

分类号	TP391
UDC	004

学校代码	10590
密 级	公开

深圳大学硕士学位论文

基于动态背景的迭代显著性检测 算法研究

学位申请人姓名	文 方 俊
---------	-------

专 业 名 称	计算机科学与技术
---------	----------

学院（系、所）	计算机与软件学院
---------	----------

指导教师姓名	雷海军 副教授
--------	---------

原创性声明

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

论文作者签名：

日期： 年 月 日

学位论文版权使用授权书

本学位论文作者完全了解深圳大学关于收集、保存、使用学位论文的规定，即：研究生在校攻读学位期间论文工作的知识产权单位属深圳大学。学校有权保留学位论文并向国家主管部门或其他机构送交论文的电子版和纸质版，允许论文被查阅和借阅。本人授权深圳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，可以采用影印、缩印或扫描等复制手段保存、汇编学位论文。

（涉密学位论文在解密后适用本授权书）

论文作者签名：

导师签名：

日期： 年 月 日

日期： 年 月 日

摘 要

我们生活在一个智能化设备大爆炸的时代,而图像作为我们人类最直观的获取信息的方式,随着数码相机和智能手机等智能图像输出设备的普及,如何让计算机能快速的从海量的图像数据中提取我们人类最感兴趣的信息是当前计算机视觉研究领域的一个亟需解决的问题。图像的显著性检测研究受到人类视觉的选择性注意机制的启发,其目的是从图像中提取人眼最感兴趣的目标图像区域,并输出一张被称为显著图的灰度图像。这样便可以把有限的计算资源分配在最显著的图像区域,进而进行后续的图像处理操作。

从动态背景和迭代显著差异的角度,本文提出了一种新颖的自底向上的图像显著性检测算法。不同于以往的任何显著检测算法,根据显著对象分割二分类的特性,从提高显著分类差异的角度来进行差异迭代,分离出了显著对象区域和背景区域。与人为的划定前景和背景的显著性检测算法以及假定显著目标一般不会出现于图像边缘的显著检测算法不同,本文提出的算法对于输入图像没有任何的限制。将图像用一条直线划分为两个动态的区域,然后选取其中一个区域作为动态背景区域,并用动态背景区域的图像特征来度量另外一个待分类的图像区域中的每一个图像块。当这条直线划过整个图像区域的时候,便对整个图像中所有的图像块都至少进行了一次显著分类。如果对每次的显著分类结果进行迭代的话,就得到了一个显著分类差异的累积,这样图像块之间的显著差异就会被放大,从而实现了显著对象区域和背景区域的自动分离。为了保证所有图像块进行次数相对均等的显著分类,本文采取了分别从上、下、左、右四个方向进行动态扫描的办法。

为了公正的评价本文算法的有效性,本文算法在两个大家公认的显著目标检测数据集MSRA-1000和MSRA-B上进行了评测,并与其他7种经典的显著性检测算法分别进行了定性的和定量的比较。算法的检测结果表现出了比现有经典算法具备更高的Precision、Recall和F-measure值。

关键字: 显著性检测; 自底向上; 动态背景; 迭代; 图像分割

Abstract

We live in a world of the explosion of artificial intelligence machines. Image, as the basis of us to acquire information, plays a vital role in our daily life. With smart digital cameras and smart cellphones widely used, it's a vital task to let computers have the ability to acquire information that we humans interested rapidly from the big amount of image data. Image saliency detection, which is used to detect the region of interest (ROI) of an image, is inspired from the selective attention mechanism of human's visual system. The output of saliency detection of an image is a gray image which we called a saliency map of the input image. Therefore, we can allocate the finite computing resources to do the detail understanding of the image content.

We develop a brightly new algorithm to detect the bottom-up saliency of an image based on the prospective of dynamic background area and iterating saliency difference. Not similar to any other salient object detection algorithms, we iterate the classification difference from the point of view of improving saliency classification difference which is based on the two-classification problem of salient object segmentation. Then we get the result of a gray image with one area as the salient object area and the other areas as the background area. We have no assumptions of the input image while some other algorithms all assume that image patches on the edge are always the background or salient object will not appear on the edge of the image. We take the input image as two dynamic areas with a moving line to separate them. We choose one area as the dynamic background area and then use its image features to classify each image patch inside another part of the image. We will do at least one saliency classification of all image patches when the moving line across the whole image area. If we iterate the results of every classification, we exaggerate the salient difference between image patches and finally we separate the salient object area and the background area automatically. To ensure all image patches do almost the same times of saliency classification, we choose the moving line to move dynamically through four directions which include top, down,

left and right.

To demonstrate the effectiveness of our algorithm, firstly, we evaluate its performance on two widely accepted image dataset MSRA-1000 and MSRA-B. And then, we do quantitative and qualitative evaluation with other seven classic saliency detection algorithms. The experimental results show that the saliency maps of our algorithm achieve a higher Precision\Recall rate and F-measure than the classic ones.

Key words: Saliency Detection; Bottom Up; Dynamic Background; Iterative; Image Segmentation

目 录

摘 要	I
Abstract	II
第1章 绪论	1
1.1 研究背景及意义	1
1.2 显著性检测研究的特点	3
1.3 国内外研究现状	7
1.3.1 自底向上的图像显著性检测	8
1.3.2 自顶向下的图像显著性检测	9
1.4 本文主要研究内容	10
1.5 本文章节关系	11
第2章 视觉显著性和本文算法的理论基础	12
2.1 图像的底层特征	12
2.1.1 颜色特征	12
2.1.2 亮度特征	12
2.1.3 纹理特征	12
2.1.4 形状特征	13
2.2 图像分割算法	13
2.3 几种经典的显著检测模型	14
2.3.1 IT算法	14
2.3.2 SR算法	15
2.3.3 RC算法	16
2.4 SLIC超像素分割	17
2.5 几种色彩空间模型	18
2.5.1 CIERGB色彩空间	18
2.5.2 CIEL*a*b色彩空间	19

2.5.3 CIEXYZ色彩空间	21
2.6 显著对象检测的计算框架	21
2.6.1 基于中心优先的显著计算框架	21
2.6.2 基于背景优先的显著计算框架	22
2.6.3 基于全局对比的显著计算框架	22
2.6.4 基于局部对比的显著计算框架	22
2.7 静态背景的概念及不足之处	23
2.8 本章小结	25
第3章 基于动态背景的迭代显著性检测算法	27
3.1 引言	27
3.2 动态背景的概念	27
3.3 本文算法的整体框架	31
3.4 区域显著性检测	32
3.4.1 基于动态背景的区域显著性检测	32
3.4.2 基于多种颜色特征的颜色对比差异的计算	33
3.4.3 基于中心优先的空间对比差异的计算	33
3.4.4 两种显著分类差异的融合	34
3.5 基于动态背景的显著差异的迭代	34
3.6 四个方向显著图的线性融合	35
3.7 本章小结	36
第4章 实验与分析	38
4.1 实验评测数据集	38
4.2 实验结果的客观评价标准	39
4.2.1 Precision和Recall曲线	39
4.2.2 F-measure曲线	40
4.2.3 MAE曲线	40
4.3 实验结果	41
4.3.1 本文算法和其它算法显著检测结果的对比	41
4.3.2 本文算法和其它算法F-measure曲线的对比	43

4.3.3 本文算法和其它算法的MAE曲线的对比.....	44
4.3.4 本文算法和其它算法的PR曲线的对比	45
4.4 实验结果分析.....	46
4.4.1 动态背景迭代的对比实验结果分析	46
4.4.2 空间的中心优先的对比实验结果分析	47
4.4.3 四个方向的对比实验结果分析	47
4.5 本章小结.....	48
第5章 总结和展望.....	49
5.1 总结.....	49
5.2 展望.....	50
参 考 文 献	52
致 谢	55
攻读硕士学位期间的研究成果.....	56

第1章 绪论

1.1 研究背景及意义

我们生活在一个智能化大爆炸的时代,而图像作为我们人类最直观的获取信息的方式,随着数码相机和智能手机等智能图像输出设备的普及,如何让计算机能快速的从海量的图像数据中提取我们人类最感兴趣的信息是当前计算机视觉研究领域的一个亟需解决的问题。受人类的视觉选择注意力机制的启发,图像的显著性检测研究是从图像中提取人类的视觉系统首先注意到的图像中的目标位置区域,其最终的检测结果是一张准确的反映了图像的各个像素的显著程度的灰度图,这张灰度图我们一般称之为图像的显著图。这样我们便可以把有限的计算资源有限分配在最显著的图像区域,进而进行后续的图像处理操作。

在当今的世界,各种智能图像输出设备产生了的海量的图像和视频数据,我们人类在过去10年里产生的数据量可能比过去几万年产生的所有数据量的总和还要多很多。面对海量的用户数据,怎么找到最能产生市场价值的数据,怎么有效地管理与分析这些数据并从中提取用户最感兴趣的信息,怎么让用户能够方便快捷地搜索到最相关的图像信息,这些成为了当前人工智能和计算机视觉领域里的热门的研究课题。如果我们仅靠人力来对图片和视频数据进行手动的标注,那么这样的方法显然是不科学的,也是代价昂贵的。因而我们必须开发出相应的智能信息处理算法,并利用计算机来自动的对人类每天产生的海量的图像和视频信息进行有效的处理,然后把最有价值的信息提取出来并建立情报数据库,从而更好的为人类服务。

心理学家们发现,在我们的日常生活中,存在多种注意的机制,根据它们分配认知资源的方式这些注意力的机制可以分为几种类型。根据Sohlberg和Mateer^[1]提出的注意力的分层模型,从最低的级别到最高的级别,依次可以划分为焦点注意、持续注意、选择注意、转换注意和分离注意。在这五种注意力中,目前的计算机视觉的研究主要专注在视觉刺激的选择注意上,也就是说,视觉注意的选择机制。

计算机视觉领域的研究科学家发现,如果能让计算机来模拟动物的视觉系统,那么我们的计算机就会具备相应的对自然界的相似的反应能力,这样我们将来制

作出来的机器人便有可能具备和人进行初步交流的能力。人们希望计算机能够模拟人类的视觉机制，像人类一样能够快速地对图像进行分析，从中提取关键的信息并分别加以存储和管理。人类的视觉系统能够快速地从复杂的场景中找出感兴趣的目标，因此，如何仿照人类的视觉注意机制来建立计算机的视觉注意模型是计算机视觉领域的研究热点之一。

我们每天所接受的80%的信息都是通过我们的视觉系统获得的。每一秒钟，我们的视网膜大概接收到了100亿比特的信息^[2]（Raichle, 2010），而我们的大脑皮层中的神经元的总数目大概只有100亿个^[3]（Shepherd, 2003）或者200亿个^[4]（Koch, 2004）。这个事实表明，我们所接受的视觉信息远远超出了我们大脑的存储能力。进一步的，大脑的处理信息能力非常有限，因此它无法同时对输入的视觉信息进行复杂的分析。为了解决这些问题，我们的视觉系统的一个中心任务就是如何高效的检测最重要的视觉子集，比如说，显著子集。这些视觉子集被快速的送入大脑并优先使用有限的大脑计算资源进行处理，同时其他的视觉子集就被抑制或者忽略掉了以加快大脑的信息处理效率。

为了能够有效地解决由大数据所带来的问题，一个很容易想到的办法就是从人类的视觉系统中学习经验。也就是说，我们能够通过模拟人类的视觉系统用来处理显著视觉子集的机制来处理大规模的图像和视频数据。用同样的办法，在计算机视觉中，为了对大量的图像数据进行高效的分析，我们可以把有限的计算资源优先分配给重要的视觉子集。更进一步说，通过采用类似人类视觉系统专注在显著视觉子集的机制，计算机分析的结果可以很好的满足人类视觉的感知特性。在这样的大背景下，视觉显著性的计算的问题被提了出来，目的是通过模拟人类视觉系统相关的机制来预测、定位和挖掘显著的视觉信息。

为了计算视觉显著性，一个关键的步骤就是通过模拟相关的人类视觉系统的机制来建立一个显著性模型。图1-1展示了基于本文提出的显著性检测模型的显著检测结果样例，其中奇数行是待检测的自然图像，偶数行是相应的显著图。从图1-1的显著图中我们可以看出，图像的大部分显著的区域都得到了凸显，同时有效的忽略了背景所在的区域。借助于本文算法所产生的显著图，我们可以做更

多后续的图像分析操作，如图像分割、图像变形等。



图1-1 本文算法显著检测实例，奇数行是原图，偶数行是显著图

1.2 显著性检测研究的特点

通过模拟人类视觉系统的选择注意机制，计算机可以像人类的视觉系统一样的处理视觉信息，经过显著性检测后的结果可以更好的满足人类的视觉感知需要。随着互联网的快速发展，人们在微博、微信、QQ及其相关的互联网社交媒体上非常活跃，图像和视频信息的总量呈现了爆发式的增长，亟需开发出智能的计算机视觉和多媒体应用来满足各种用户的需求。通常来说，一个应用可以被称之为智能应用，它应该要能提供对各种用户重要的和兴趣的图像和视频内容相关的服务。首先，从图像显著性的定义来看，显著性检测就是要检测图像或视频中显著的视觉子集。通过集中关注这些显著的视觉子集，图像或视频处理的结果可以紧密的关联在图像和视频的重要内容上。其次，人们感兴趣的目标通常是视觉显著的^[5]。因而视觉显著性可以在某种程度上反映用户的兴趣点。通过专注在图像和视频中的显著的内容，我们可以根据用户的兴趣为用户提供更多精准定位的服务。显著性研究在以下六大类上都有很好的应用，包括图像重定向、广告、图像和视频提取、图像融合、图像压缩和图像识别。

今天，人们手上使用的是显示各种各样的不同分辨率的图像或视频的设备，如智能手机、数码相机等等。在这些设备上，在过去显示不同分辨率的图像或视频信息的方法是通过添加黑框或者使用同一的分辨率显示图像或视频。然而，添加黑框不能很好的利用整个屏幕，使用统一的分辨率可能丢失重要的图像或视频信息。因此，非常有必要开发出自适应的图像或视频内容的显示技术，比如可以在保留图像或视频中显著信息的前提下对图像或视频的内容进行重定向来增强用户的观赏体验。为了解决图像内容重定向的问题，以色列的两位教授Shai和Ariel^[6]，提出了一种新的缩放裁剪图像的方法（如图1-2所示），进而我们可以对图像进行任意的改变但是却不会让图像中的特定内容变得扭曲。

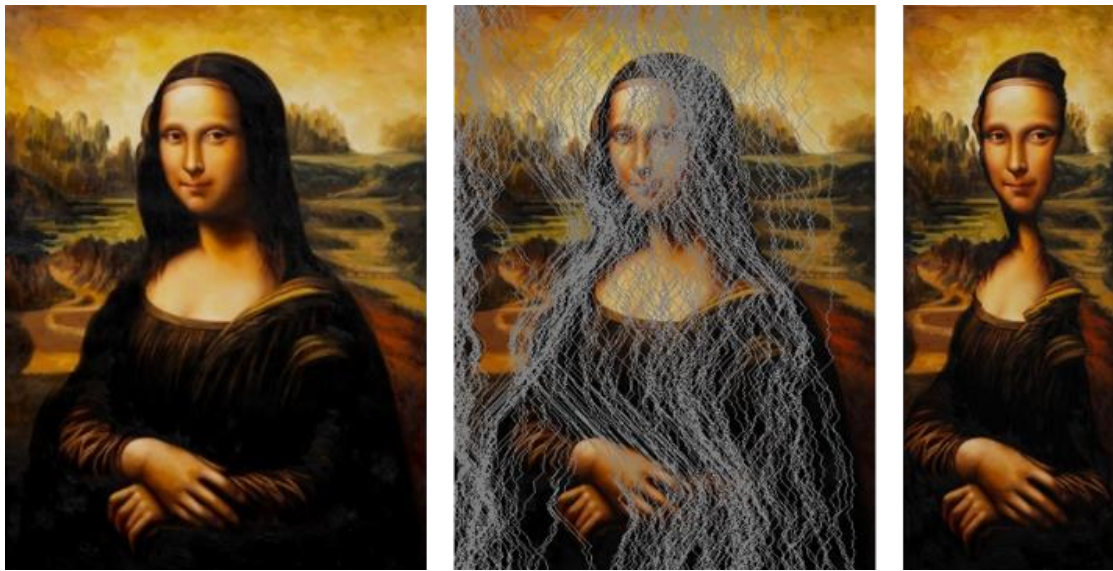


图1-2 图像重定向实例

和图像重定向技术相比，视频重定向技术更加具有挑战性，因为它要考虑更多的空间重定向结果的时间的一致性。更进一步的说，每一帧的显著的内容都是通过空间和时间因素联合决定的，因而很有必要同步处理多个后续的视频帧。因此，视频重定向技术通常采用一个优化的框架并且计算复杂度比图像重定向技术要复杂的多。Li^[7]提出了一种比较好的视频重定向的方法。在这种方法中，局部的时空显著性被应用于衡量视频的每个像素的重要程度。视频重定向的目标是去掉不显著的内容成分并保留最显著的内容。

最近这些年，互联网上的视频总量增长非常的凶猛。因此，提供个性化的视频推荐和在线的视频内容关联就成为了一件很有必要的事情。视频内容关联所指向的关联到附加资源（文本、图片和视频点击）的服务通过视频内容丰富了用户

的观看体验，因而诞生了一个关键的应用就是基于内容的视频广告投放服务。这种新的广告业务模式仅仅只针对性的提供用户最关心的相关内容。因此，基于视频内容的广告给企业提供了巨大的商机并通过强制的特定内容输入扩大了广告的投放能力^[8]。目前来说，主要有两种提供基于视频内容的广告方式，包括条幅广告和流媒体广告。

我们知道视觉显著性可以帮助定位图像和视频中的重要内容。一些研究采用了显著线索来提升对图像和视频内容的提取。其主要的思想是：从显著对象中提取出的视觉特征比从背景区域中提取的视觉特征可以更好的代表图像或视频本身。Fu^[9]提出了一种基于注意力驱动的图像内容提取系统。在他的方法中，图像被解读为包含多个可以感知的显著对象和背景的对象集合，集合中的每一个对象都有一个显著值。潜在的注意对象的显著值被映射成为影响因子并依据影响因子来进行图像内容的提取。在提取的过程中，当要强调显著对象的时候，图像上相关的内容是匹配上的。还有其他的一些相同的方法也被用作于视频内容的提取。比如，Li和Lee^[10]第一次根据运动和内容复杂度从场景中提取了一系列的关键帧。时空显著性检测算法被应用到了每一个关键帧上，并从中提取出一张显著图和一系列的注意区域。最后，视频点击的匹配的结果可以通过关键帧之间的相似性来衡量。在Fu^[9]中提出关键帧之间的相似性可以通过计算注意力区域之间的相似性来获得。

