计算整幅图中其他像素点与这个像素的差异：

$$D(I_i) = \|f_i - f_s\| \quad (4.2)$$

这里 f_i 是第*i*个像素点的特征向量， f_s 是采样点的特征向量。 $\|\cdot\|$ 为欧氏距离算子。在我们的工作中，我们选取Lab颜色空间的颜色向量作为特征向量。然后我们通过归一化，将其归一化到0和1之间，得到distance map。

由于得到的距离中，存在一些异常高的值，因而使用通常的MIN-MAX归一化会使某些单一的像素高亮，而整幅图像比较暗淡。因此，我们采用下面的方法进行归一化：

$$D = \begin{cases} 1 & d \geq a \\ (d - u)/(a - u) & u < d < a \\ 0 & d \leq u \end{cases} \quad (4.3)$$

这里d是由公式4.2计算得到的距离， a 和 u 都是根据自适应的图像参数。 a 被设置为使至少有5%的像素值大于 a ， u 被设置为至少有5%的像素值小于 u 。

4.3.2 生成初始二值标记图

我们通过阈值化distance map获得二值标记图，如下所示：

$$B(I) = \text{THRESH}(D(I), \theta) \quad (4.4)$$

然而，如何选取一个合适的阈值是非常困难的事情。正如之前所说的，我们的方法结合了三种空间特征，这里，我们使用紧致性和连通性来确定一个合适的阈值。

根据我们的观察，一个二值图标记的显著区域通常比较集中，且形成一个连通的区域。我们将二值图在空间上分布的方差作为考量紧致性的指标：

$$BinMap_{Var} = Var_X + Var_Y \quad (4.5)$$

其中

$$Var_X = \frac{1}{N} \sum_{i=1}^N |x_i - x_c| \quad (4.6)$$

$$Var_Y = \frac{1}{N} \sum_{i=1}^N |y_i - y_c| \quad (4.7)$$

N 是被标记为显著的像素数量， (x_i, y_i) 则是第*i*个像素的坐标， (x_c, y_c) 是所有被标记为显著的像素的中心。

对于连通性，我们考虑显著像素的8邻域内显著像素的数量，数量越多，说明连通性越好：

$$BinMap_{Con} = \frac{1}{N} \sum_{p \in I_{Sal}} \sum_{(x,y) \in N_p} 1(p_{xy}) \quad (4.8)$$

其中， I_{Sal} 为二值图中显著像素的集合， N 是该集合的大小， N_p 是像素*p*的8邻域的坐标， $1(p_{xy})$ 则是判定函数，用于判定在 (x, y) 的像素点 p_{xy} 是否为显著像素。

显然，高连通性，空间分布越集中的二值图更好，因此，我们将阈值通过以下公式确定：

$$Criterion = \frac{BinMap_{Con}}{BinMap_{Var}} \quad (4.9)$$

在我们的实现中，我们选取了均匀选取了9个不同的阈值，并获得了9副不同的显著图，然后，我们通过上面的式子，*Criterion*最高的显著图将被选取为初始的二值标记图。

4.3.3 生成最终的二值标记图

通过包络性，我们可以进一步优化二值标记图。根据Gestalt原理[8]，一个有着封闭轮廓的区域更加易于被人眼理解为一个整体的实物，因而抓住人眼的注意力。

我们将包络性定义为二值图中有着封闭轮廓的区域，在这个定义下，任何与图像边缘连通的区域将被标记为0（背景区域），其余的区域则被标记为1（前景区域）。我们可以通过泛洪法填充，在 $O(n)$ 的时间内标记出所有的像素点。

在图4.3中，上面一行为原图像，中间一行为初始二值标记图，下面一行为通过包络性优化过的最终的二值标记图，可以看到，通过这一简单的优化，可以大大提高二值标记图的质量。

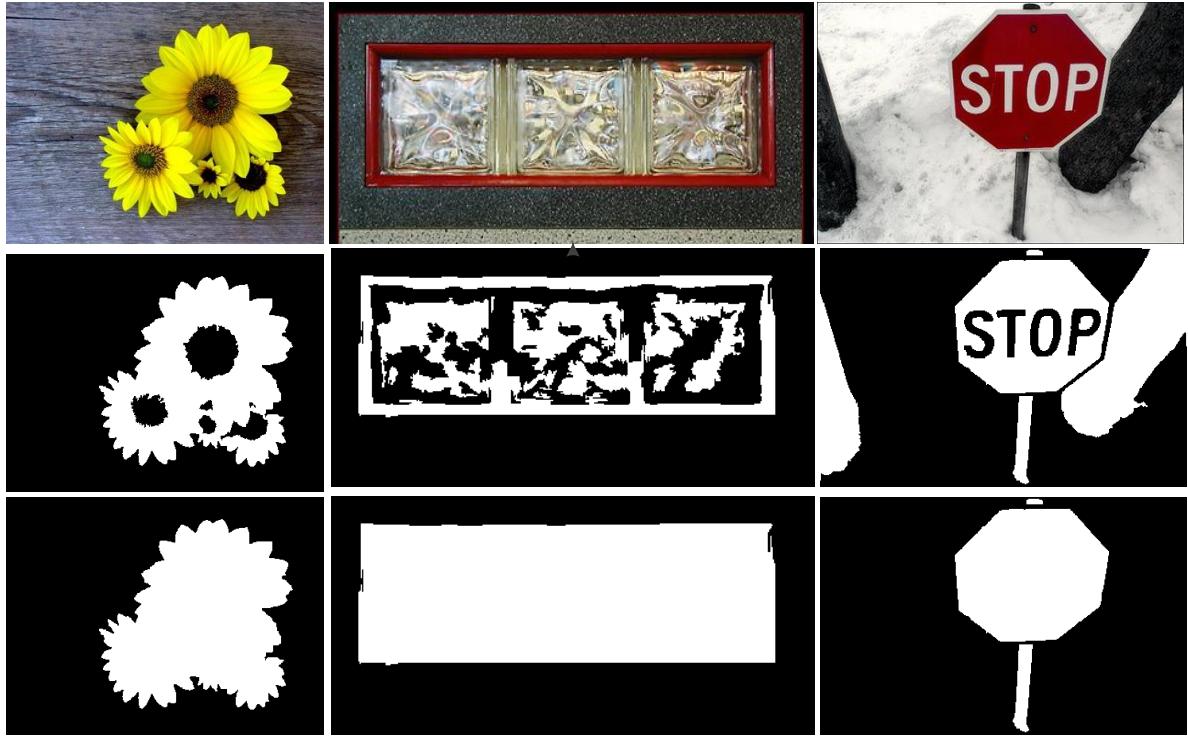


图 4.3: 包络性的作用

4.3.4 生成显著图

在每一次采样后，我们都会得到一幅二值标记图，其中1标记为这次采样中，认为该像素为显著区域。根据蒙特卡洛采样原理，一个像素为显著值（即它被标记为显著的概率）最终可以用整个采样过程中，其被标记为1的频率替代（当采样次数足够多时）。因此，我们最后简单的将所有得到的二值图进行加和平均，即：