图像和视频的融合作为一个重要的视觉注意力和显著模型的应用，Ma^[11]构建了一个视频融合的系统并分别从有效性、健壮性和通用性三个方面阐述了注意力框架。在这种方法中，通过使用各种特征（如运动、外观、对象和语音）首先生成了一系列的注意力模型，这些注意力模型包括运动注意模型、静态注意模型、面部注意模型、相机注意模型、听觉注意模型、语音和音乐注意模型。Ma^[11]提出的注意曲线生成的流程图，如图1-3所示：

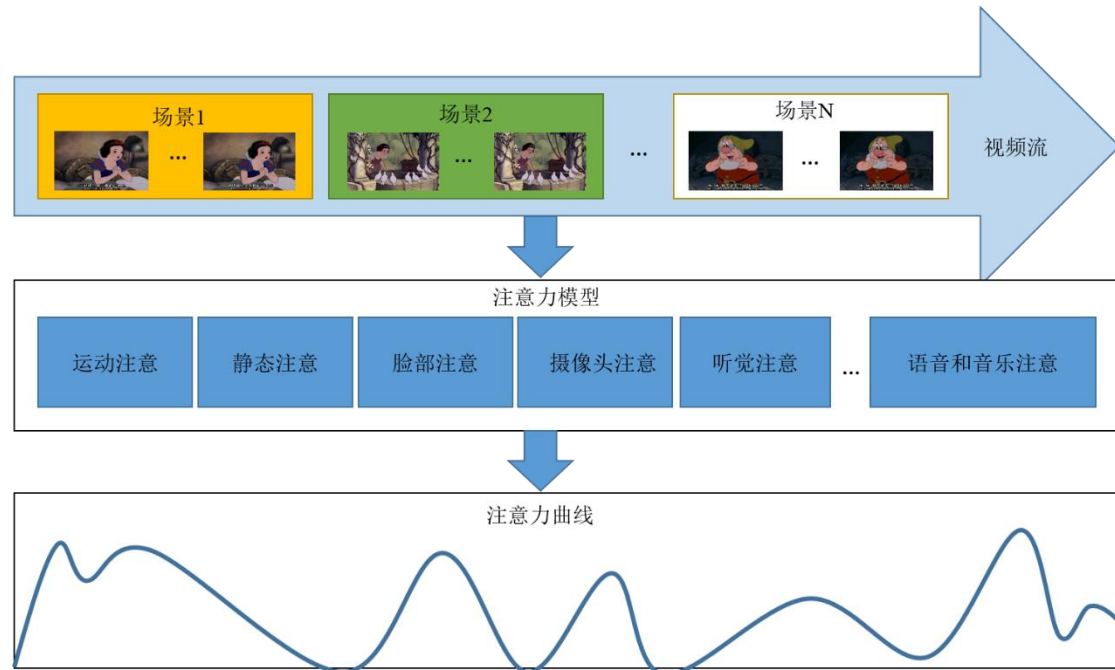


图1-3 Ma提出的注意曲线生成的流程图

给出全部的注意模型后，一个特别的注意曲线能够通过合成所有的注意曲线来构造。这种融合可以是线性的也可以是非线性的。依据给定的注意曲线，视频可以静态的融合也可以动态的融合。具体的流程图如图1-4所示。在融合的过程中，关键帧和视频浏览记录可以依据注意曲线的峰值提出取来，这就意味着这些视频片段很可能是最吸引用户注意的视频片段。为了获得精确的注意曲线峰值的位置，我们生成一条一阶微分曲线，那么这条曲线上取值为零的点就是峰值所在的位置。

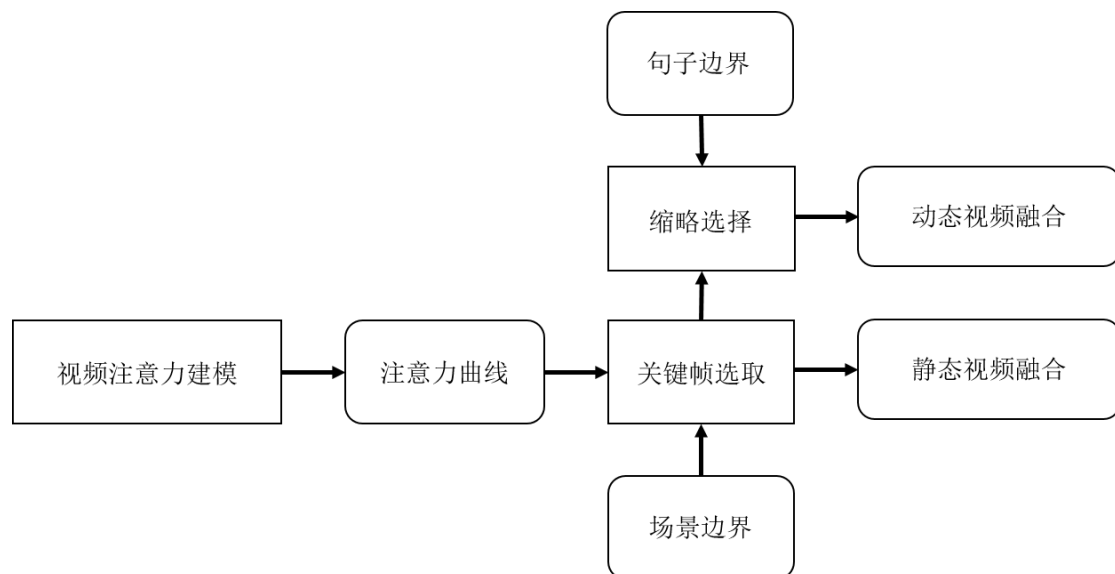


图1-4 Ma提出的视频融合方法的流程图

Liu^[12]提出了一种基于显著性的位分配策略进行视频压缩。其主要的思想是使用基于经典的中心周边对比差异的显著模型所生成的显著图来指导位分配。通过使用一种新的眼球追踪加权评价准则，与标准的编码方法按比率相比的话，由上面所提出的方法得到的超过90%的视频编码剪辑比标准的编码方法的结果质量要好。

本节我们介绍了一些相关的基于显著性的应用，然而显著性同时也可以应在其他的一些领域。比如说，Zhu^[13]采用了多尺度的结构化的显著性来做签名检测。Mahadevan and Vasconcelos^[14]提出了使用一个显著模型从高速的动态场景中移除背景的方法。Chalmond^[15]使用视觉显著性来分析大量的来自于远程传感设备的图片数据。Li和Clark^[16]提出了利用视觉显著性来帮助提取不变的特征。Siagian和Itti^[17]提出了利用视觉显著性来提升场景分类的性能。Yang^[18]提出了利用视觉显著性来做视频取证分析。在这些应用中，视觉显著性可以帮助分配有限的计算资源到重要的视觉内容上，从而可以对这些重要的视觉内容进行高效的分析，而这些分析的结果可以很好的满足人类视觉系统专注在视觉子集上的视觉感知特性。更一般的说，我们使用视觉显著性的最主要的目的即使提升用户的体验，比如增强用户的阅读体验的图像重定向、精准的广告投放、对用户感兴趣的视频内容的融合等等。值得注意的一点是，视觉显著性的这些应用，其客观的性能在考虑了视觉显著性后变化不大，但其主观评价却发生了巨大的变化，因为基于视觉显著性的分析可以更好的满足人类的视觉系统的感知需要。

1.3 国内外研究现状

在人类进化的非常漫长的这么一个经历中，人类对复杂和实时的场景的理解能力获得了很大的提升。人类是怎么使用大脑的神经元来实现这一复杂的认知过程的呢？为了深入的了解了人类视觉系统的选择性注意机制，心理学家和生理学家通过深入的研究认知反应时间和信号在生物途径扩张的传输速度，进而建立了计算机视觉的显著性检测的基本理论。在计算机视觉的显著性检测理论中，研究者认为人类的视觉系统只对图像中局部的区域感兴趣，而其余大部分的图像区域都是无效的输入信息^[19, 20]。在静态图像中，能够引起人类视觉注意的信号刺激主

要包括：新异刺激、适当刺激和预期刺激。根据这个，认知心理学家Triesman和Gelade^[19]与神经生物学家Koch和Ullman^[20]为代表的研究者们从人类视觉认知的角度将视觉注意机制分为两个阶段：

- (1) 第一个阶段为下意识的和自底向上驱动的显著目标的提取；
- (2) 第二个阶段为任务相关的和自顶向下驱动的显著目标的提取。

1.3.1 自底向上的图像显著性检测

自底向上的图像显著性检测算法主要关注由刺激信号本身所引起的视觉反应，这种方法单纯由数据驱动，与具体目标无关，是下意识的选择，因而比自顶向上的显著性检测更加快速。由于各种显著性检测方法非常多，这里主要介绍国际上大家公认的最有影响力的研究工作。

1998年，加州理工大学的Laurent Itti等人^[21]提出了一种基于生物启发模型^[20] (biologically plausible architecture) 和特征整合理论^[19] (feature integration theory) 的视觉显著性计算方法。这个方法首先从图像中提取颜色、亮度和方向等图像特征并生成相应的特征图。然后，在每个特征图内部，不同的位置按照赢者通吃 (winner take all) 的准则进行显著性竞争，以便使得局部最为突出的位置得以保留。最终，所有的特征图被按照自下而上的顺序融合成一张显著图。这张显著图可以有效的反映图像局部的显著程度。

2003年，微软亚洲研究院的Ma和Zhang^[22]提出了一种基于图像局部对比度分析的方法，并采用模糊增长的策略从显著图中提取用户感兴趣的图像区域。

2007年，上海交通大学的Hou和Zhang^[23]提出了一种基于频谱残留的方法来进行图像的视觉显著性检测。该方法通过分析输入图像的对数谱(log spectrum)，从图像的频域提取谱残留(spectral residual)，并在对应的空间域构建显著性图。

同年，美国加州理工大学的Jonathan Harel^[24]等人提出了一种基于图结构的视觉显著性检测算法。该方法首先为每个图像位置提取特征向量，然后利用这些特征向量构建动标图(activation maps)，最后对这些动标图进行归一化以突出和其它动标图相兼容的显著性组合。

2008年，Achanta^[25]等人使用图像在CIE L*a*b色彩空间中的亮度和色彩等底层信息，并建立了三个不同尺度的滤波器获取了三张不同对比度的显著图。最后，把三张显著图进行线性融合作为最终的显著图输出。

2009年, Achanta等人^[26]提出使用频率调整的方法, 这种方法直接定义显著性为图像中单个像素和整张图像的平均像素的颜色差异。

2010年, Goferman^[27]提出了一种基于上下文注意的显著性检测算法。该算法基于四个人类视觉注意的原则来检测图像区域中所代表的场景。在该算法中, 中心与周边的特征分布的对比是基于KL距离 (Kullback-Leibler divergence) 来进行显著对象检测的。

2011年, Cheng^[28]考虑在整张图片上的全局区域对比和跨区域的空间关系对比来计算图像的显著图。

2012年, Perazzi^[29]提出了一种可以通过使用高维的高斯滤波用一种统一的方式来进行完整的对比和显著值估计的显著性检测算法。

2012年, Shen和Wu^[30]提出了一种新的显著检测方法, 该方法基于低秩矩阵恢复的图像显著性检测模型并同时使用了低阶的视觉特征和高阶的视觉特征来进行图像的显著性检测。

2013年, MR^[35]提出了一种基于静态背景的显著检测方法, 即假定图像边缘的像素所在的超像素为静态的图像背景区域。该方法基于显著对象分割二分类的特性, 并使用机器学习中的流型排序来对图像块进行显著分类。在四个静态背景的作用下, 取得了不错的显著检测效果。

1.3.2 自顶向下的图像显著性检测

自顶向下的显著性检测方法通常都是根据具体任务对自底向上的检测结果进行尺度、方位、大小、形状、特征数目、阈值、类型、组合参数、描述形式等进行一定的调整而实现的^[31]。目前的比较流行的方法是根据任务示例或样本, 自动建立显著性检测模型。

2007年, 微软亚洲研究院的Tie Liu等人^[32]提出了一种基于条件随机场的学习模型来优化三种不同的显著图的组合系数。这三种不同的显著图分别是: 反映局部显著性的多尺度对比度 (multi-scale contrast)、反映区域显著性的中心邻域直方图 (center-surround histogram) 和反映全局显著性的颜色空间分布 (color spatial distribution)。示例样本中含有关于目标物体的大量指导信息, 能够有效地指导算法过程对潜在显著性物体进行正确的判断。但是, 不同的目标物体往往需要不同的物体特征、大量的训练样本和不同的学习方法, 因而不能得到更具普遍意义的

显著性物体区域。

1.4 本文主要研究内容

对于图像显著性检测研究而言,由于受到多方面因素的影响,目前该研究还处于探索的阶段。随着智能摄像设备的普及,产生了大量的图像和视频数据,目前人们对图像和视频数据的理解还处于探索阶段,从而使得显著性检测的研究具有广泛的应用前景和经济价值。本文根据该技术的研究现状、研究意义以及学术价值,针对图像的显著性检测问题,进行了一些探索性的研究。

本文的研究目标是设计一种快速通用的图像显著性检测算法。在本文算法的指导下,可以对任意的图像进行显著性检测,并取得不错的显著检测效果。

本文提出了一种基于动态背景区域的迭代显著性检测算法。主要的工作包括以下几个方面:

(1) 由于静态背景区域存在一些不足之处,因而本文提出了全新的动态背景区域。在本文中,动态背景区域的定义是根据一条人为设定的直线来对整张图片进行一个动态的区域划分,图像中直线左边划分出来的固定图像区域作为动态的背景图像区域,而直线的右边划分出来的图像区域则作为动态的待分类图像区域。

(2) 由于本文算法采用的是动态的背景区域,因而无需考虑图像中对象和背景的分布状况。动态背景区域的位置是灵活的,没有人为的划分图像的背景区域和对象所在的图像区域,并且同一个图像区域,在一种动态扫描方式上是背景图像区域,而在另一种动态扫描方式下则是待分类的图像区域。然后本文在动态的背景区域之上定义了动态的背景图像特征。由于动态背景区域本身是时刻变化的,因而作为参照的动态背景区域的平均色彩特征向量也是时刻变化的。而使用动态的图像特征其最主要的目的就是用来进行图像块之间的显著差异分类。也就是说,在本文所提出的算法中,动态背景区域的图像特征能够更好的描述作为参照的客观背景色彩特征向量的变化情况,因而在客观背景的选取上比使用固定图像背景特征的显著检测方法要先进。

(3) 在动态背景区域的基础上定义了两种显著分类差异。其中一种本文称之为基于色彩对比的显著分类差异,另外一种本文称之为基于空间对比的显著分

类差异。

(4) 在两种显著分裂差异之上，定义了显著差异的迭代计算。

(5) 本文在图像的上、下、左、右四个方向上进行了动态扫描，并基于每一个方向上的动态背景生成了一个显著图空间。

(6) 为了公正的评价本文所提出的算法，本文算法在大家公认的图像显著检测数据集MSRA-1000和MSRA-B上分别进行了PR曲线、F-measure曲线和MAE曲线的计算。

1.5 本文章节关系

本文共分为五个章节，其中每个章节的内容安排如下：

第一章为本文的绪论部分，介绍了本文的研究背景及意义、显著性检测研究的特点、图像显著性的国内外研究现状、本文主要研究内容以及本文的章节关系。

第二章首先介绍了自底向上的图像显著性检测研究所需要依赖的一些相关知识，包括图像的底层特征、图像分割算法以及几种经典的显著检测模型。随后简单介绍了与本文算法有关的一些图像显著性的理论知识。其中包括SLIC超像素分割、色彩空间模型以及图像显著对象检测的几种计算框架以及静态背景的概念。

第三章详细地阐述了本文所提出的基于动态背景的迭代显著性检测算法的计算框架。主要介绍了动态背景的概念、本文算法的整体框架、区域显著性检测（基于动态背景的区域显著性检测、基于多种颜色特征的颜色对比差异的计算、基于中心优先的空间对比差异的计算和两种显著分类差异的融合）、基于动态背景的显著差异的迭代和四个方向显著图的线性融合。

第四章是本文的实验和实验分析部分。主要包括实验评测数据集的介绍、实验结果的客观评价标准、实验结果以及实验结果分析。在大家公认的图像数据集MSRA-1000和MSRA-B上，本文将本文提出的算法与7个经典的世界公认的权威图像显著性检测方法和几个最近几年发表的图像显著性检测算法进行了定量和定性的对比，实验结果表明，本文提出的算法具有一定的先进性。

第五章对本文进行了一个总体的概括，并指出了将来进一步的研究方向。

第2章 视觉显著性和本文算法的理论基础

2.1 图像的底层特征

在图像的显著性检测中，一般都是使用图像的底层特征来构造特征向量，几乎所有的计算机视觉算法中都或多或少的使用了颜色、亮度、纹理、方向和边缘等图像特征中的一种或几种。在本节中，我们将对这些图像特征进行简单的描述。

2.1.1 颜色特征

颜色特征是图像显著性检测中使用最为广泛的一种图像特征。因为和其它图像底层特征相比，人眼对于颜色的反应非常敏感，这是我们与生俱来的本能。

随着现在机器学习的兴起，人们对智能的分类算法越来越感兴趣，而图像分类的前提就是要找到分类的特征向量，对图像而言这个特征向量的成分最好就是颜色了。一般使用颜色特征有两种用途，一种是用于用欧氏距离来度量像素间或图像块之间的色彩差异，另一种是用于计算图像某个区域的颜色平均值，把整个图像区域用这个均值代替。把这个区域当成一个像素点，从而带来速度上的极大提升。

2.1.2 亮度特征

亮度特征是图像中最基本的特征之一，根据RGB色彩空间与HIS色彩空间之间的映射关系，图像的亮度特征可以通过公式（2.1）得到。

$$I = \frac{r + g + b}{3} \quad (2.1)$$

其中 r , g , b 分别代表在RGB色彩空间下，图像在三个色彩通道上的颜色值， I 是亮度值，对于亮度特征的使用，局部区域亮度来表示该区域，然后用于计算区域间的亮度差异。

2.1.3 纹理特征

纹理是指图像中小范围的形状很小的或者有规律地排列的图案。纹理特征是指图像的灰度等级的变化，这种变化是与空间统计相关的。图像的纹理特征反应了图像本身的属性。

纹理特征的提取方法分为基于结构的提取方法和基于统计的提取方法。在计算机视觉的应用中，LBP（Local binary patterns）方法是一种常用的图像特征分类的方法。LBP方法在1994年首先由Ojala, Pietikainen^[36]和Harwood^[37]提出，用于纹理特征提取。有人将HOG特征分类器和LBP方法混合在一起使用，提升了该方法在数据集^[38]上的检测有效性。

作为物体表面的一种特性，纹理无法准确的反应出物体的本质属性，因而，使用纹理特征无法进行高级的图像理解。纹理必须对多个像素点进行统计运算，因而和基于单个像素的颜色特征不同。在图像的模式匹配中，纹理特征要优于单个像素的颜色特征，基本不会出现因为局部误差而引起的模式匹配失败。但是，纹理是一种具有旋转不变性的图像的统计特征，因而对与噪声不怎么敏感。

2.1.4 形状特征

形状特征和颜色、纹理等底层特征不一样，图像中的区域或者物体的划分是形状特征表达的基础。一般说来，形状特征可以划分为两大类，一类是以轮廓特征为基础的，另一类是以区域特征为基础的。图像的局部特征和物体的整个形状都有关系，但轮廓特征主要和物体的外边界有关联。几种常见的形状特征描述方法有：边界特征法、傅立叶形状描述符法、几何参数法和形状不变矩法等。

2.2 图像分割算法

在计算机视觉的各种高级应用中，图像分割是十分重要的预处理。本文算法中所使用的SLIC超像素分割也是图像分割方法中的一种。但是，目前各种图像分割方法依赖的只是图像中像素的亮度和颜色。因此，在实践中仍然存在各种不足。也就是说，图像分割是仍然需要进一步研究的技术。常用的图像分割方法有基于梯度下降的方法（Watershed^[39]，Mean shift^[40]，Quick Shift^[41]，Turbopixels^[42]，SLIC^[43]，SEEDS^[44]）和基于图论的分割方法（Normalized cuts^[45]，Graph-based segmentation^[46]，Graph cuts^[47]，GCa10 and GCb10^[48]，Entropy Rate Superpixel Segmentation^[49]，Superpixels via Pseudo-Boolean Optimization^[50]，GrabCut^[51]）。

2.3 几种经典的显著检测模型

2.3.1 IT算法

IT算法^[21]是一种基于检测局部不一致性的算法。在这个方法中，一个位置视觉显著性被量化为它和它的周边区域在多个特征和多个尺度上的差异。Itti^[21]模型中，显著值是像素点在颜色、亮度、方向方面与周边背景的对比值，该算法的主要的计算框架可以分为3个主要的部分，其中包括：1)早期注意特征的提取 2)多尺度中心优先的计算 3)对比显著图的整合。其计算框架如图2-1所示：

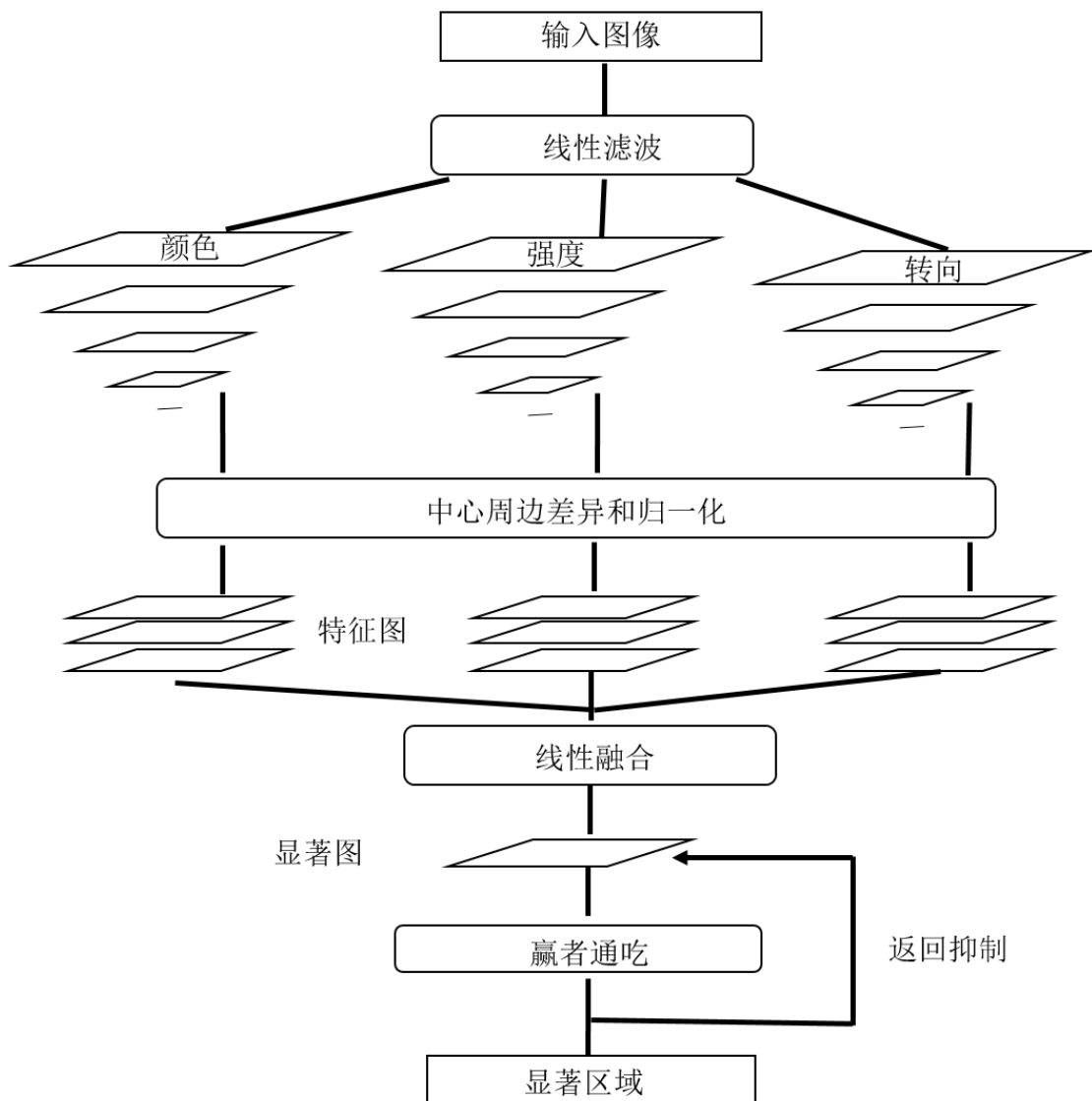


图2-1 Itti算法的主要计算框架

2.3.2 SR算法

侯晓迪在2007年提出了一种和图像特征独立的方法（SR^[23]算法），并命名为频谱残差法。通过分析输入图像的1幅度谱，他在图像的频域提取出了输入图像的频谱残差信息，并提出了一种快速的显著性检测方法通过在空间域来构造输入图像的显著图。他通过分析图像的幅度谱从而获得了图像的频谱残差。然后把频谱残差转换到空间域上面，进而就得到了图像的显著图。

SR算法^[23]效果图，如图2-2所示：

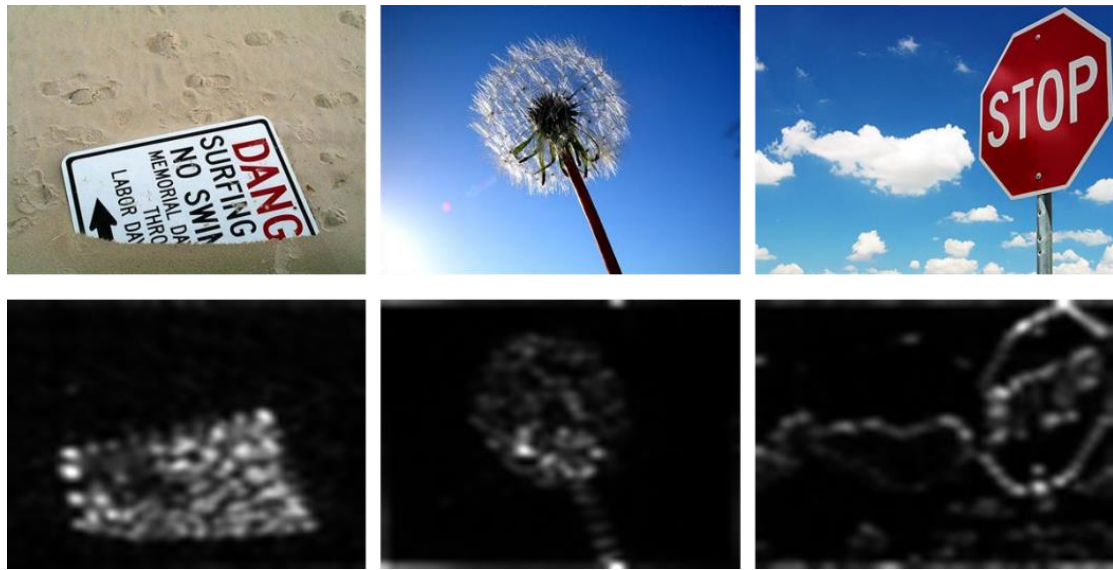


图2-2 SR算法显著检测实例，

第一行代表数据集中的图像，第二行代表SR算法的显著检测结果

他认为，图像中相似的信息就是多余的。他通过实验进一步得出结论，自然图像的幅度谱不是没有规律的，从图2-3中可以看出不同的输入图像的对数幅度谱都具备大致相同的走向。

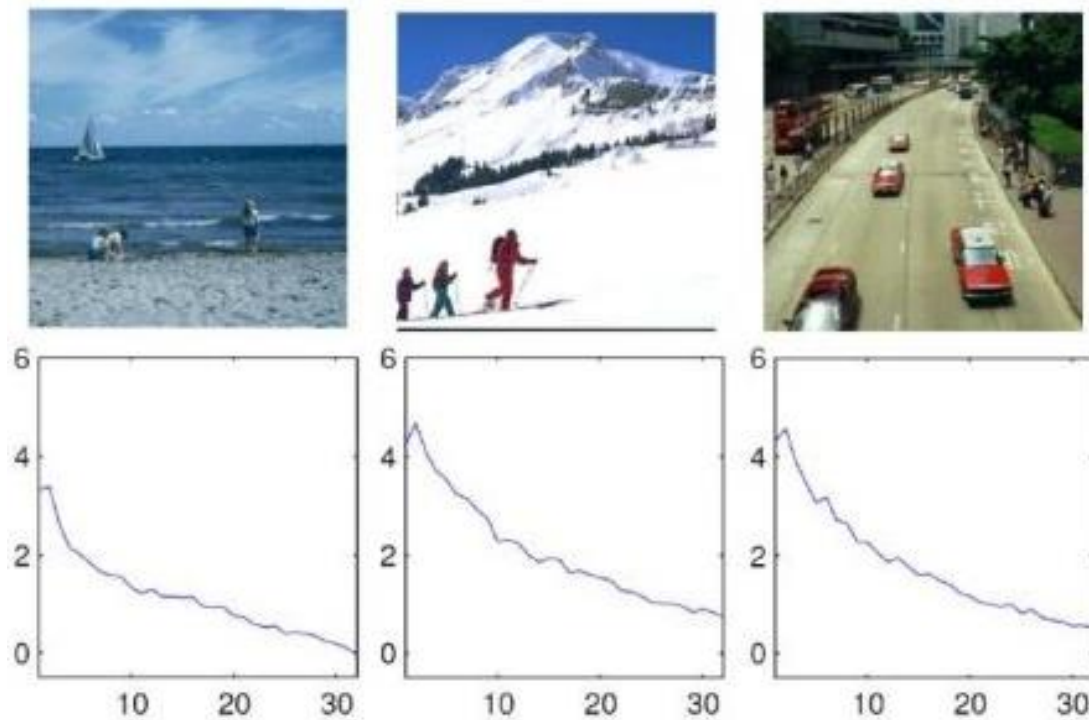


图2-3 自然图像的幅度谱的相似性

2.3.3 RC算法

RC算法^[28]是程明明在2011年提出的一种基于全局对比的显著检测方法。该方法为HC算法^[28]的改进版。

HC方法^[28]根据图像中的每一个像素与所有其他像素的颜色对比差异来定义单个像素的显著程度，进一步产生出相应的图像所对应的显著性图。程明明提出使用量化后的图像的直方图来表示图像的颜色特征，从而极大的提高了算法的运算速度。

RC算法^[28]作为HC方法的改进，作者结合空间关系提出了RC方法。首先将图像分割成区域，再为每个区域分配显著性值，从而形成基于区域对比度的(Region based Contrast, RC方法)显著性图。区域的显著性值由全局对比度值计算得到，其中，全局对比度值由当前区域相对于其它区域的对比度以及空间距离来计算。这种方法更好地反映了图像区域关系与显著性确定之间的联系。



图2-4 RC算法的显著检测实例

2.4 SLIC超像素分割

2003年Ren和Jitendra^[52]第一次提出了超像素的概念，他们认为超像素是具有相似纹理、色彩和亮度等特征的由相邻像素构成的具有一定视觉含义的形状不规则的像素块。它根据像素之间的相似程度来对像素进行分组，这样的话就可以用单个的超像素来表达图像的特征，因而大大降低了图像分割算法的复杂度。一般来说，研究者们都是使用超像素分割算法作为图像分割算法的预处理阶段。

Achanta^[53]等人提出了SLIC (simple linear iterative clustering)，即简单的线性迭代聚类。SLIC超像素分割算法在2010年首次提出，是一种被显著性检测领域广泛认可的作为图像分割的方法。该方法基于CIEL*a*b颜色空间和XY坐标位置构造一个复合的5为特征向量，并使用k-means对图像中的像素进行局部的聚类。SLIC所生成的超像素，具有排列均匀、一致的特征，而且综合性能也不错，比较符合人们对分割效果的期望。

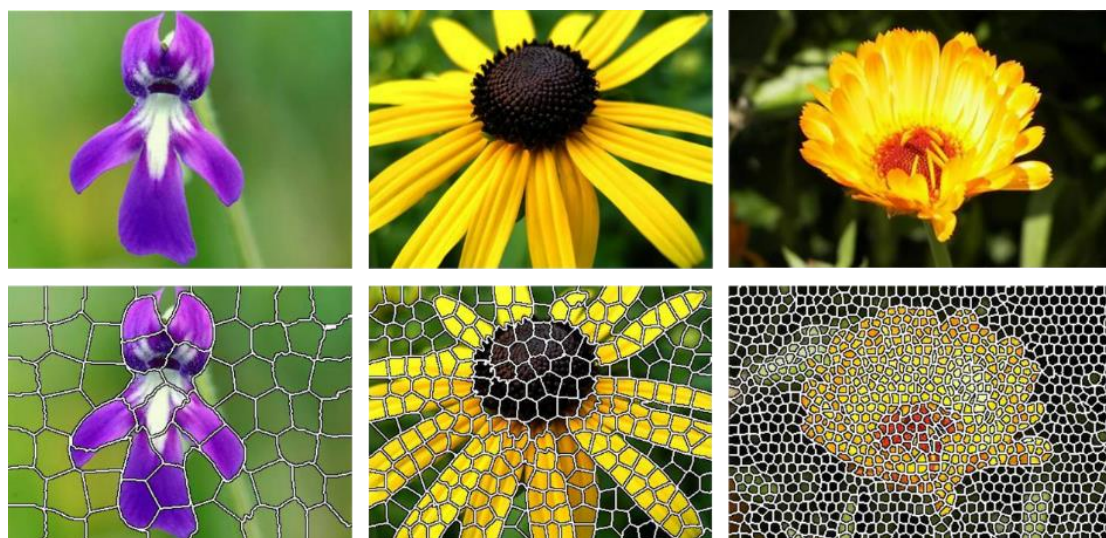


图2-5 第一行为原图像，
第二行的超像素的大小依次取64像素、256像素和1024像素

2.5 几种色彩空间模型

对图像的感官刺激是人眼的天生的本能。我们的视觉系统可以准确的分辨出图像中哪些物体是我们第一时间关注到的（也就是最感兴趣的），哪些物体不那么重要。而在这个过程中，起决定性作用的就是图像中的色彩特征，因而图像的色彩特征成为了人类视觉系统所能够感知到的最基本的图像特征之一。

2.5.1 CIERGB色彩空间

RGB色彩模型是是目前运用最广的颜色系统之一，也是一种在工业界广泛采纳的颜色标准。在RGB色彩系统中，各种颜色的获得是通过对红（R）、绿（G）和蓝（B）三个颜色通道的叠加而起作用，几乎人类视觉所能感应出的所有的颜色都能通过RGB色彩模型表示出来。

在我们日常所使用的各种设备的显示屏幕上，大多都是采用了基于RGB颜色标准。目前主流的计算机显示器都能够显示32位的颜色，也就是说都能显示超过一千万以上中颜色的色彩变化。而计算机显示器一般也都是采用了基于RGB的颜色标准，屏幕上显示的所有颜色都是这R（红）、G（绿）和B（蓝）三种颜色的复合色。在色彩领域，R（红色）、G（绿色）和B（蓝色）通常也被称为三原色。而我们日常经常会听到，对电脑的“亮度”进行调整，这里的亮度指的就是R（红）、

G（绿色）和B（蓝色）这三种三原色的平均值。

由RGB色彩空间的定义可以看出，256个阶的RGB色彩一共能产生超过1678万种色彩，即 $256 \times 256 \times 256 = 16777216$ 。通常也被简称为1600万色或千万色。也称为24位色（2的24次方）。而在我们现在日常所使用的LED电视机中，LED采用了三合一点阵全彩技术，其中每一个发光单元是一个由RGB三色晶片构成的全彩像素。LED显示技术使得图像的显示画面更加逼真，色彩更加鲜艳，因而LED显示技术相关的产品也赢得了广大的市场份额。