$$Sal(I) = \frac{1}{T} \sum_{i=1}^N BinMap_i(I) \quad (4.10)$$

这里 T 是采样的次数， $BinMap_i(I)$ 是在第*i*次采样过程中得到的二值图。在我们的实现中，我们设置 $T = 400$ 。

4.4 效率问题

倘若直接在原图像上去进行上面的计算，由于像素数量之多，会导致时间复杂度将十分巨大。为了加速上述采样过程，我们使用了一下两个工具用于加速：

1. SLIC超像素分割算法[5]。我们将整幅图像分割为超像素，之后所有的计算都以超像素为单元进行计算，这样可以使得像素数量下降许多。

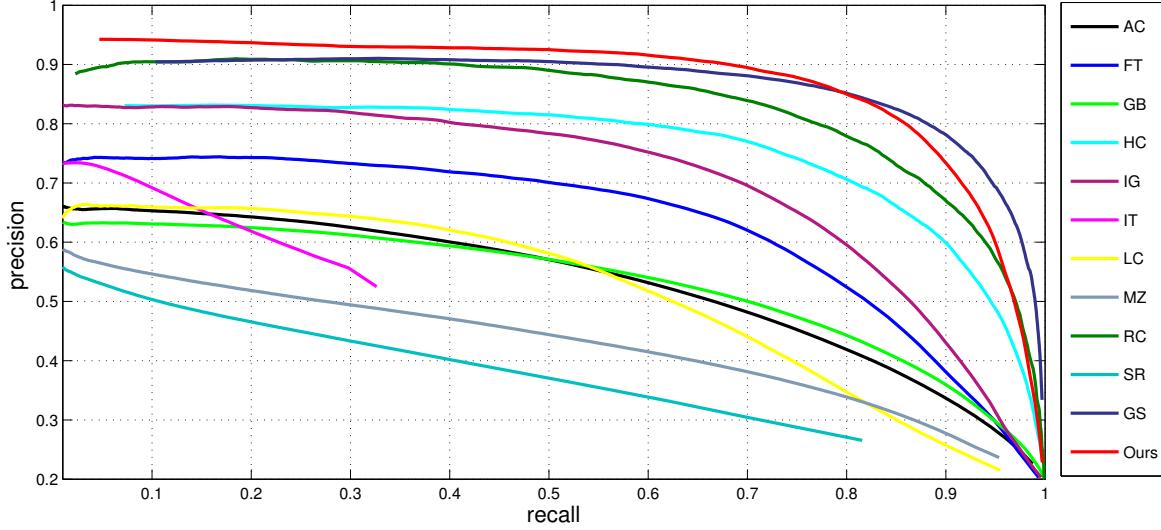


图 4.4: ASD数据集上与其他10种方法的对比：准确率-召回率曲线

2. 并行采样。由于每一次采样过程都是相互独立的，因此可以非常方便的将所有采样过程并行计算。由于我们的测试环境为4核cpu，我们将采样的线程数设置为4。

4.5 实验结果

我们分别在ASD[4]和ECSSD[46]这两个公开数据集上进行了我们的实验。同时与11中国际经典算法进行了对比，包括IT98[22], MZ03[30], LC06[48], GB06[19], SR07[21], AC08[3], FT09[4], IG09[4], HC11[13], RC11[13], and GS12[45]。

4.5.1 在ASD数据集上的结果

可以看到，我们的方法在ASD数据集上有很好的表现，与GS方法有非常接近的性能。尽管GS方法在召回率大于0.8的时候有更高的准确率，我们的方法在召回率低于0.8时，拥有更高的性能。同时，从F-score中可以看到，我们的方法在很大的阈值范围内均有非常良好的表现，这说明了我们的方法对阈值的选取不敏感，有一定的鲁棒性。

4.5.2 在ECSSD数据集上的结果

ECSSD是一个更加有挑战性的数据集，包含了1000幅背景结构复杂的图片。由于找不到AC,IG,GS方法在这个数据集上的结果，作者也没有公开源码，因此我们仅仅比较剩余的方法。

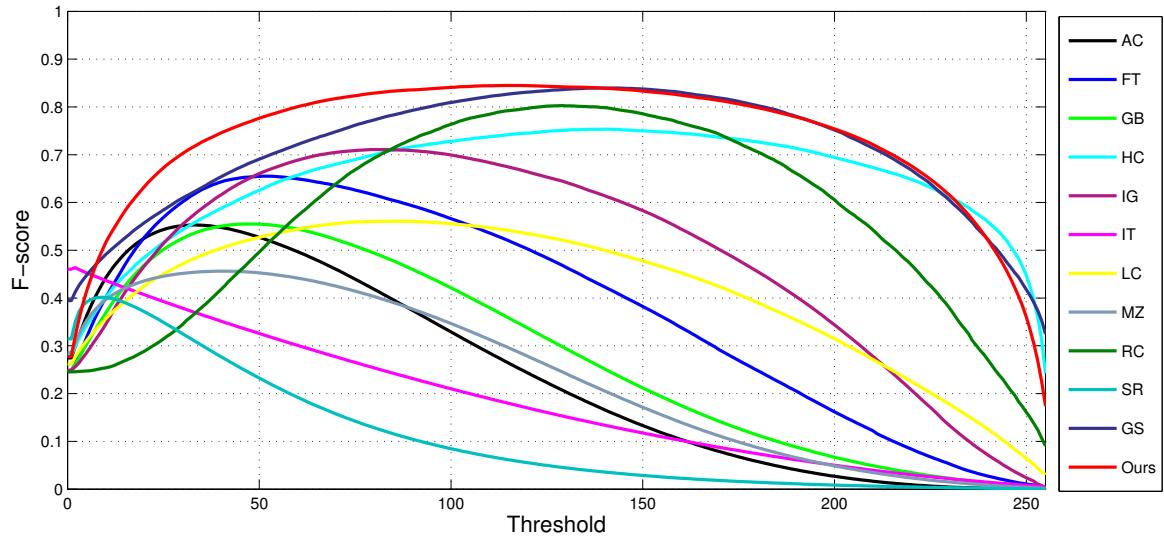


图 4.5: ASD数据集上与其他10种方法的对比: F-score曲线

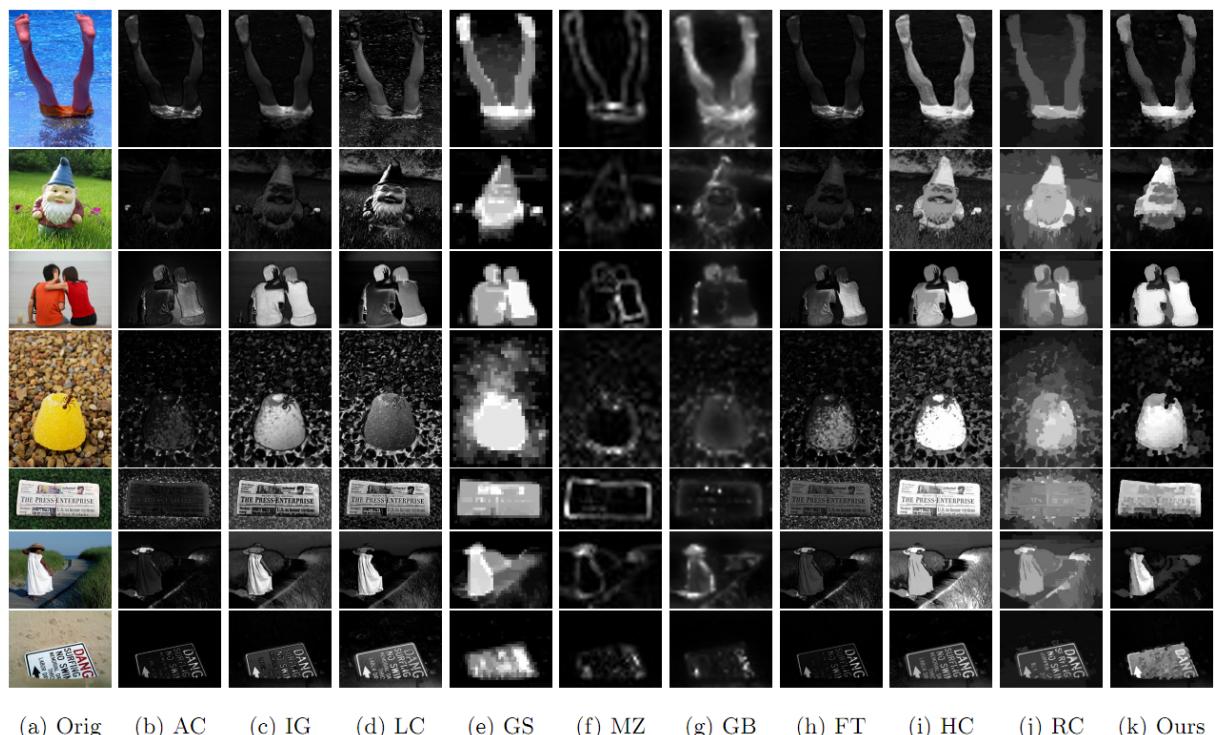


图 4.6: ASD数据集上与其他10种方法的主观视觉对比

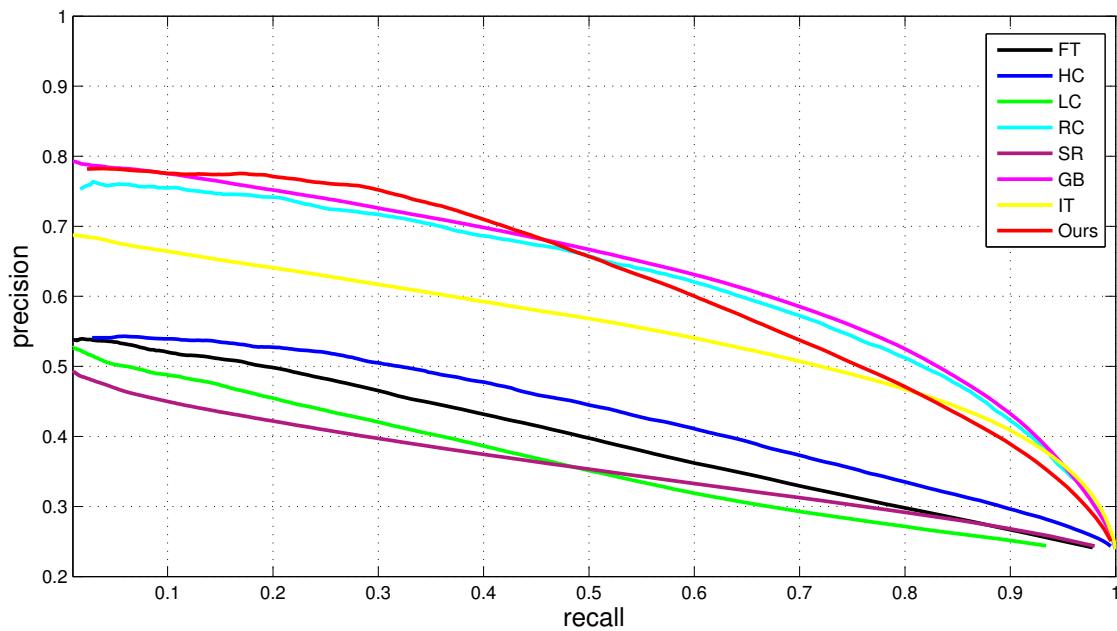


图 4.7: ECSSD数据集上与其他10种方法的对比: 准确率-召回率曲线

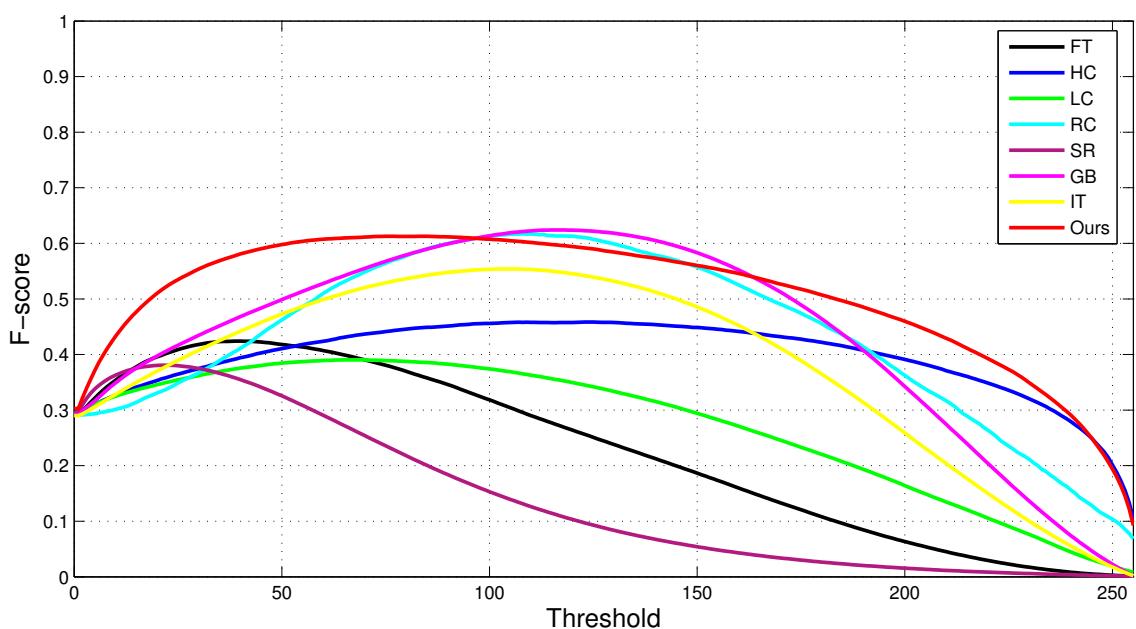


图 4.8: ECSSD数据集上与其他10种方法的对比: F-score曲线

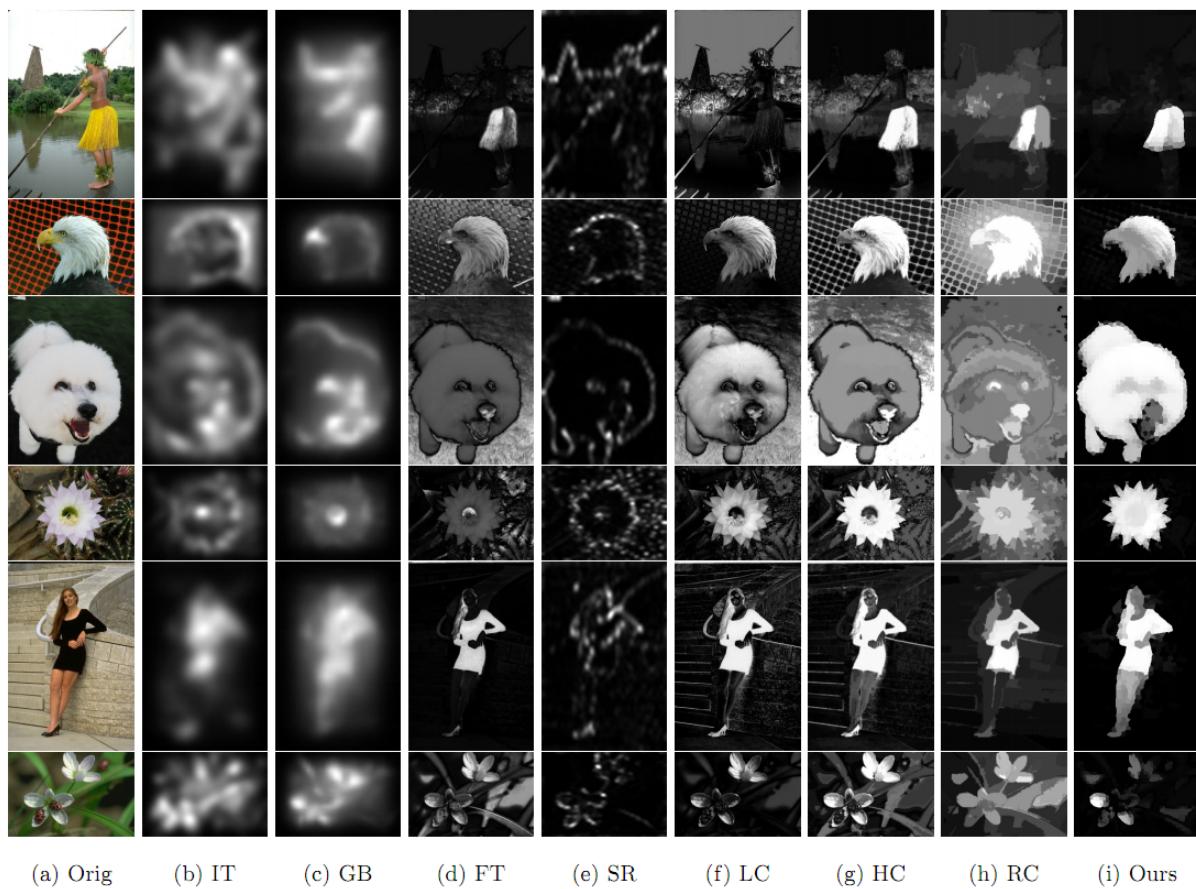


图 4.9: ECSSD数据集上与其他10种方法的主观视觉对比

从图4.7和图4.8可以看到，我们的方法在这个数据集上依然保持竞争力。显然，由于数据集的背景结构更加，所有的方法在这个数据集上的表现都有所下降。在图4.9中，我们列出了两个我们的方法表现较差的图像（第一行和最后一行）。在这样的图像中，前景图像要么对比度很低（比如第一行中人的身体），要么空间先验不起作用（比如最后一行中，花朵与图像的边缘连接在了一起）。在这类图像中，底层特征已经失去了判别力，无法体现人的视觉显著性。

第五章 显著区域检测在图像检索上的应用

5.1 引言

5.1.1 基于内容的图像检索

基于内容的图像检索 (Content-based image retrieval) 是在给定查询图像的前提下，依据内容信息或指定查询标准，在图像数据库中搜索并找出符合查询条件的相应图片[49]。

随着网络以及多媒体技术的迅速发展，数字图像的数量正在不断迅速的增长，对数字图像的自动化管理与检索，也成为新时代迫切的需求。然而，传统的基于关键字检索的方式，需要人工对图像内容进行标注，不仅工作量巨大，同时也存在人工标注的文字歧义等问题。在这样的背景下，基于内容的图像检索技术应运而生。