CIERGB的色彩空间模型，如图2-6所示，

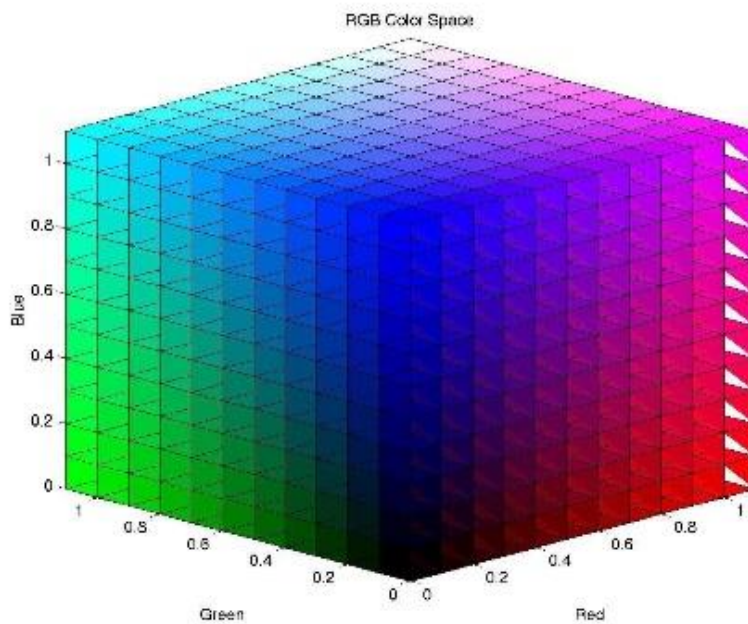


图2-6 CIERGB色彩空间模型

2.5.2 CIEL*a*b色彩空间

CIEL*a*b颜色空间没有CIERGB使用的那么广泛。1931年，国际照明委员会（CIE）制定了关于CIEL*a*b的颜色度量国际标准。CIEL*a*b是一种基于人类的视觉特征的颜色系统，因而也是与设备无关的。换句话说，在CIEL*a*b色彩空间中，人的视觉系统对色彩的敏感程度是可以用数字来度量的。在CIEL*a*b中，L分量代表的是图像中单个像素的亮度，其取值区间为 $[0,100]$ ，意思是从纯黑到纯白。L=50时，就相当于50%的黑；a分量代表的是颜色在红色和绿色之间的变化区间，其取值区间为 $[127,-128]$ ；b分量代表颜色在黄色和蓝色的之间的变化区间，其取值区间为 $[127,-128]$ 。图2-7所示为CIEL*a*b颜色空间的图示。

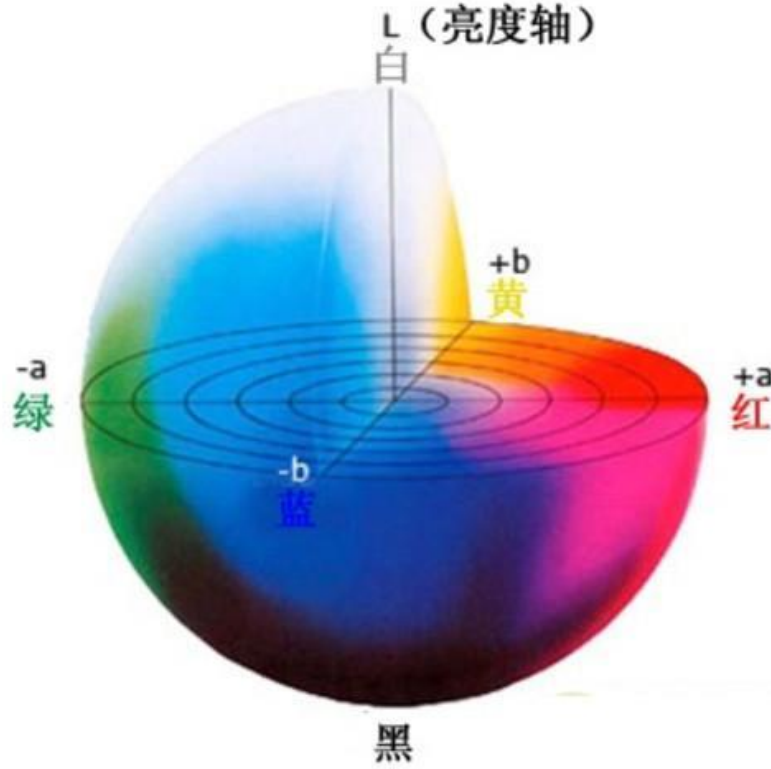


图2-7 CIE L*a*b*颜色空间

CIE L*a*b*色彩模型具有两大优点：色域宽阔和不依赖于设备。它能表示 CIERGB和CMYK的所有色域，同时也能表示这两个色彩空间无法表示的颜色。CIE L*a*b*色彩模型可以表示几乎人的肉眼能感知的所有色彩。在图像的显著性检测中，由于我们是要模拟人类的视觉系统的行为机制，因而CIE L*a*b*颜色模型是最佳的选择。

CIE L*a*b*色彩空间无法从CIERGB色彩空间直接进行转换，常规的做法是先将 CIERGB色彩空间转换为CIEXYZ色彩空间，再从CIEXYZ色彩空间转换到CIE L*a*b*色彩空间。其转换的公式如下：

$$\begin{cases} L^* = 116f\left(\frac{Y}{Y_n}\right) - 16 \\ a^* = 500\left[f\left(\frac{X}{X_n}\right) - f\left(\frac{Y}{Y_n}\right)\right] \\ b^* = 200\left[f\left(\frac{Y}{Y_n}\right) - f\left(\frac{Z}{Z_n}\right)\right] \end{cases} \quad (2.2)$$

其中，

$$f(t) = \begin{cases} t^{\frac{1}{3}}, t > \left(\frac{6}{29}\right)^3 \\ \frac{1}{3}\left(\frac{29}{6}\right)^2 t + \frac{16}{116}, t \leq \left(\frac{6}{29}\right)^3 \end{cases} \quad (2.3)$$

$X_n = 0.950456, Y_n = 1.0, Z_n = 1.088754$ 为白色点在CIE XYZ色彩空间中的三色刺激值，n代表normalized，即归一化。

2.5.3 CIEXYZ色彩空间

CIE XYZ色彩空间是从1920年代后期Wright^[53] (Wright 1928) 和Guild^[54] (Guild 1931) 做的一系列实验中得出的。在三色加色法模型中，如果某一种颜色和另一种混合了不同份量的三种原色的颜色，均使人类看上去是相同的话，我们把这三种原色的份量称作该颜色的三色刺激值。CIE1931色彩空间使用X、Y和Z定义了颜色的三色刺激值。

CIEXYZ色彩空间与CIERGB色彩空间的之间的转换关系如下：

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \frac{1}{0.17697} \begin{bmatrix} 0.49 & 0.31 & 0.20 \\ 0.17697 & 0.81240 & 0.01063 \\ 0.00 & 0.01 & 0.99 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (2.4)$$

为了综合评估CIEL*a*b、CIERGB色彩空间对图像的显著性检测效果的影响，A.Borji等人^[56]在四个图像数据集上进行了实验。实验结果表明各种颜色空间各有各的优劣。从提高显著性检测的效果的角度看，在有些数据集上CIEL*a*b色彩空间要优于CIERGB色彩空间，而在另一些数据集上CIERGB色彩空间要优于CIEL*a*b色彩空间。但在所有数据集上的实验结果都一致表明，同时使用两种或两种以上色彩空间进行图像显著性检测要比仅使用单一色彩空间可以获得更高的图像显著性检测的准确率。在作者的研究成果的指导下，我们在本文的显著性检测算法中同时使用了CIE RGB色彩空间、CIE L*a*b色彩空间和CIE XYZ色彩空间三种色彩空间来定义图像的颜色特征。

2.6 显著对象检测的计算框架

2.6.1 基于中心优先的显著计算框架

人们在使用数码相机拍照的时候，他们通常都会把自己感兴趣的对象定格在

镜头的中央，因而显著的对象通常都会分布在图像的中央。根据这个假设，Tilke Judd^[57]提出了一种基于中心优先的视觉显著性检测方法，并建立了一个包含1003张视觉跟踪图像的数据集。

2.6.2 基于背景优先的显著计算框架

不同于基于中心优先的显著性检测方法，背景优先的显著性检测方法从对立面来考虑图像的显著性检测的问题。基于中心优先的显著性方法考虑的是显著对象在哪的问题，即假设显著对象通常都分布在图像的中心，而基于背景优先的方法考虑的是背景在哪的问题，即当我们把背景检测出来后，留下的自然就是对象了。Wei^[33]提出了一种基于背景优先的测地距离的显著性检测方法，他提出了两种基于背景优先的检测背景的方法（边界优先和连接优先），“边界优先”是基于显著对象有可能偏离了中心图像区域，但是它们很少触及图像的边界。“连接优先”是基于现实生活中的自然图像中的背景区域都是很大一片并且是外观相似的。这两种方法有效的检测出来了潜在的背景区域，并在主流数据集上都取得了不错的效果。

2.6.3 基于全局对比的显著计算框架

用单个像素或一个区域和整张图像作对比来计算图像的显著性的方法，研究者称之为基于全局对比的显著性检测算法。Achanta等人^[26]在频率的基础上定义了像素级别的显著性，该方法根据单个像素和整张图像的平均色差的来确定单个像素的显著性值。Zhai和Shah^[58]在像素级别定义了全局对比的显著性，基于某个像素和其余像素的对比度来计算像素级别的显著性。Chen^[28]定义了基于像素块级别的显著性，并同时在像素块级别定义了基于空间对比的显著计算，分别利用颜色信息和空间信息定义了图像的显著性。

2.6.4 基于局部对比的显著计算框架

局部对比的显著检测方法使用单个图像区域相对于局部邻域的显著程度来检测显著性区域。Itti^[21]提出了一种基于生物启发模型（biologically plausible architecture）和特征整合理论^[19]的视觉显著性计算方法。Ma和Zhang^[22]提出了另一种基于局部对比度分析方法，并采用模糊增长的策略从图像的显著图中提取用户感兴趣的图像区域。Liu等人^[32]提出了一种基于条件随机场的学习模型来优化

三种不同的显著图的组合系数。Goferman等人^[27]提出了一种基于上下文注意的图像显著性检测算法，该算法的目的是检测场景中重要的代表性的图像区域。这些基于局部对比的显著性检测方法更多的是检测出图像所在边缘部分的显著程度，而不是对显著对象所在的整体区域进行检测。

2.7 静态背景的概念及不足之处

显著的目标一般不会出现在图像的边界上，因而Wei^[33]根据这个假设提出了一种新的通过检测背景所在的区域间接的检测出显著对象所在的区域的图像显著性检测模型。他所提出的主要的观点是一对背景区域中的图像块之间的距离一定会小于一个显著对象区域中的图像块和一个背景区域中的图像块之间的距离。在这个基础上他定义了背景优先和连接优先，并把这种方法称之为基于测地距离的图像的显著性检测。但是，这种方法有两个不大好的地方，一个是一旦显著对象正好占到了图像的边界上，那么这个边缘的图像块就会被认定是不显著的。另一个是，小块的孤立的背景区域会被认定是显著的区域。出现这两个的问题的原因都在于，在自然图像中，显著对象和背景所在的位置都是无法事先估计的，像这样用预设的固定区域的图像特征来度量图像中区域的显著性显然不是特别的合理。

受Wei^[33]所提出的背景优先概念的启发，Lu^[35]提出了一种基于固定背景的计算图像显著性的方法，即假定图像边缘上的像素所在的超像素为固定的背景区域，从而用机器学习中的流型排序对图像中剩下的超像素进行排序，最后把排序所得到的值作为超像素的显著值，从而得到最终的四个方向上的基于背景的显著图。Lu^[35]中所提出的一张图像的静态背景的如图2-8所示，其中静态背景分为上边背景（上边界上的像素所在的超像素所代表的图像区域）、下边背景（下边界上的像素所在的超像素所代表的图像区域）、左边背景（左边界上的像素所在的超像素所代表的图像区域）和右边背景（右边界上的像素所在的超像素所代表的图像区域）。

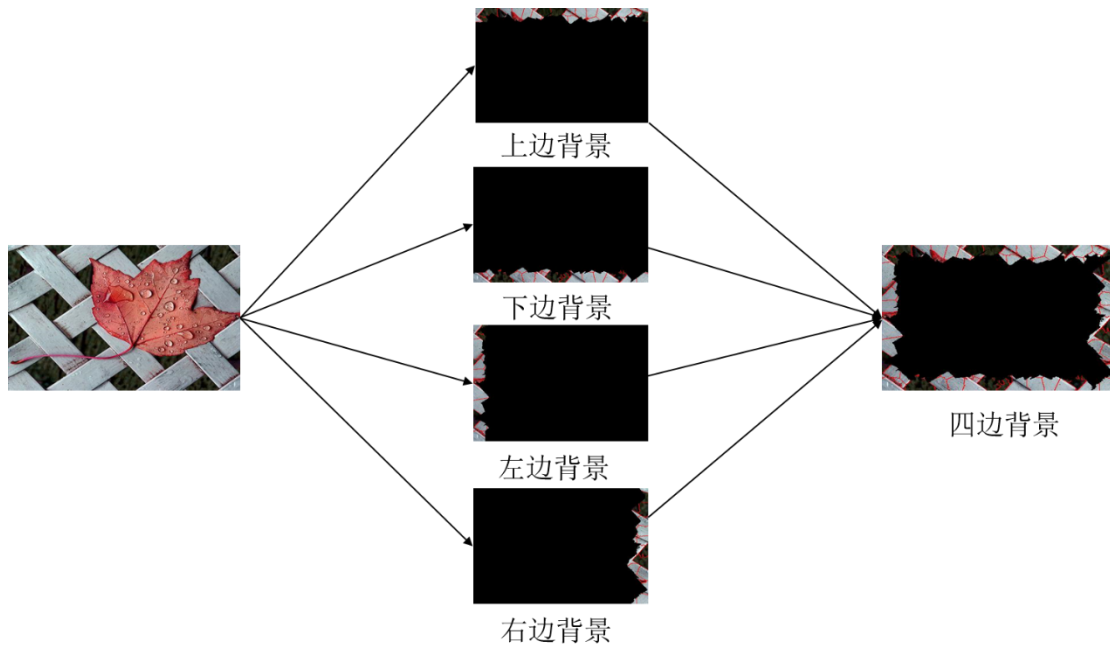


图2-8 Lu中提出的静态背景

Qin^[60]提出了一种基于元胞自动机的显著检测算法。和Lu^[35]一样，该算法首先假设图像的边缘像素所在的超像素为固定的背景区域，进而定义相关的显著分类算法，最后使用元胞自动机对相邻图像块之间的显著程度进行平行进化。

静态的背景有一些不足之处，比如说，当显著对象占据图像的边缘（图2-9中的第一行的图像）或者边缘像素所在的超像素的图像特征和显著对象极为相似的时候（图2-9中的第二行的图像），那么MR^[35]和BSCA^[60]中所提出来的显著检测算法就会失效。图2-9所示的是MR^[35]、BSCA^[60]的错误检测实例与本文算法的显著检测结果的对比图。对于图2-9的第一行中的图像，由于在图像的边缘存在30个像素左右的黑框，MR^[35]采取的办法是，直接去除边缘所包含的30个像素黑框。在去除了黑框之后，对象所在的区域已经占据了图像的边缘，而MR^[35]把边缘像素所在的超像素看作是背景区域，因而未能成功的检测出图像中间的左边白色的显著部分。又由于把边缘上的皮肤所在的超像素看作是背景区域，恰好这块超像素又包含了一只右边的眼睛，因而把右边的一只眼睛检测为背景区域，而把另外一只远离边界的眼睛检测为显著区域。同样的，在BSCA^[60]中也出现了这样的问题。而本文所提出的算法，由于采用的是动态的背景区域，因而像素块的显著程度只和像素块本身被计算显著差异的次数有关，而和图像块在图像中的具体的分布位置无关。对于图2-9的第二行中的图像，图像的上边界的像素所对应的超像素的

颜色和图像中的显著对象的颜色是一致的。这种情况下，如果我们采用静态的背景来度量图像中其他图像块的显著差异的话，就会出现误检测的情形。对MR^[35]来说，图2-9中第二行中的第二张图片就出现了显著对象的大部分区域由于和上边界的背景的颜色高度一致，而没有被检测出来的情形。对BSCA^[60]来说，同样存在和MR^[35]类似的问题，但由于采用了元胞自动机对相邻的图像块进行显著值动态调整的缘故，导致被检测出来的显著的区域几乎占据了除去人为设定为背景区域的整张图像。对于这种图像，由于背景的一部分与显著对象的颜色高度一致，我们很难通过一些常规的图像特征做到只检测出显著对象而不检测出背景。目前所有的显著检测算法中，只有少数的几个图像显著检测算法能正确的检测出这类图像中的显著对象，但一般也是采用多尺度的方式来实现的，如DSR^[61]。

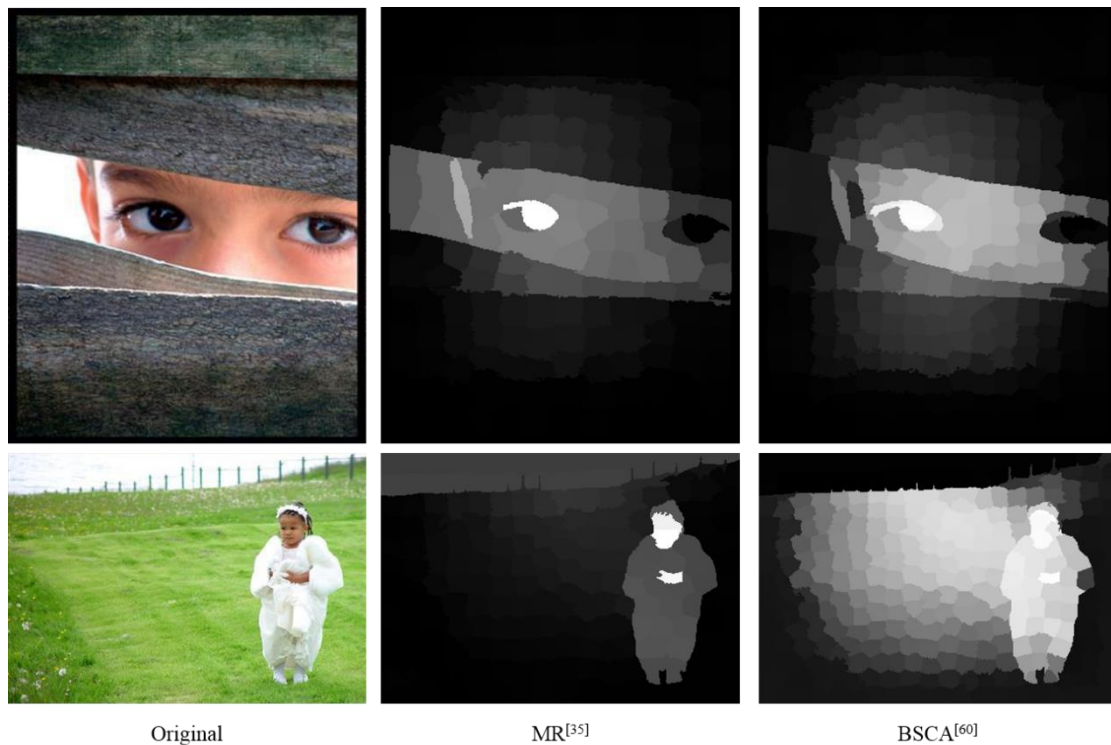


图2-9 MR、BSCA的错误检测实例

2.8 本章小结

目前自底向上的图像显著性检测研究一般都是采用图像的颜色特征、亮度特征、纹理特征、频谱特征等基本的图像底层特征来构造显著分类的特征向量，事实上这些基本的图像特征也是我们人眼能够直接分辨出来的最明显的图像特征。