最早的基于内容的图像检索应用成功的是IBM开发的QBIC[2](Query By Image Content)系统，通过用户指定按例子查询或者按绘制草图查询，它利用了颜色、纹理、形状等特征对图像进行分析，从而查找出符合用户意图的图像。

由于基于内容的图像检索技术从图像内容本身出发，无需人工干预或者主动标记，大大减轻了多媒体管理人员的负担，因而被广泛的应用于电子图书馆、医学分析、博物馆等领域；与此同时，基于内容的图像检索也常常用于网购、拷贝检测等大众产品之中。

5.1.2 研究动机与方法

基于内容的图像检索通常对整幅图像进行分析，提取特征，然后在数据库中查找相似图像。然而，用户的查询意图常常并非充斥于整幅图像之中。对整幅图像进行特征提取的同时，也会将非用户意图部分（如图像背景）计算在内，从而影响最终的查询结果与精度。

显著区域检测则恰好可以解决这个问题，通过显著区域检测标记出用户感兴趣的区域，仅仅在用户感兴趣的区域提取特征并检索，能减少用户意图鸿沟，提高检索效率与效果。

在本章的接下来的部分，我们将首先介绍经典的BOW模型，并根据此模型实现一个图像检索系统。接下来，我们将显著区域检测与该系统相结合，最后，我们通过实验验证效果，以原系统为baseline，对比现系统添加显著区域检测后的提升。



图 5.1: BOW模型

5.2 基于词袋模型的图像检索系统

在大规模图像检索系统中，词袋模型(BOF)是目前为止最为成功的模型之一[6]，相比其它模型(如VLAD[7], FisherVector[36])，BOF具有易于实现，扩展众多的优点，非常利于工程化实现调优。

词袋模型在图像上提取局部特征点，并通过kmeans等方法将特征向量聚类，形成视觉单词的码书。在线检索时，通过将提取到的局部特征点映射为视觉单词，则可以将整个图像看做是由视觉单词组成的文本，从而可以应用文本检索的经典方法来对图像进行检索。下面我们将对基于BOF模型的图像检索系统进行简单的介绍。

5.2.1 SIFT特征提取

SIFT，即尺度不变性特征(Scale-invariant feature transform, SIFT)，是图像处理与计算机视觉领域广泛使用的一种局部特征描述子[29]。其具有以下一些特性：

1. SIFT特征是图像的局部特征，对旋转、尺度缩放、亮度变化等均具有不变形，对视角变化、仿射变换、噪声也具有一定程度的稳定性；
2. 独特性好，信息量丰富，适于在海量特征数据库中进行快速、准确的匹配；
3. 多量性，即使少数的几个物体也可以产生大量的SIFT特征向量；

4. 高速型，经优化的SIFT匹配算法甚至可以达到实时的要求；
5. 可扩展性，可以很方便的与其他形式的特征向量进行联合；

正是因为SIFT具有以上特性，BOF模型通常都选用SIFT对图像局部特征进行描述。SIFT特征检测主要包括以下4个基本步骤：

1. 尺度空间极值检测：搜索所有尺度上的图像位置，通过高斯微分函数来识别潜在的对于尺度和旋转具有不变性的兴趣点。
2. 关键点定位：在每个候选位置上，通过一个拟合精细的模型来确定位置和尺度，依据候选点的稳定程度选取关键点。
3. 方向确定：基于图像局部的梯度方向，分配给每个关键点位置一个或多个方向，后面对关键点的特征描述都相对于关键点的方向，尺度和位置进行变化，从而保证特征描述的不变性。
4. 关键点描述：在每个关键点周围的领域内，根据所选尺度，在图像局部计算梯度直方图，并连接为一个128维的向量。

5.2.2 建立码书

对于每个图像提取得到若干的SIFT特征点向量，然后对所有图像的特征向量进行聚类，通常使用kmeans聚类，最后将聚类得到的聚类中心作为视觉单词，形成码书。假设要形成 k 个视觉单词，kmeans算法描述如下：

1. 适当选择 k 个类的初始中心；
2. 在第 i 次迭代中，对任意一个样本，求其到 k 个中心的距离，将该样本归到距离最短的中心所在的类；
3. 利用均值等方法更新该类的中心值；
4. 对于所有的 k 个聚类中心，如果重复利用2, 3步骤进行迭代更新后，值保持不变（或者指定一个变化阈值），则迭代结束，否则继续迭代。

在实际操作中，通常在另外一个数据集上提取特征并建立码书，视觉单词数量 k 通常选取 $20k$ 至 $200k$ 之间。

	文档1	文档2	文档3	文档4	文档5
词汇1	✓			✓	
词汇2		✓	✓		
词汇3				✓	
词汇4	✓				✓
词汇5		✓			
词汇6			✓		

图 5.2: 倒排索引示例

5.2.3 倒排索引

在线实际检索时，由于通常图像库很大（例如百万级别），此时如果通过线性比较计算图像之间的相似性，速度是不可忍受的，为此需要建立某种形式的索引，以提高检索效率。通常采用倒排索引。如图5.2所示为一个单词-文档矩阵，其中打勾则代表包含关系。从纵向即文档的维度来看，每列代表文档包含了哪些单词，即正向索引。而横向即从单词的维度来看，每行代表了哪些文档包含了某个单词。倒排索引即从单词的维度来看，建立的单词-文档矩阵。如此建立索引后，每次在线查询时，只需要将得到的局部特征映射为视觉单词，然后通过倒排索引的单词-文档矩阵，就可以查找到所有包含该单词的文档，最后通过tf-idf进行评分，即可将相关的文档进行评分。倒排索引高效的关键在于，每个文档包含的视觉单词只是整个码书很小的一部分，即单词-文档矩阵是稀疏的，这样在检索时，只需要计算码书中非0部分的评分，从而避免了在整个码书上计算。

5.2.4 整体系统框架

整体框架如图5.3所示，离线训练时，我们从一组图像集合中提取局部特征，然后对其进行kmeans聚类，我们将聚类中心作为视觉单词，形成码书。在线检索时，我们首先提取图像的局部点特征(SIFT)，然后将每一个点特征进行量化（查找与码书中的视觉单词欧氏距离最近的单词），形成词频向量，最后通过倒排索引，对相关图像进行评分，排序后，输出相关文档（图像）。

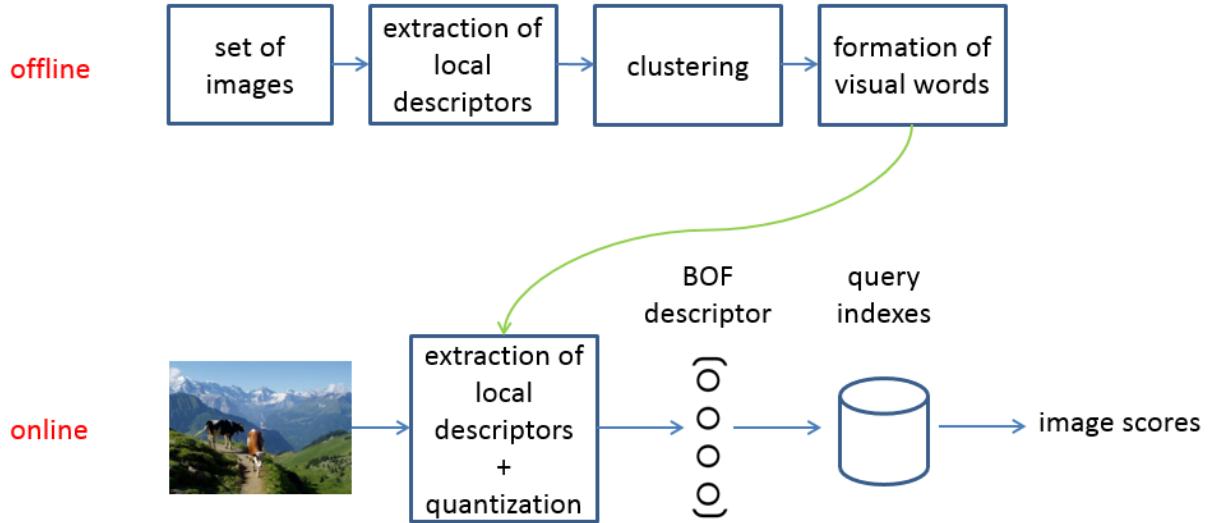


图 5.3: BOW模型框架图

5.3 利用显著区域检测优化结果

5.3.1 问题来源

如图5.4所示，第一行左边的图像为输入图像，第二行为检索结果的前三幅图像，可以看到第三幅图像是错误的。第一行右边的图像上表示了检测到的SIFT关键点，可以看到，在该图像的背景上也提取到了很多SIFT点，这些SIFT点与第三幅图像背景上的点匹配上了，因而影响了检索的精度。

可以看到，背景元素通常与用户的检索意图无关，然而背景元素上提取到的特征点，却同样参与了图像相关性的评分，因而会影响到检索的精度。

5.3.2 优化方案

我们采用显著区域检测做预处理，将前景和背景分离出来，提取特征点的时候，只在显著区域计算特征点，而忽略背景区域的特征点。这样做的好处有两个：

1. 剔除了背景上的特征点，提高检索精度，使得检索结果更加贴合用户意图。
2. 提升了检索速度。首先特征提取只需在显著区域进行计算，其次，得到更少的特征后，需要在倒排索引上检索的次数同时也减少了。

在实验中，我们选取了前一章中介绍的基于蒙特卡洛采样的显著区域检测算法。



图 5.4: 检索示例

	KS		mAP	
K=	20k	200k	20k	200k
BOF(baseline)	2.31	2.81	0.65	0.74
BOF + Salinecy	2.52	3.00	0.70	0.78

表 5.1: 实验结果

5.4 实验结果

5.4.1 数据集和评价指标

我们使用了ukbench数据集[34]，这个数据集包含2550种不同的物体或场景，每一个物体（场景）都包含4幅从不同角度拍摄的图像，整个数据集一共有10200幅图像。对于这个数据集，我们采用两种评价指标：mAp（mean Average Precision）和KS（Kentucky Score），KS指标是该数据集的作者针对该数据集提出的评价指标，即取top 4的结果中，正确结果的平均值。另外，为了更加客观的比较优化方案的优化效果，我们分别取码书大小为10k和100k的结果做比较。

5.4.2 实验结果

如表所示，可以看到，在添加了显著区域检测进行特征点的过滤后，KS值提高了0.2左右，同时mAp提高了5个百分点左右，实验证明，显著区域检测对图像检索确实有一定的提升作用。

第六章 总结与展望

6.1 研究工作总结

随着互联网和多媒体技术的普及，图像的显著性区域检测逐渐成为一个研究热点。在最初的阶段，图像的显著区域检测主要基于局部对比度的计算，通过挖掘图像块的局部稀缺性来标明显著区域。然而，这种方法很容易高亮图像的边缘而并非整个显著性区域。为了克服这些缺陷，研究者又发明了基于全局对比度的计算方法。然而，由于全局对比度的计算复杂度极高，如果在整个颜色空间上做，在现有的硬件资源条件下，计算时间将不可忍受，因而在最初的实现中，只仅仅实现了在亮度分量上的计算。随后，Cheng等人[13]发明了颜色量化与色彩空间平滑两个工具，从而将其扩展到了整个颜色空间上。除以上两种方法，还有基于频域的计算，通过将图像转化到频域空间计算其中区块的稀有性。这种方法具有良好的理论基础，然而被研究者们证明其相当于基于局部对比度的方法加上一个高斯平滑，因而与基于局部对比度的方法有一样的缺陷。另外，随着机器学习的火热，近年来也出现了很多将机器学习应用于显著区域检测的方法，但是这些方法训练复杂度高，而且其检测效果常常具有bias，即在不同数据集上的效果相差较大。

为了进一步提升显著区域检测的效果，我们从两方面展开了研究。一方面，我们探索了显著性特征，即什么样的图像被认为是显著的，在此基础上，我们提出了多个显著性特征，并应用在我们的方法中。另一方面，我们研究了如何进行特征的融合，即怎样结合这些特征，产生互补的效果，从而提升算法的性能。在这两个思路的引导下，我们开发了两个算法：基于条件随机场的显著区域检测，以及基于蒙特卡洛采样的显著区域检测。

最后，我们探索了显著区域检测的应用。我们选取了图像检索这个任务，并将显著区域检测算法应用在图像检索中。通过实验对比，我们验证了显著区域检测算法对图像检索效果的提升。