如IT算法^[21]采用了颜色、亮度、方向作为基本的显著分类的特征向量。SR算法采用的是图像的频率特征。IT算法^[21]和都是在像素级别进行图像的显著性检测研究。而RC^[28]算法则采用图像块在CIEL*a*b上的颜色信息作为基本的显著分类的特征向量。本文算法采用了图像在三个色彩空间（CI RGB、CIEL*a*b和CI XYZ）的颜色信息以及超像素块的中心坐标作为基准的图像特征向量。在基于色彩信息的特征向量上本文定义了基于欧氏度量的局部对比显著性，而在基于位置坐标的特征向量上本文定义了基于位置的中心优先的显著性。另外，本章也同时指出了静态背景的来源以及采用静态背景会带来的误检测的问题。

第3章 基于动态背景的迭代显著性检测算法

3.1 引言

本章提出了一种基于动态背景的迭代显著性检测算法。与基于固定背景区域的显著性检测算法不同，我们无法预先估计图像背景在图像中的分布状况，因而这个假设有时可能不成立。本章首先由静态背景引申出了动态背景的概念，然后本文在动态背景的基础上定义了两种显著分类差异，其中一种本文称之为基于图像色彩特征的颜色对比差异，另一种本文称之为基于中心优先的空间对比差异。另外，由于单次的显著差异的度量并不能很好的反应出图像块的显著差异状况，因而本文定义了基于这两种显著分类差异的显著差异迭代。其次，单个方向上的动态背景无法保证所有的图像块都接受大致相同次数的显著分类，因而本文定义了基于四个方向上的动态背景。最后，对四个方向上的动态背景所生成的单张的显著图，本文对其进行了线性融合并最终得到了单张的显著图。

3.2 动态背景的概念

由于静态背景有一定的局限性，比如图2-9中的第一张图像的周围有一个大概30个像素宽的黑色矩形框。同时由于采用静态背景会带来一些由于静态背景本身所带来的误检测的问题，如图2-9中出现的问题等。本文受Wei^[33]和Lu^[35]中提出的静态背景概念的启发，为了解决由于静态背景所带的一些问题，从而本文提出了动态背景的概念。在本文中，动态背景是对图像的一个动态的区域划分。动态背景的通俗定义是：对于一幅图像，使用一条直线将图像划分为两个区域，那么随着直线的运动，图像会被自动的划分为两个动态区域。在这两个图像区域中，我们称直线左边的图像区域为动态的背景区域，而直线右边的图像区域为动态的待分类的图像区域。图3-1、图3-2、图3-3和图3-4是本文中所提出的四个方向上的

动态背景区域和待分类的图像区域。

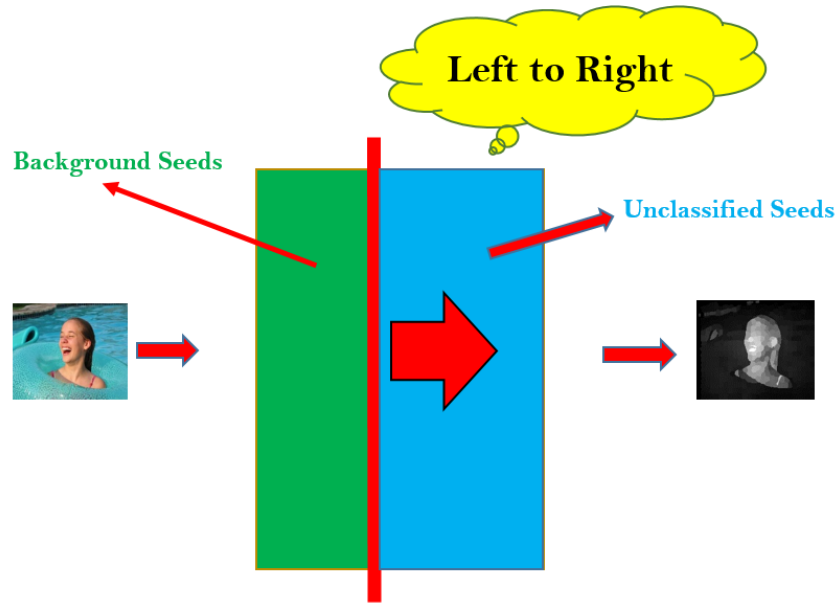


图3-1 从左至右扫描的动态背景区域

从图3-1中，由箭头的指向可以看出，当前所定义的是从左往右移动的动态背景，而且这条直线把整个图像区域从左往右划分为两个动态的图像区域。随着直线从左往右的慢慢移动，直线左边的动态的背景区域的面积越来越大，而直线右边的待分类的图像区域的面积则越来越小。这样的话，用直线左边图像区域的平均色彩特征来度量直线右边的图像区域中的图像块，会导致越是处在直线右边图像区域中的图像块，其被计算显著分类的次数也就越多。越是处于直线左边图像区域中的图像块，其被计算显著分类的次数也就越少。

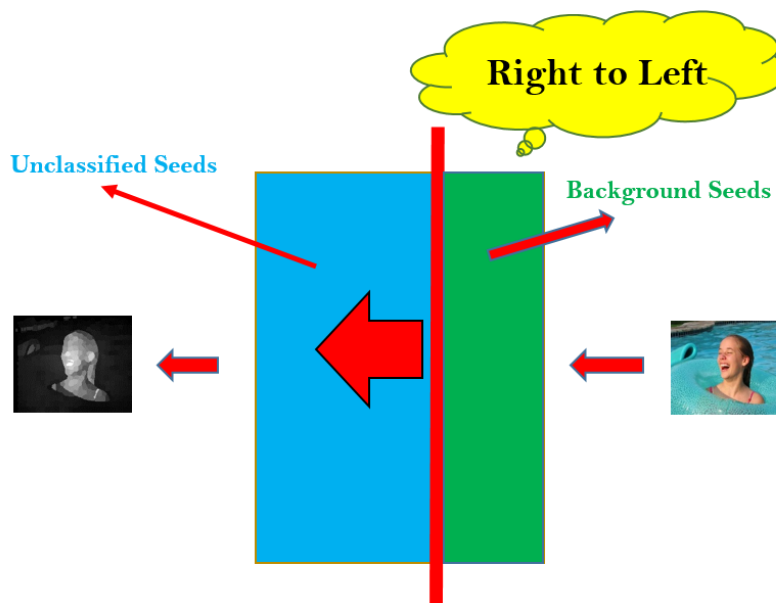


图3-2 从右至左扫描的动态背景区域

从图3-3中，由箭头的指向可以看出，当前所定义的是从上往下移动的动态背景，而且这条直线把整个图像区域从上往下划分为两个动态的图像区域。随着直线从上往下的慢慢移动，直线上边的动态的背景区域的面积越来越大，而直线下边的待分类的图像区域的面积则越来越小。这样的话，用直线上边图像区域的平均色彩特征来度量直线下边的图像区域中的图像块，会导致越是处在直线下边图像区域中的图像块，其被计算显著分类的次数也就越多。越是处于直线上边图像区域中的图像块，其被计算显著分类的次数也就越少。

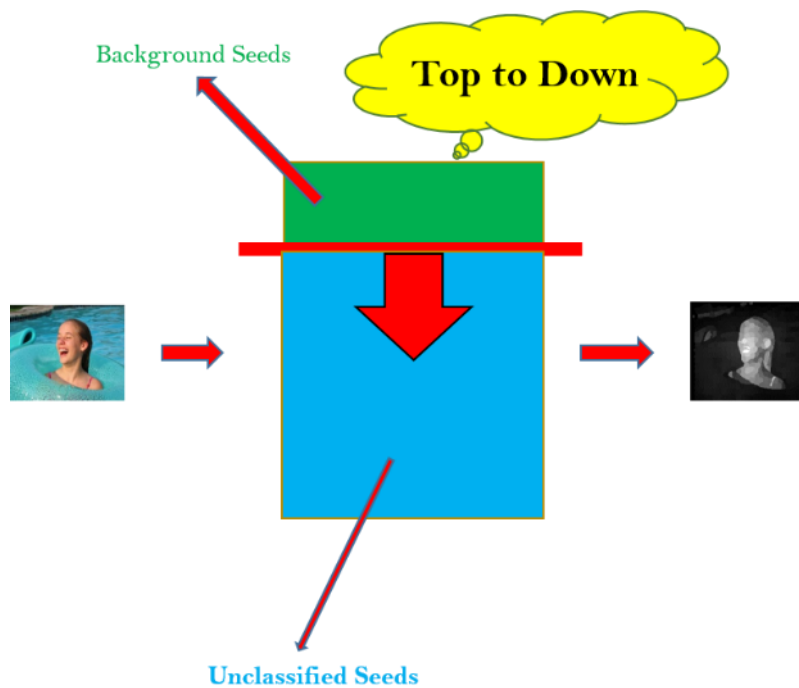


图3-3 从上至下扫描的动态背景区域

从图3-4中，由箭头的指向可以看出，当前所定义的是从下往上移动的动态背景，而且这条直线把整个图像区域从下往上划分为两个动态的图像区域。随着直线从下往上的慢慢移动，直线下边的动态的背景区域的面积越来越大，而直线上边的待分类的图像区域的面积则越来越小。这样的话，用直线下边图像区域的平均色彩特征来度量直线上边的图像区域中的图像块，会导致越是处在直线上边图像区域中的图像块，其被计算显著分类的次数也就越多。越是处于直线下边图像区域中的图像块，其被计算显著分类的次数也就越少。

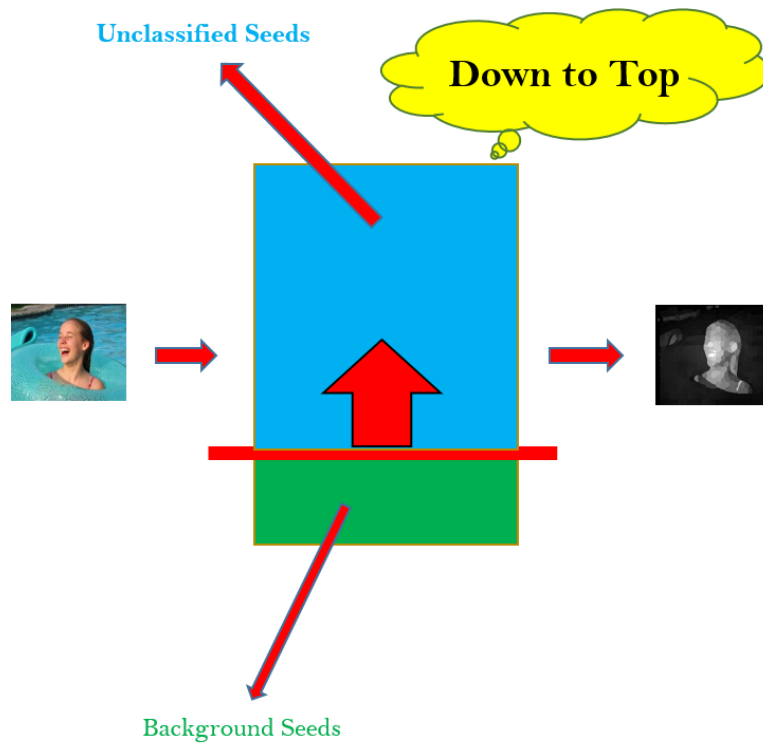


图3-4 从下至上扫描的动态背景区域

本文在动态背景的基础上定义了一个拓扑空间 T ，基于动态背景所生成的每一张显著图，本文把它看作是一个拓扑子空间 $T_i (i=1,2,3,4,...,k)$ ，也可以把 T 称作显著空间，而把 T_i 称作显著子空间。

接下来本文在动态背景的基础上定义了两种显著分类差异的计算，其中一种是基于动态背景本身色彩特征和待分类的图像区域的色彩差异对比；另一种是基于待分类区域自身的中心优先的空间差异对比；定义这两种差异对比的目的都是为了增大待分类区域中的图像块之间的显著分类差异，以获得精确度更高的图像块之间的显著分类。

为了解决MR^[35]和BSCA^[60]中采用固定背景图像区域特征所带来的显著分类失败的问题，本文算法在采用动态背景区域的基础上，提出了迭代显著分类误差和四个方向扫描的概念。图3-5为MR^[35]、BSCA^[60]的错误检测实例与本文算法检测结果对比图。

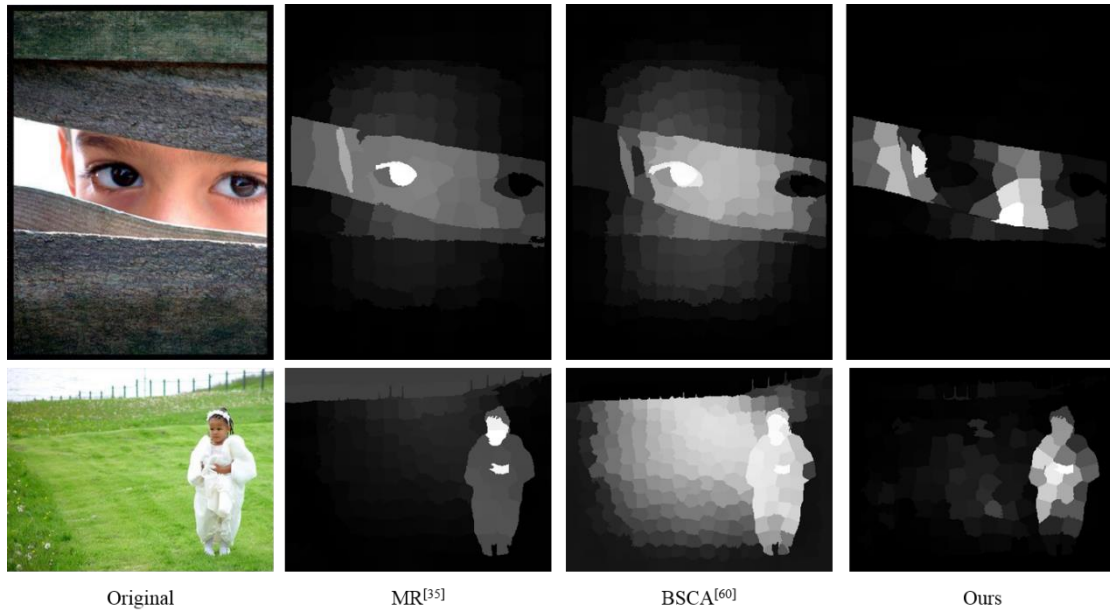


图3-5 MR、BSCA错误检测实例与本文算法检测结果对比图

3.3 本文算法的整体框架

在本章我们将介绍本文所提出的算法，该算法使用了动态背景和基于动态背景的迭代显著差异来计算图像的显著性。该算法的设计目的是在超像素的级别上进行图像的显著性识别。该算法利用了图像分割中的超像素分割，在超像素级别上分别对每一个超像素进行显著差异的度量，最后通过迭代每次的显著分类差异后作为最终的显著图。本文算法的流程图如图3-6所示：

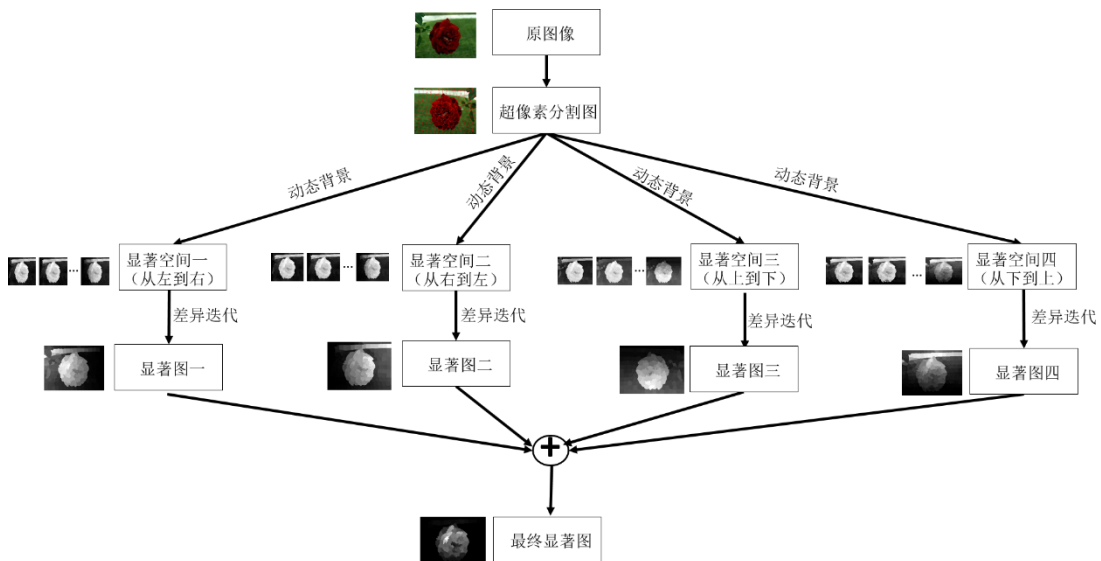


图3-6 本文算法的流程图

具体的来说, 本文的算法分为四个步骤: (1) 对图像进行超像素级别的图像分割, 并将分割的结果作为图像特征提取的基础; (2) 基于动态背景, 构造一个动态的背景空间, 并基于这个动态的背景空间生成一系列的显著图的度量空间; (3) 对每一个显著图子空间的生成, 本文利用背景的平均特征向量作为基准特征向量, 以此来度量另一待分类的图像区域中的每一个超像素块的显著程度, 具体的方法是, 对于背景空间和待分类的空间之间的显著差异关系的度量, 本文采取了基于色彩特征向量的欧氏距离度量; 对于待分类空间内部的超像素之间的关系度量, 本文采取了基于空间权重的空间差异对比。把色彩差异和空间差异之间的乘积一起作为待分类图像区域中的每一个超像素的显著差异值。(4) 最后, 本文把动态背景空间的概念分别应用到从左到右、从右到左、从上到下和从下到上四个方向上, 每一个方向上的动态扫描都会生成(3)中所说的显著图空间。由于每一个方向上每次所选取的动态背景中的超像素块的平均色彩向量的特征向量都不同, 因而在每一个方向上都能反映出不同的色彩差异, 从而可以更好的反应图像的特征, 进而提升了超像素块之间的差异程度。但在每次动态扫描的过程中, 由于动态背景空间中的超像素块是作为客观参照的, 为了给每一个超像素进行次数相对均等的显著分类, 本文从四个不同的方向上分别进行了显著分类, 最后本文采取了线性融合的方式对最终的四个方向生成的显著图进行线性融合, 并以此得到最终的单张的显著图。

3.4 区域显著性检测

3.4.1 基于动态背景的区域显著性检测

在动态背景的基础上, 我们提出了两种差异对比的计算, 一种是基于动态背景和待分类图像区域之间的颜色差异对比的计算, 我们使用动态背景在多个色彩空间上的平均颜色特征作为特征向量并使用欧氏度量来计算这种色彩差异。这种对比我们称之为基于动态背景的颜色差异对比; 另一种是基于待分类图像区域内部之间的空间差异对比的计算, 我们考虑到显著对象通常都分布在图像的中央, 根据中心优先的理论, 因此考虑对待分类的图像区域进行一种空间差异对比的计算。这个基于空间差异对比的计算结果和前面提到的基于颜色差异对比的计算结

果相累积,就可以有效的加大对待分类图像区域中的超像素块之间的对比差异的计算值,并且在MSRA-1000和MSRA-B上验证了本文的算法,实验结果表明本文提出的算法是有效的。

3.4.2 基于多种颜色特征的颜色对比差异的计算

颜色特征是一种非常有效的图像特征,也是人眼最为关注的图像底层特征之一。Borji和Itti^[56]在四个主流显著检测数据集上进行了实验。实验结果表明各种颜色空间各有各的优劣。但同时也都一致表明,使用多个色彩空间比只使用一个色彩空间结果要好。因此,本文所提出的算法中,同时使用了CIERGB、CIEL*a*b和CIEXYZ三个色彩空间,并使用这些色彩空间上的分量构造了一个复合的颜色特征,作为图像的基本视觉特征。并把这个复合颜色特征向量记为 $f_i = (r_i, g_i, b_i, l_i, a_i, b_i, x_i, y_i, z_i)$, $i = 1, 2, 3, \dots, N$. 其中 r_i, g_i, b_i 代表图像在CIERGB色彩空间上色彩分量, l_i, a_i, b_i 代表图像在CIEL*a*b色彩空间上的色彩分量, x_i, y_i, z_i 代表图像在CIEXYZ色彩空间上的色彩分量。我们基于动态背景的平均颜色特征定义了如下的颜色差异对比计算公式,

$$C_i = d \left(f_i, \sum_{j=1}^{L_{bg}} f_j \right),$$

$$i = 1: N \& i \neq j. \quad (3.1)$$

, 其中 C_i 代表的是第 i 个超像素块的显著差异值, f_i, f_j 表示的是图像的色彩特征, L_{bg} 表示的是动态背景中超像素块的总数目。 d 代表的是距离度量, N 代表的是图像分割出来的超像素块的总数目。

3.4.3 基于中心优先的空间对比差异的计算

受Judd^[57]等人提出的中心优先的启发,在本文提出的算法中,对待分类的区域中的图像块我们进行了局部的中心差异对比。这种差异对比是基于待分类的区域的动态的空间中心来进行的。但其原理上来说,还是依据显著对象最有可能分布在图像的中心的假设。

在此基础上,本文分别定义了动态的待分类图像区域的中心 l_c , 其公式为

$$d_i = \|\bar{l}_i, l_c\|,$$

$$S_i = \exp^{-d_i},$$

$$i = 1: L_N \quad (3.2)$$

其中 \bar{l}_i 代表的是待分类区域中的每一个超像素块的位置中心, d_i 代表的超像素的中心与之间的距离, S_i 代表的是第 i 个超像素块的空间差异的度量值。 L_N 代表的是待分类图像区域中超像素块的总块数。

3.4.4 两种显著分类差异的融合

首先, 基于动态的背景空间, 本文提出了颜色差异对比, 这种差异对比能够很好的对所有待分类区域的超像素块进行一个初步的显著分类。为了很好的反映待分类图像区域中的超像素块之间的空间位置关系, 本文在此基础上定义了基于局部对比的空间差异。接下来, 本文对每一个待分类图像区域中的超像素块都同时做了这两种分类, 然后把分类的结果进行累积, 这样我们便通过一个简单的计算方式同时对每一个超像素块进行了一个显著分类。实验结果证明, 同时使用两种对比差异比只是用一种差异对比的结果要好。

两种显著分类差异融合的计算公式如下,

$$\begin{aligned} D_i &= C_i \times S_i, \\ i &= 1:L \end{aligned} \quad (3.3)$$

其中 D_i 代表的是待分类区域中每一个超像素块的显著程度。 L 代表的是待分类区域中超像素的总数目。

3.5 基于动态背景的显著差异的迭代

本文首先构造了一个显著图的显著的拓扑空间 W 。其中每一张显著图代表一个拓扑子空间, 本文表示为 $W_i = (i=1, 2, 3, \dots, k)$, 其中 k 表示的是动态背景的总数目。也就是说, 在每一个动态背景上, 本文定义了一张显著图 W_i 。另外, 本文使用 d_j 表示第 j 个待分类的超像素块所代表的显著值, n 表示超像素块的总数目。那么可以得到如下的计算公式,

$$\begin{aligned} (W_i)_j &= d_j, \\ (W_{i+1})_j &= (W_i)_j + d_j, \\ j &= 1:n \& i = 1:(k-1). \end{aligned} \quad (3.4)$$

基于动态背景的显著差异的迭代可以更好的把每次所得到的显著分类误差继承了下来, 从而有效的实现了超像素块之间的显著差异的放大, 最终为实现显著目标与背景的自动分离的准备好了条件。

3.6 四个方向显著图的线性融合

本文算法在公式(3.4)的基础上得到了一张显著单个方向进行动态扫描的显著图。这张显著图能够比较好的凸显了显著的目标，同时又抑制了背景。但是，我们观察到，仅仅从一个方向进行动态扫描的话，会使得区域分割线左边的区域被显著分类的次数最少，而区域分割线右边的区域中的超像素块被显著分类的次数最多。由于本文对最终的结果采取的是迭代累加的方式，因此必须保证对所有的像素块做次数相对均等的显著分类，这样得到的结果才会比较客观。因此，本文在这个假设的前提下，对算法进行了改进，从四个方向（从左往右、从右至左、从上到下、从下到上）分别进行动态扫描，从而保证了所有的超像素块都获得次数相对均等的显著分类的机会。因而最终所得到的显著分类的结果会更加的客观。实验证明，本文的假设是成立的，同时这么做还有一个好处，在凸显潜在的显著对象所在图像区域的同时也能够比较好的抑制背景所在的图像区域。对于每一个方向所生成的一张显著图，本文目前采取的是线性融合的方式。目前使用最普遍的融合多张显著图的方法是线性加权的方式，这么做的好处是使得显著图中原本显著的地方更加显著，原本不显著的地方更加不显著，从而实现了显著对象和背景的自动分离。图3-7阐述的是本文算法中四个方向上的显著图的融合的过程。经过四个方向的融合后，我们获得了本文算法的最终显著图。

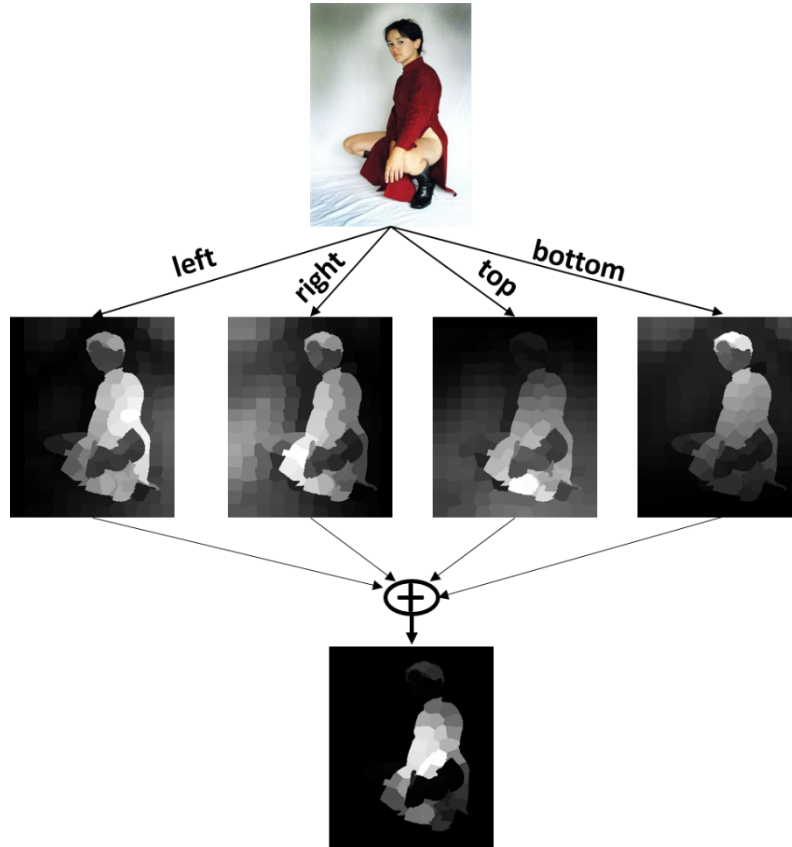


图3-7 四个方向的显著图融合

四个方向上的显著图的线性融合有助于更好的把每个方向中的显著图中的显著的部分继承下来，同时更好的抑制背景中的成分。从实验的结果中可以观察到，如果本文只进行一个方向上的动态扫描，那么会导致区域分割线左边的图像区域中的像素块的显著分类的次数更少，而区域分割线右边的图像区域中的像素块的显著分类的次数更多，从而使得最终的图像中的所有像素块的显著分类次数不均衡，因此，会使得一个方向的显著图的最终结果呈现出一部分比较亮，而其他的部分比较暗的情形。为了实现凸显图像中的显著目标，本文对四个方向上的显著图进行了线性融合，从而达到了目的。实验证明，本文算法中所提出的四个方向的显著图融合的显著图比三个方向的显著图融合的显著图、两个方向上的显著图融合的显著图或单方向上的显著图的效果都要好。最后，本文通过四个方向的显著图融合的对比如实验验证了这一点。

3.7 本章小结

由于自然图像中存在显著对象占据了图像的边缘、图像在边缘上有大面积的

背景的色彩信息和显著对象所在的图像区域的色彩信息高度一致或者图像的边缘存在一些几十个像素的黑色矩形边框,这时候基于静态背景的显著检测算法就会失效。为了解决这几个问题,本文提出了一种基于动态背景的迭代显著性检测算法。该算法首先针对静态背景中出现的问题进行了分析,在此基础上本文构造了动态的背景区域。在动态的背景图像区域之上又构造了两种显著分类计算。其中一种本文称之为基于多种颜色特征的颜色对比差异,另外一种本文称之为基于中心优先的空间对比差异。基于多种颜色特征的颜色对比差异能够很好的反映出待分类的图像区域中的超像素块之间的颜色信息的差异程度,而基于中心优先的空间对比差异能够很好的反映出待分类图像区域中的超像素块与待分类图像区域中心之间的距离关系。针对每一个动态背景区域,本文算法便对待分类图像区域中所有的像素块都进行了一次显著分类。为了使得这种显著分类的效果足够让显著对象所在图像区域自动的凸显出来,本文算法在动态背景的基础上定义了多次的显著差异的迭代,通过显著差异迭代把每次显著分类的结果都累积起来,从而最终实现显著对象与图像背景的自动分离。为了使得图像中所有的超像素块都进行相对均等的显著分类差异计算,本文在上、下、左、右四个方向上进行了动态扫描,并最终对每个方向上生成的单张显著图进行了线性融合。

第4章 实验与分析

我们在两个获得广泛认可的图像显著检测数据集MSRA-1000和MSRA-B上进行了详细的评测，实验结果表明，本文所提出的算法在定量和定性的分析上都取得了非常不错的效果。我们总共选取了7种经典的显著性检测方法和几个最近发表的显著性检测方法作为参照，和经典的显著检测方法相比，本文所提出的算法在Precision/Recall、F-measure和MAE上都取得了不错的效果。和最新的显著检测算法相对比，本文所提出的算法在Precision上也是最好。首先，我们介绍了本文所选择的两个主流显著评测数据集的基本情况；接着，我们对本文所提出的算法在两大主流数据集上的评测结果依次进行了阐述。最后，我们对实验的结果进行详细的对比实验验证和分析。

本文算法是在一台装有Windows 7 Professional的PC上完成的，该机器的硬件和软件配置如下：Intel Core i3-3110M CPU @ 2.40GHz、4GB RAM和MATLAB R2011b。

4.1 实验评测数据集

本文算法在两个获得世界各地显著性检测研究者广泛认可的图像数据集上进行了显著性检测，其中一个MSRA-1000^[26]，另一个是MSRA-B^[32]。MSRA-1000^[26]是在2009年由Achant^[26]从MSRA-B中所提取出来的一个数据集，数据集总共包含了1000张自然图像，该数据集最主要的特点是显著目标一般都位于图像的中心，并且背景区域和显著对象区域的对比比较明显，是当前被广泛使用的显著检测数据集。MSRA-B^[32]由Liu^[32]于2007年提出，Liu^[32]第一次提出了显著对象检测的概念，该图像数据集包含了5000张自然图像，而且大部分的显著目标也都是比较突出的，因而也获得业界的一致认可。图4-1是我们分别从MSRA-1000和MSRA-B中提取的部分图像实例，其中MSRA-1000图像数据集中的场景要相对简单一些。

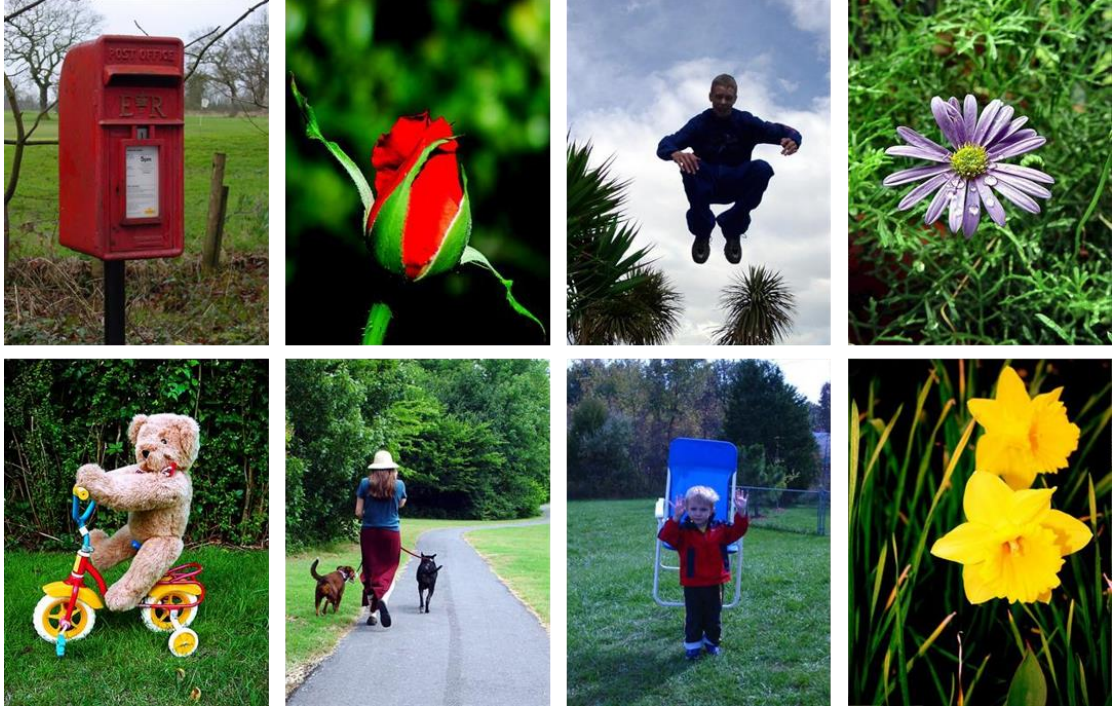


图4-1 从上到下分别为MSRA-1000、MSRA-B中的图像实例

4.2 实验结果的客观评价标准

本文采用了三种广泛认可的、标准的并且容易理解的方法来评测显著对象检测模型。其中Precision-Recall是基于人工标注对象和显著检测对象的覆盖面积来计算的。而F-measure是由Precision和Recall联合计算出来的。我们同时也介绍了MAE，MAE直接由的显著图和人工标注的真值图之间的平均绝对误差计算得到。我们用一张归一化到 $[0,255]$ 的灰度图来代表显著图， G 代表显著对象的二值标记的真值图。对于一张二值图，我们用 $|\cdot|$ 代表二值图中的非零部分。

4.2.1 Precision和Recall曲线

对于一张显著图 S 所对应的二值图 M ，我们通过计算Precision和Recall曲线来比较 M 和真值 G 之间的关系，其中

$$Precision = \frac{|M \cap G|}{M}, Recall = \frac{|M \cap G|}{G} \quad (4.1)$$

从Precision和Recall的计算公式中，我们可以看出显著图 S 所对应的二值图是评价中一个关键的计算步骤。一般来说，我们有两种流行的方式来做二值化。第

一种方式, Achanta^[26]提出一种和图像独立的自适应阈值的方法来二值化显著图, 这个自适应的阈值被定义为显著图S的平均显著值的二倍。

$$T_a = \frac{2}{W \times H} \sum_{x=1}^W \sum_{y=1}^H S(x, y), \quad (4.2)$$

在公式(4.2)中, W和H分别代表显著图的宽度和高度。第二种方式, 我们使用一个固定的从0到255变化的固定阈值来二值化显著图。在每一个阈值, 一对(Precision和Recall)值被计算了出来, 这两个值最终绘成一条PR曲线来反映显著检测模型在不同情形下的性能。

4.2.2 F-measure曲线

通常情况下, 和Recall曲线都不能很好的Precision反映显著图的质量。因此, 我们对Precision和Recall做了一个加权的调和平均, 这里面使用到了一个非负的权重 β^2 ,

$$F_\beta = \frac{(1 + \beta^2) \text{Precision} \times \text{Recall}}{\beta^2 \text{Precision} + \text{Recall}}. \quad (4.3)$$

在许多的显著对象检测的文章中(比如Achanta^[26]), β^2 通常都被设置为来提高Precision的权重值。Liu^[32]提到, 提高的Precision原因是Recall没有Precision更重要。我们可以很容易得到100%的Recall, 比如说, 我们可以把整个图像区域都设置为前景。