6.2 未来工作的展望

目前显著区域检测在简单背景的情况下工作良好，然而当背景复杂时，仅仅通过像素级的底层信息，还难以达到很好的检测效果。因此在未来的工作中，我们将考虑加入适度的高层次特征，比如人脸识别算子，车牌识别算子等等。另外，目前的显著区域检测算法大多基于单目标物体的先验，当图像中存在多个前景物体时，通常检测效果较差，因此后期的工作也会适度放在解决多目标的情况。

除了显著区域检测算法本身，我们还需将注意力放在应用上。脱离了实际应用的算法，将失去其实际价值，所以我们将从两个方面开展研究：其一，应用的深度问题，如何将显著区域检测与应用深度结合，而并非简单的进行预处理，从而更好的发挥显著区域检测的效果；其二，应用的广度问题，除了本文已探索的图像检索领域，其它领域是否也可以尝试加入显著区域检测以提升效果呢（譬如图像分割，目标识别等等）。

总之，显著区域检测还有很多方面等待改进，同时其应用场景的挖掘也极其具有吸引力和挑战性。未来我们将从算法改进和应用两方面着手继续进行研究工作。

参 考 文 献

- [1] <http://www.di.ens.fr/~mschmidt/software/ugm.html>.
- [2] <http://www.qbic.almaden.ibm.com/>.
- [3] Radhakrishna Achanta, Francisco Estrada, Patricia Wils, and Sabine Süsstrunk. Salient region detection and segmentation. In *Computer Vision Systems*, pages 66–75. Springer, 2008.
- [4] Radhakrishna Achanta, Sheila Hemami, Francisco Estrada, and Sabine Susstrunk. Frequency-tuned salient region detection. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 1597–1604. IEEE, 2009.
- [5] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine Süsstrunk. Slic superpixels. *École Polytechnique Fédéral de Lausanne (EPFL), Tech. Rep*, 149300, 2010.
- [6] Relja Arandjelovic and Andrew Zisserman. Three things everyone should know to improve object retrieval. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 2911–2918. IEEE, 2012.
- [7] Relja Arandjelovic and Andrew Zisserman. All about vlad. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 1578–1585. IEEE, 2013.
- [8] David C Beardslee and Michael Ed Wertheimer. Readings in perception. 1958.
- [9] Christopher M Bishop et al. *Pattern recognition and machine learning*, volume 4. springer New York, 2006.
- [10] Ali Borji, Dicky N Sihite, and Laurent Itti. Salient object detection: A benchmark. In *Computer Vision–ECCV 2012*, pages 414–429. Springer, 2012.
- [11] Tao Chen, Ming-Ming Cheng, Ping Tan, Ariel Shamir, and Shi-Min Hu. Sketch2photo: internet image montage. In *ACM Transactions on Graphics (TOG)*, volume 28, page 124. ACM, 2009.
- [12] Ming-Ming Cheng. *Saliency and Similarity Detection for Image Scene Analysis*. PhD thesis, Tsinghua University, Beijing, China, 2012.

- [13] Ming-Ming Cheng, Guo-Xin Zhang, Niloy J Mitra, Xiaolei Huang, and Shi-Min Hu. Global contrast based salient region detection. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 409–416. IEEE, 2011.
- [14] Charilaos Christopoulos, Athanassios Skodras, and Touradj Ebrahimi. The jpeg2000 still image coding system: an overview. *Consumer Electronics, IEEE Transactions on*, 46(4):1103–1127, 2000.
- [15] Joerg Deigmoeller, Takebumi Itagaki, Gerhard Stoll, and Norbert Just. A context-based approach to crop and scale video for broadcast applications. In *Broadband Multimedia Systems and Broadcasting (BMSB), 2010 IEEE International Symposium on*, pages 1–9. IEEE, 2010.
- [16] Min-Yuan Fang, Yu-Hsin Kuan, Chung-Ming Kuo, and Chaur-Heh Hsieh. Effective image retrieval techniques based on novel salient region segmentation and relevance feedback. *Multimedia Tools and Applications*, 57(3):501–525, 2012.
- [17] Viswanath Gopalakrishnan, Yiqun Hu, and Deepu Rajan. Salient region detection by modeling distributions of color and orientation. *Multimedia, IEEE Transactions on*, 11(5):892–905, 2009.
- [18] Junwei Han, King Ngi Ngan, Mingjing Li, and Hong-Jiang Zhang. Unsupervised extraction of visual attention objects in color images. *Circuits and Systems for Video Technology, IEEE Transactions on*, 16(1):141–145, 2006.
- [19] Jonathan Harel, Christof Koch, and Pietro Perona. Graph-based visual saliency. In *Advances in neural information processing systems*, pages 545–552, 2006.
- [20] Xiaodi Hou, Jonathan Harel, and Christof Koch. Image signature: Highlighting sparse salient regions. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 34(1):194–201, 2012.
- [21] Xiaodi Hou and Liqing Zhang. Saliency detection: A spectral residual approach. In *Computer Vision and Pattern Recognition, 2007. CVPR’07. IEEE Conference on*, pages 1–8. IEEE, 2007.
- [22] Laurent Itti, Christof Koch, and Ernst Niebur. A model of saliency-based visual attention for rapid scene analysis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 20(11):1254–1259, 1998.

参 考 文 献

- [23] Huaizu Jiang, Jingdong Wang, Zejian Yuan, Tie Liu, Nanning Zheng, and Shipeng Li. Automatic salient object segmentation based on context and shape prior. In *BMVC*, volume 3, page 7, 2011.
- [24] Hongchang Ke, Hui Wang, and Degang Kong. A target tracking technology based on visual salient features. In *Proceedings of the 2012 Third International Conference on Mechanic Automation and Control Engineering*, pages 1349–1352. IEEE Computer Society, 2012.
- [25] Wolf Kienzle, Felix Wichmann, Bernhard Schölkopf, and Matthias Franz. A non-parametric approach to bottom-up visual saliency. 2007.
- [26] Byoung Chul Ko and Jae-Yeal Nam. Object-of-interest image segmentation based on human attention and semantic region clustering. *JOSA A*, 23(10):2462–2470, 2006.
- [27] Christof Koch and Shimon Ullman. Shifts in selective visual attention: towards the underlying neural circuitry. In *Matters of Intelligence*, pages 115–141. Springer, 1987.
- [28] Tie Liu, Zejian Yuan, Jian Sun, Jingdong Wang, Nanning Zheng, Xiaou Tang, and Heung-Yeung Shum. Learning to detect a salient object. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(2):353–367, 2011.
- [29] David G Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004.
- [30] Yu-Fei Ma and Hong-Jiang Zhang. Contrast-based image attention analysis by using fuzzy growing. In *Proceedings of the eleventh ACM international conference on Multimedia*, pages 374–381. ACM, 2003.
- [31] Long Mai, Yuzhen Niu, and Feng Liu. Saliency aggregation: A data-driven approach. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 1131–1138, 2013.
- [32] Kevin P Murphy. *Machine learning: a probabilistic perspective*. MIT press, 2012.
- [33] Vidhya Navalpakkam and Laurent Itti. An integrated model of top-down and bottom-up attention for optimizing detection speed. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 2, pages 2049–2056. IEEE, 2006.