根据二值化显著图方式的不同, F-measure的计算方式也相应的不一样。当我们用自适应的阈值来进行二值化的时候, 我们由每张图像得到了单个的 F_β 值, 最终的F-measure是所有 F_β 的平均值。如果我们用一条PR曲线来计算F-measure的话, 那么就要对每一对Precision和Recall分别计算F-measure, 然后再对F-measure求平均值。在公式(4.3)中, 我们从F-measure的计算公式可以看出, F-measure是Precision和Recall的加权调和平均, 因此也和Precision、Recall一样都在 $[0, 1]$ 的范围之内取值。

4.2.3 MAE曲线

基于覆盖的评测方法(如Precision、Recall和F-measure)都没有考虑到正确标注的像素却被认定为不显著的情形。这样的话, 就支持了一些把很高的显著值赋给显著的区域但同时不对不显著区域判定失效的方法。而且, 在一些应用场景中

Avidan^[59], 加权、连续的显著图的质量可能比二值的真值图更重要。为了得到一个更综合的评价, 通常情况下都要求评价连续的显著图 S 和二值的真值图 G 的一个平均绝对误差 (MAE), 显著图 S 和真值图 G 都是归一化到 $[0,1]$ 的范围内。MAE的定义如下,

$$MAE = \frac{1}{W \times H} \sum_{x=1}^W \sum_{y=1}^H \|S(x, y) - G(x, y)\|. \quad (4.4)$$

4.3 实验结果

我们将本文算法与7种经典的和最近几年发表的显著性检测方法进行了对比实验。它们分别是对IT^[21]、MZ^[22]、GB^[24]、SR^[23]、AC^[25]、FT^[26]、RC^[28]、LRMR^[30]、GS^[33]、SF^[29]、MR^[35]、BSCA^[60]评测的PR曲线、F-measure曲线和MAE曲线, 我们采取的都是固定的阈值。以1为步长, 我们从 $[0, 255]$ 的范围中选取了256个阈值。然后用这些阈值给显著图进行二值化, 那么每张显著图都会生成256张二值的图像。这样的话, 我们就可以对每一张二值图像进行评价。最后, 在完成所有的二值图像的评价指标的计算后, 我们对二值图像按照阈值归类, 这样的话共有256类, 最后计算每一类的评价指标的平均值。以下是本文算法和其它算法的实验结果对比图。

4.3.1 本文算法和其它算法显著检测结果的对比

图4-2表示的是本文提出的算法与经典的显著检测方法检测结果的对比。图4-3所示的是本文算法和几个最近几年的显著检测方法的显著检测结果的对比。从图4-2中, 我们可以看出, 本文算法所产生的显著检测结果明显优于经典的显著检测算法的检测结果。并且本文算法所产生的结果对背景的抑制效果非常好, 并且同时凸显了图像中的显著对象所在的图像区域。

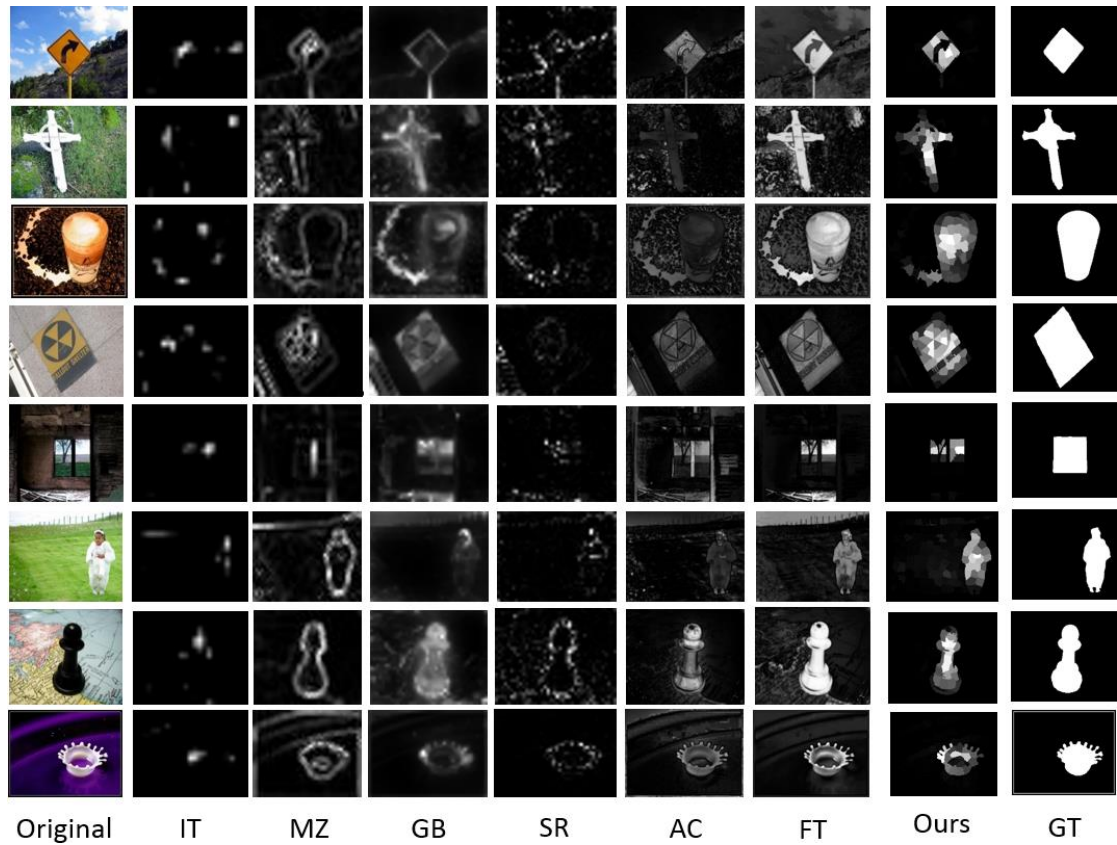


图4-2 本文算法与经典显著检测算法检测结果对比图

从图4-3中，我们可以看出，和最近几年提出的最新的显著检测方法对比，本文算法的检测结果也是比较好的。比如，从图4-3中的第六张图像的显著检测结果本文发现，现有的优秀的显著检测方法只有少数几个算法取得了不错的显著检测效果。而采用固定的静态背景区域的显著检测算法MR^[35]、BSCA^[60]都出现了大面积的误检测的情形，经过我缜密的分析和细致的判断之后，本文发现原来是背景节点的选取有误所致。因而在接下来的过程中，在设计本文算法的时候，本文就考虑到了这一点，这是由于选取固定的超像素块作为背景种子节点所导致的错误的显著检测情形，因此本文提出了基于动态背景的显著检测方法。在BSCA^[60]中，本文发现图4-3中的第六张图像出现了大面积的高亮的现象，而在BSCA^[60]中，该方法也是假设边界的超像素为背景种子节点，并依据元胞自动机来对局部超像素的显著检测结果进行差异进化，而对于图4-3中的第六张图而言，显著对象所在的超像素本身的颜色特征和边界所代表的超像素的颜色特征极为类似，因而也出现了严重的误检测情形。因此，本文算法在这一点上具有先进性，本文成功的解决了这个问题，动态背景的概念使得本文算法在检测这一种类型的自然图像的时候，不会出现MR^[35]、BSCA^[60]中所出现的情形。

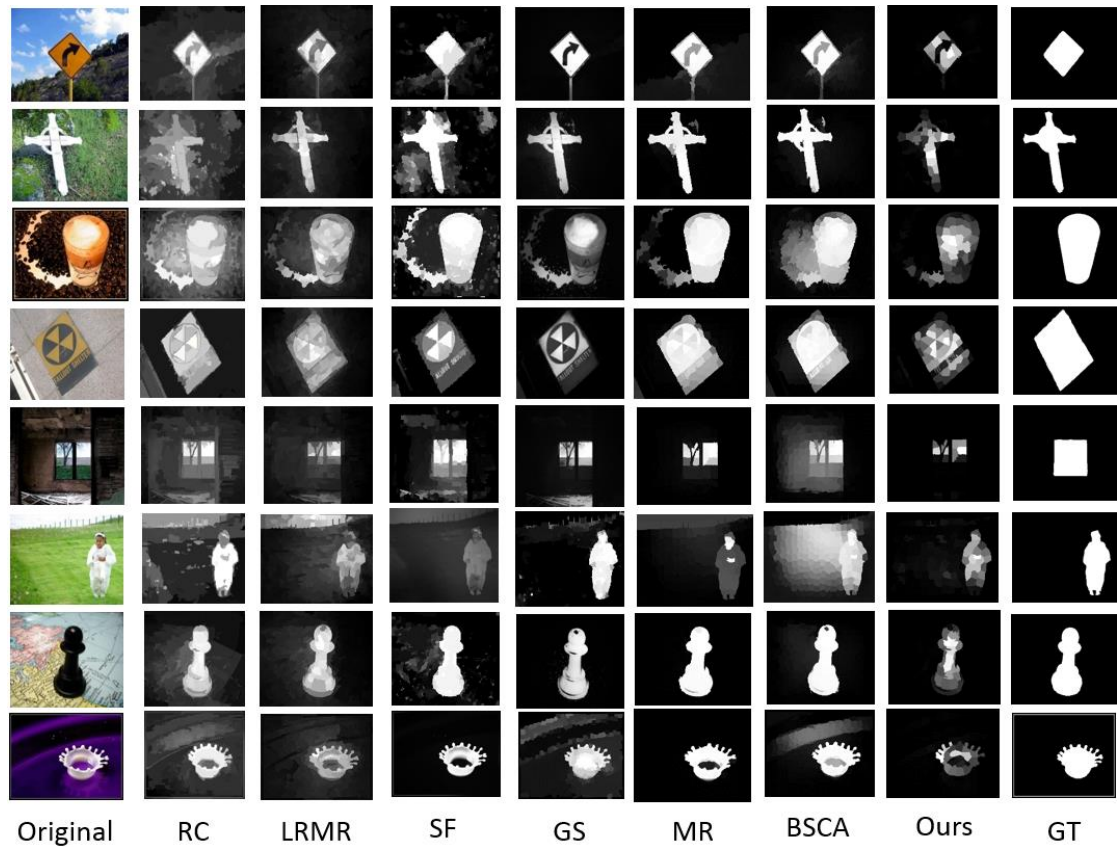


图4-3 本文算法和几个最近发表的显著检测方法的检测结果对比

4.3.2 本文算法和其它算法F-measure曲线的对比

从图4-4中可以看出，本文所提出的算法在F-measure曲线的评测上，本文算法的Precision比其它算法都要好。Liu^[32]提出，在显著性检测中，最重要的指标就是Precision，Precision比Recall和F-measure都要重要。他进一步指出，Recall曲线对于显著检测的评价意义并不大。举例来说，我们可以把整张图像都检测为前景，从而得到100%的Recall值。因此，一个获得高Recall值的算法倾向于选择大块的显著区域，从而牺牲了Precision值。显著对象检测的一个最主要的目的就是要准确的定位显著对象所在的位置，而且准确度越高越好。然而，对于包含一大块显著对象的图像来说，很高的精确度通常很难达到。也就是说，对于一张显著对象占据了整张图像面积80%的图像来说，我们通过简单的选取整张图像都作为显著对象可以获得80%的Precision和100%的Recall。因此，对于显著对象检测真正的挑战是如何在小的显著目标上获得很高的准确度（Precision）。

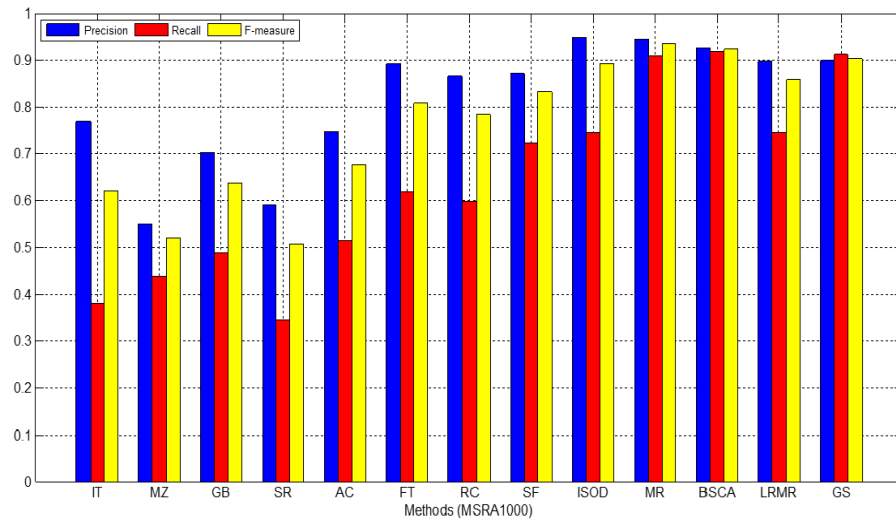


图4-4 本文算法（ISOD）和其它算法的F-measure曲线对比图

4.3.3 本文算法和其它算法的MAE曲线的对比

从本文算法那和各种对比方法的MAE（Mean Absolute Error）对比曲线中可以看出，本文所提出的算法在MAE的评测上，本文算法的MAE比经典的显著检测算法都要好。图4-5为本文算法和其他算法在MSRA-1000上的MAE对比曲线图。图4-6为本文算法和其它显著检测算法在MSRA-B上的MAE对比曲线图。

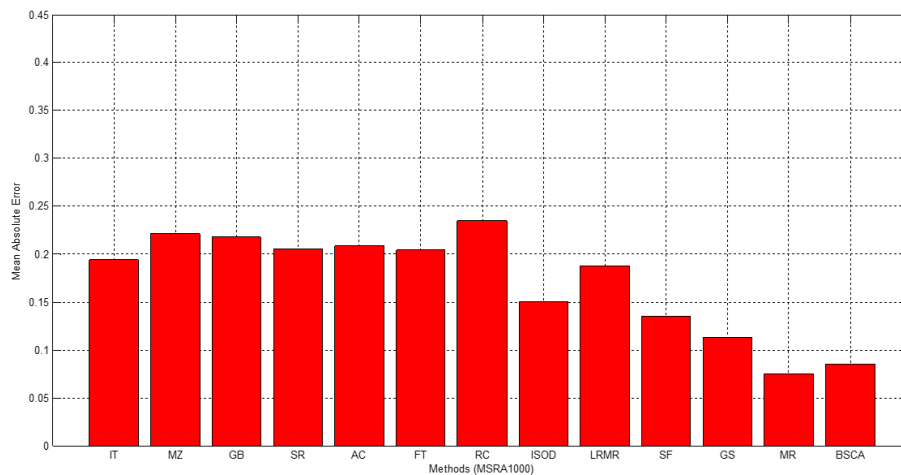


图4-5 本文算法与其他算法在MSRA-1000上的MAE对比曲线图

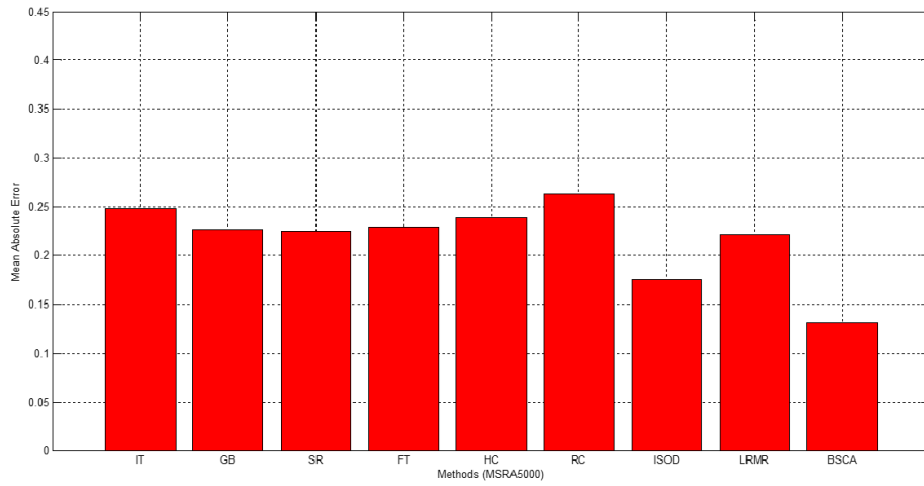


图4-6为本文算法与其他算法在MSRA-B上的MAE对比曲线图

4.3.4 本文算法和其它算法的PR曲线的对比

图4-7展示了本文算法在MSRA-B上与经典显著检测方法的检测效果PR曲线对比结果。实验表明，本文算法具有一定的先进性，在准确度上一致超越了所有经典的显著检测方法，并且本文所提出的方法具有简单、高效的特点。

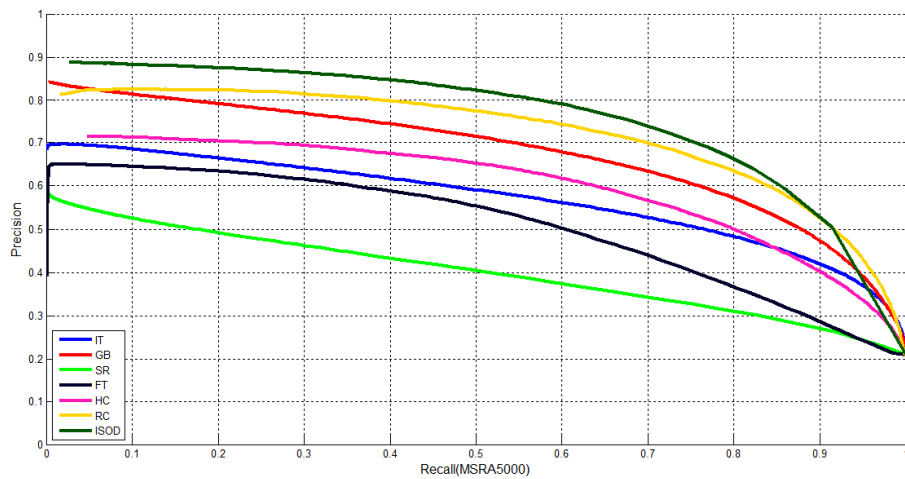


图4-7 本文算法在MSRA-B上与经典方法显著检测结果的PR对比曲线图

4.4 实验结果分析

4.4.1 动态背景迭代的对比实验结果分析

我们发现，在动态背景的选取中，采取不同的大小的动态背景可以获得不同的显著检测效果。当动态背景的大小为图像大小0.25倍时，本文算法取得了最佳的Precision值。因此，本文算法最终采取了这种策略。图4-8为在不同大小的动态背景下，本文算法在MSRA-1000上的显著检测结果的对比曲线图。