- [34] David Nister and Henrik Stewenius. Scalable recognition with a vocabulary tree. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 2, pages 2161–2168. IEEE, 2006.
- [35] Federico Perazzi, Philipp Krahenbuhl, Yael Pritch, and Alexander Hornung. Saliency filters: Contrast based filtering for salient region detection. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 733–740. IEEE, 2012.
- [36] Florent Perronnin, Yan Liu, Jorge Sánchez, and Hervé Poirier. Large-scale image retrieval with compressed fisher vectors. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 3384–3391. IEEE, 2010.
- [37] Anthony Santella, Maneesh Agrawala, Doug DeCarlo, David Salesin, and Michael Cohen. Gaze-based interaction for semi-automatic photo cropping. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 771–780. ACM, 2006.
- [38] Boris Schauerte and Rainer Stiefelhagen. Quaternion-based spectral saliency detection for eye fixation prediction. In *Computer Vision–ECCV 2012*, pages 116–129. Springer, 2012.
- [39] A-H Shabani, John S Zelek, and David A Clausi. Human action recognition using salient opponent-based motion features. In *Computer and Robot Vision (CRV), 2010 Canadian Conference on*, pages 362–369. IEEE, 2010.
- [40] Elya Shechtman, Daniel R Goldman, and David E Jacobs. Methods and apparatuses for generating co-salient thumbnails for digital images, May 16 2013. US Patent 20,130,120,438.
- [41] Keyang Shi, Keze Wang, Jiangbo Lu, and Liang Lin. Pisa: Pixelwise image saliency by aggregating complementary appearance contrast measures with spatial priors. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 2115–2122. IEEE, 2013.
- [42] Benjamin W Tatler. The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. *Journal of Vision*, 7(14), 2007.
- [43] Yao Hong Tsai. Hierarchical salient point selection for image retrieval. *Pattern Recognition Letters*, 33(12):1587–1593, 2012.

参 考 文 献

- [44] Paul Viola and Michael Jones. Rapid object detection using a boosted cascade of simple features. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, volume 1, pages I–511. IEEE, 2001.
- [45] Yichen Wei, Fang Wen, Wangjiang Zhu, and Jian Sun. Geodesic saliency using background priors. In *Computer Vision–ECCV 2012*, pages 29–42. Springer, 2012.
- [46] Qiong Yan, Li Xu, Jianping Shi, and Jiaya Jia. Hierarchical saliency detection. CVPR, 2013.
- [47] Tiancai Ye, Dongming Zhang, Ke Gao, Guoqing Jin, Yongdong Zhang, and Qingsheng Yuan. Salient region detection: Integrate both global and local cues. In *Multimedia and Expo (ICME), 2014 IEEE International Conference on*, pages 1–6. IEEE, 2014.
- [48] Yun Zhai and Mubarak Shah. Visual attention detection in video sequences using spatiotemporal cues. In *Proceedings of the 14th annual ACM international conference on Multimedia*, pages 815–824. ACM, 2006.
- [49] 陈媛媛. 图像显著区域提取及其在图像检索中的应用. PhD thesis, 上海交通大学, 2007.

致 谢

本论文的工作是在导师张冬明副研究员的悉心指导下完成的。从科研方向的选题，到工程项目中的设计实现，张老师都以其高超的学术造诣与丰富的实践经验，为我指明了道路，提供了许多精妙的建议。同时，张老师严谨求实的治学态度、乐观积极和平易近人的工作作风，都给我树立了学习的榜样。另外，还要非常感谢张老师在我学习阶段中所犯各种错误的宽容与谅解，不断的指点与帮助，特此向敬爱的张冬明老师表达我最诚挚的敬意与感谢！

另外还要感谢张勇东研究员为实验室提供的优质科研环境，张勇东研究员以其勤恳、朴实的工作态度深深的影响着实验室的每一位同学。感谢实验室的顾晓光，靳国庆老师，与他们的合作讨论使我受到了很多启发。感谢高科老师经常教我一些为人处世的道理，其乐观向上的态度深深感染了我。感谢张磊、周仁浩、姚涵涛等同学对我的帮助。

最后，还要对我的父母表示真挚的敬意与感激。他们一直以来都默默支持和鼓励着我，使我顺利的完成了学业，克服了成长中的各种困难与障碍。

感谢所有给予过我帮助与支持的老师同学们！

作 者 简 历

基本情况

姓名：叶天才 性别：男 出生日期：1990年05月 籍贯：湖北孝感

教育经历

2008年9月—2012年7月，华中科技大学，电子与信息工程系，学士

2012年9月—2015年7月，中国科学院，计算技术研究所，硕士

【攻读硕士学位期间发表的论文】

- [1] Tiancai Ye, Dongming Zhang, Feng Dai, and Yongdong Zhang . “Fast mode decision algorithm for intra prediction in HEVC,” in Proceedings of the Fifth International Conference on Internet Multimedia Computing and Service (ICIMCS). ACM, 2013: 300-304.
- [2] Tiancai Ye, Dongming Zhang, Ke Gao, Guoqing Jin, Yongdong Zhang, Qingsheng Yuan, “Salient Region Detection : Integrate Both Global and Local Cues,” in IEEE International Conference on Multimedia and Expo (ICME),. IEEE, 2014:105-111.
- [3] Tiancai Ye, Dongming Zhang, Guoqing Jin, Ke Gao, Xiaoguang Gu, and Yongdong Zhang, “Monte Carlo Sampling based Salient Region Detection,” in Proceedings of International Conference on Multimedia Retrieval (ICMR). ACM, 2014: 97.

【攻读硕士学位期间参加的科研项目】

- [1] 互联网视频中台标识别系统（242国家重大安全项目）
- [2] 大规模图像检索系统