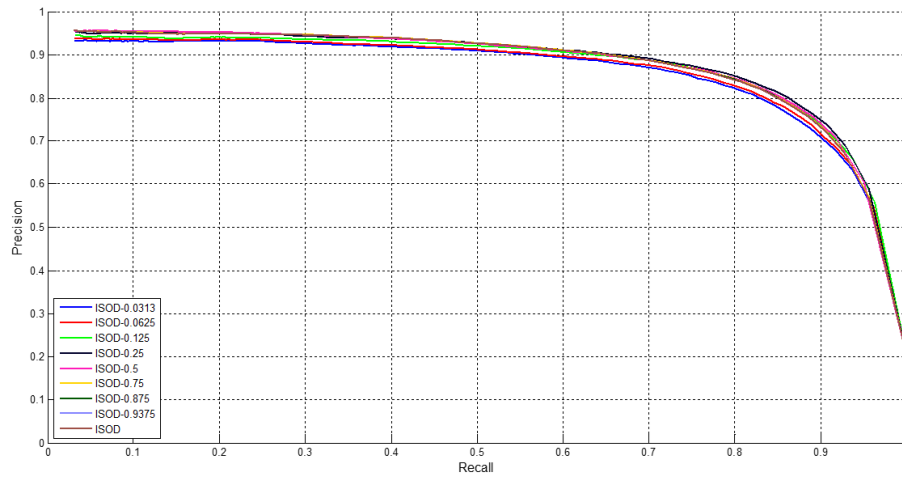


图4-8 不同大小的动态背景下，本文算法的显著检测结果对比图

图4-9为图4-8的显著检测结果的曲线在右上角的局部放大图。

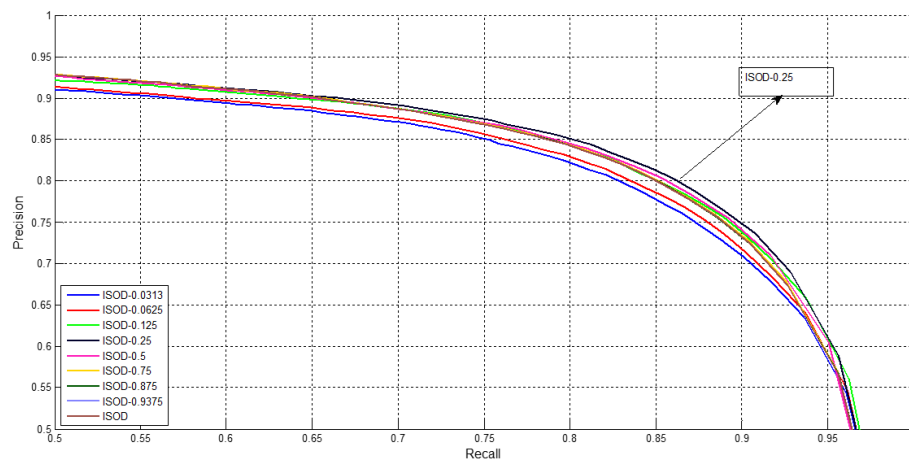


图4-9 局部放大图

4.4.2 空间的中心优先的对比实验结果分析

Tilke Judd^[57]首次提出了“人们在使用数码相机拍照的时候，他们通常都会把自己感兴趣的对象定格在镜头的中央，因而显著的对象通常都会分布在图像的中央位置。”从图4-10中，我们可以看出，中心优先的假设对图像的显著检测结果具有一定的促进作用，也就是说，显著的中心优先的理论从某种程度上来说，是有一定的道理的。因而本文算法也采纳了这种策略。对比实验证明，该策略对于MSRA-1000来说完全成立。

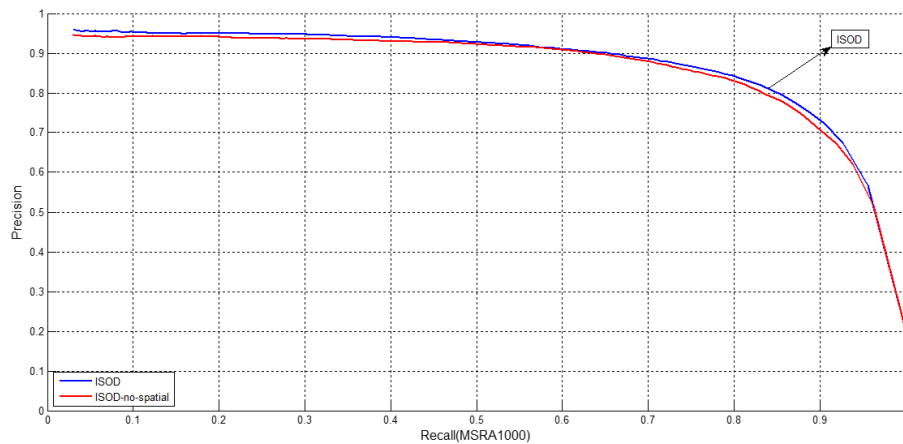


图4-10 空间的中心优先的对比实验结果分析

4.4.3 四个方向的对比实验结果分析

实验表明，本文算法采取的四个方向的显著图融合的策略是非常有效的。当我们只从一个方向进行显著检测的时候，我们发现图像显著检测的结果没有同时融合两个或两个以上的方向上的图像显著检测的结果的效果好。因此根据对比实验的结果，本文算法最终采取了同时对四个方向的显著检测结果进行融合的策略。图4-11为本文算法在MSRA-1000上的四个方向上的显著图融合对比实验结果。

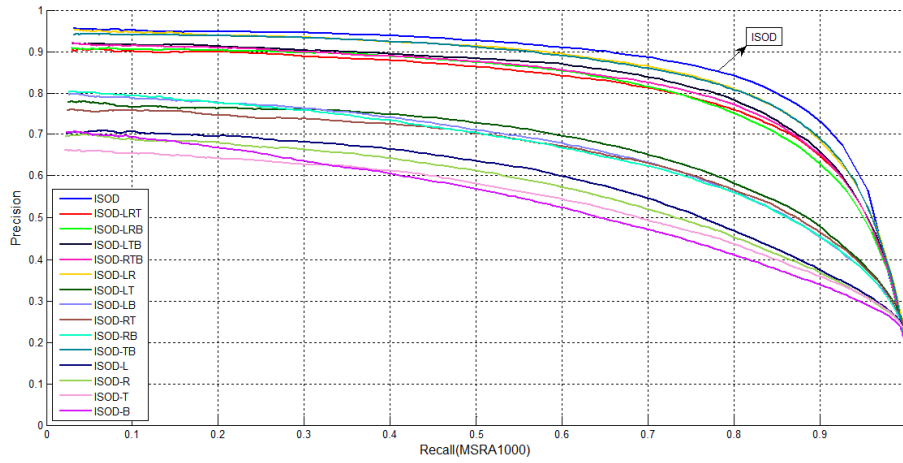


图4-11 本文算法在四个方向进行显著图融合的对比实验结果的曲线图

4.5 本章小结

本章对本文提出的算法的PR、F-measure、MAE在两个数据集MSRA1000和MSRA-B上都进行了详细的对比实验。本文算法在准确率上一致超过了经典的显著检测算法和几个最近几年发表的优秀的显著检测算法。而由Liu^[32]提出，图像的显著目标检测中最重要的一个指标就是显著检测的准确率。因此从性能评测的角度来说，本文算法具有一定的先进性。另外，本文对提出的两个创新点都分别进行了对比实验分析。其中一个是动态背景大小的选取实验，另一个是四个方向上的显著图的融合实验。由动态背景大小选取的实验本文发现，动态背景的大小也不是选的越大越好，而是大小为图像的0.25倍至0.5时，能够取得几乎差不多的检测效果。但在0.25倍图像大小时，显著差异迭代的次数最少，因而对图像进行显著检测的速度最快，而且在准确率上和0.5倍图像大小时很接近。由四个方向上的显著图融合的对比实验，本文发现只有四个方向上的显著图同时进行线性融合，才会取得最佳的显著检测效果。最后，对本文中所提出的基于空间的中心优先本文也做了相应的对比实验，实验发现，采用中心优先比不采用中心优先的效果要好一些，但不是太明显。综上，本章所提出的基于动态背景的迭代显著检测算法是有效的，并且是高效的。

第5章 总结和展望

5.1 总结

我们生活在一个智能化设备大爆炸的时代，而图像作为我们人类最直观的获取信息的方式，随着数码相机和智能手机等智能图像输出设备的普及，如何让计算机能快速的从海量的图像数据中提取我们人类最感兴趣的信息是当前计算机视觉研究领域的一个亟需解决的问题。图像的显著性检测研究受到人类视觉的选择性注意机制的启发，其目的是从图像中提取人眼最感兴趣的目标图像区域，并输出一张被称为显著图的灰度图像。这样我们便可以把有限的计算资源分配在最显著的图像区域，进而进行后续的图像处理操作。

本文在第一章介绍了显著性检测研究的研究背景及意义、显著性检测研究的特点、国内外研究现状、本文主要研究内容和课题内容安排及章节关系。第二章首先介绍了视觉性显著性的理论基础，包括图像的底层特征（颜色特征、亮度特征、纹理特征和形状特征）、图像分割算法以及几种经典的图像显著性检测模型（IT算法、SR算法和RC算法）。接着介绍了本文算法所使用到的一些相关的算法以及图像特征，其中包括基于超像素的图像分割、色彩空间模型（CIERGB色彩空间、CIEL*a*b色彩空间以及CIEXYZ色彩空间）和显著对象检测所使用到的计算框架（包括显著的中心优先、显著的背景优先、全局对比以及局部对比）。第三章，本文介绍了基于动态背景的迭代显著检测算法，该部分包括引言、动态背景的概念、本文算法的主要计算步骤、区域显著检测（其中包括：基于动态背景的区域显著性检测、基于多种颜色特征的颜色对比差异的计算、基于中心优先的空间对比差异的计算以及显著差异的融合）、基于动态背景的显著差异的迭代以及四个方向的显著图的线性融合。第四章是本文算法的实验部分。我们首先介绍了实验所使用到的评测数据集（包括MSRA-1000和MSRA-B）。然后，我们介绍了实验所使用到的相关显著检测评测标准，其中包括（Precision和Recall曲线、F-measure曲线以及MAE曲线）。最后，对本文算法和其它显著检测算法的结果以及评测曲线都进了详尽的分析。

实验结果表明，首先，本文中所提出的“基于动态背景的迭代显著性检测算

法”是有效的，而且也在Precision/Recall、F-measure和MAE上一致超越了经典的显著检测方法。最重要的是，本文所提出的算法是一种全新的计算图像显著性的计算框架，而不是基于以往的图像显著检测方法所做的优化或其他改进。其次，本文算法中所提出来的两个重要创新点（动态背景区域和迭代显著差异）是不同于以往任何的显著检测方法的显著计算理论，而且也获得了比较好的性能评价，所以本文算法在研究的方法论上具有一定的先进性。最后，本文算法没有做任何优化，是一个计算图像显著性的原生算法，并且本文算法所使用到的计算手段也是非常的简单和高效。

5.2 展望

本文所提出的算法在图像显著性的计算上取得了超越经典显著检测方法性能的效果，这同时也凸显了本文所提出的新的显著计算框架的新颖性和有效性。但本文算法所基于的动态背景和迭代显著差异是一种新的计算图像显著性的计算理念，因此，我认为至少还可以在以下几个方面进行后续更深入的研究。

第一，本文算法所获得的显著检测结果都是原始的灰度图，并没有经过任何的优化。而且，在迭代的过程中，每次所选取的动态背景下所获得的对应的显著图，也没有经过相应的优化，因此，在后续实验中，我会考虑对本文算法的显著检测的结果进行相关的优化操作。

第二，本文算法中所使用的显著差异度量方法也是非常的简单，我们仅仅使用一个最简单欧氏度量来度量图像块之间的色彩差异程度。因此，在后续的研究中，我会考虑去寻找更好的度量图像的色彩块之间的差异的方法，那么，实验的结果也会获得相应的提升。

第三，本文算法所使用的动态背景的概念，是否可以应用在现有的显著检测方法之上，这个也是未来的研究方向之一。

第四，本文算法中使用到了一个显著对象的中心优先的假设，并做出了相应的对比实验图，实验发现在性能上确实有一些提升，但不是很明显。因此，在后续实验中，我会考虑对空间对比的度量进行重大的改进，这也是未来的研究方向之一。

最后，基于动态背景可以极大提高显著检测的精度，而且不需要对输入图像

的类型进行限制。但同时也带来一些问题，比如说，由于迭代的次数过多而带来的性能下降的问题，本文算法目前的处理速度为2秒钟处理完一张 300×400 像素的自然图像。目前的手机等智能设备的处理器的处理速度的平均水平一般为1.5GHz左右，用本文的算法处理一张 300×400 的自然图像的话，就需要3秒多。因此，如果未来本文所提出的算法要达到实用的阶段的话，如何在精确度和性能之间找到一个最佳的平衡点，这个是一个非常值得考虑的问题。

参 考 文 献

- [1] Sohlberg M.M., Mateer C.A. Introduction to cognitive rehabilitation: Theory and Practice[M]. New York: Guilford Press, 2014,56(1):17-27.
- [2] Raichle M.E. The brain's dark energy[J]. Scientific American Magazine, 2010,302(3):1249-1250.
- [3] Shepherd G.M. The Synaptic Organization of the Brains, 3th[M].New York:Oxford University Press, 1990,3(1):145-162.
- [4] Koch C. Biophysics of Computation: Information Processing in Single Neurons, 1st[M]. New York: Oxford University Press,2004,37(10):1213-1214.
- [5] Elazary L, Itti L. Interesting objects are visually salient[J]. Journal of Vision,2008, 8(3),3-15.
- [6] Avidan S., Shamir A. Seam carving for content-aware image resizing[J]. ACM SIGGRAPH,2007, 26(3):10.
- [7] Li Y, Tian Y, Yang J, Duan L, Gao W. Video retargeting with multi-scale trajectory optimization[C].Proceedings of the International Conference on Multimedia Information Retrieval, 2010:45-54.
- [8] Gao W, Tian Y, Huang T, Yang Q. Vlogging: A survey of videoblogging technology on the web[C]. ACM Computing Surveys,2010, 42(4):2623-2627.
- [9] Fu H, Chi Z, Feng D. Attention-driven image interpretation with application to image retrieval[J]. Pattern Recognition,2006,39(9):1604-1621.
- [10]Li S, Lee M. Efficient spatiotemporal-attention-driven shot matching[C]. Proceedings of the 15th Annual ACM International Conference on Multimedia[C],2007:178-187.
- [11]Ma Y, Hua X, Lu L, Zhang H. A generic framework of user attention Model and its application in video summarization[J]. IEEE Transactions on Multimedia,2005,7(5):907-919.
- [12]Li Z, Qin S, Itti L. Visual attention guided bit allocation in video compression[C].Image Vision Computing,2010,29(1):1-14.
- [13]Zhu G, Zheng Y, Doermann D, Jaeger S. Multi-scale structural saliency for signature detection[C]. Preceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2007:1-8.
- [14]Mahadevan V, Vasconcelos N. Background subtraction in highly dynamic scenes. Preceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2008:1-6.
- [15]Chalmond B, Francesconi B, Herbin S. Using hidden scale for salient object Detection[J]. IEEE Transactions on Image Processing,2006,15(9):2644-2656.
- [16]Li M, Clark J. Selective attention in the learning of invariant representation of Objects[C]. Preceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2005:93-95.
- [17]Siagian C, Itti L. Rapid biologically-inspired scene classification using features shared with visual attention[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2007,29(2):300-312.
- [18]Yang R, Tian Y, Huang T. Dct-based videoprinting on saliency-consistent regions for detecting video copies with text insertion[C]. Advances in Multimedia Information,2009,5879:797-806.

- [19] Triesman A, Gelade G. A feature-integration theory of attention[J]. *Cognitive Psychology*, 1980, 12(1):97–136.
- [20] Koch C, Ullman S. Shifts in selective visual attention: towards the underlying neural circuitry[J]. *Human Neurobiology*, 1985, 4:219–227.
- [21] Itti L, Koch C, Niebur E. A model of saliency-based visual attention for rapid scene analysis[J]. *IEEE Transactions on Pattern Recognition and Machine Intelligence (TPAMI)*, 1998, 20(11):1254–1259.
- [22] Ma Y, Zhang H. Contrast-based image attention analysis by using fuzzy growing[C]. *ACM International Conference on Multimedia*, 2003:374–381.
- [23] Hou X, Zhang L. Saliency detection: a spectral residual approach[C]. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2007:1–8.
- [24] Harel J, Koch C, Perona P. Graph-based visual saliency. *Advances in Neural Information Processing Systems*, 2007, 19:545–552.
- [25] Achanta R, Estrada F. Salient region detection and segmentation[C]. *Computer Vision Systems*, 2008:66–75.
- [26] Achanta R, Hemami S, Estrada F. Frequency-tuned salient region detection[C]. *CVPR*, 2009, 22(9–10):1597–1604.
- [27] Goferman S, Zelnik-Manor L. Context-aware saliency detection[J]. *CVPR*, 2010:2376–2383.
- [28] Cheng M, Mitra N, Huang X, Torr P, Hu S. Global contrast based salient region detection[J]. *TPAMI*, 2015, 37(3):409–416.
- [29] Perazzi F, Krahenbuhl P, Pritch Y, Hornung A. Saliency filters: Contrast based filtering for salient region detection[C]. *CVPR*, 2012:733–740.
- [30] Shen X, Wu Y. A unified approach to salient object detection via low rank matrix recovery[C]. *CVPR*, 2012, 23(10):853–860.
- [31] 张鹏, 王润生. 静态图像中的感兴趣区域检测技术[J]. *中国图象图形学报: A 辑*, 2005, 10(2):142–148.
- [32] Liu T, Sun J, Zheng N, Tang X, Shum H. Learning to detect a salient object[C]. *CVPR*, 2007, 33(2):1–8.
- [33] Wei Y, Wen F, Zhu W, Sun J. Geodesic saliency using background priors[C]. *ECCV*, 2012, 7574(1):29–42.
- [34] Xie Y, Lu H. Visual saliency detection based on Bayesian model[C]. *ICIP*, 2011:653–656.
- [35] Yang C, Zhang L, Lu H, Ruan X, Yang M. Saliency detection via graph-based manifold ranking[C]. *CVPR*, 2013, 9(4):3166–3173.
- [36] Ojala T, Pietikäinen M, Harwood D. Performance evaluation of texture measures with classification based on Kullback discrimination of distributions[C]. *ICPR*, 1994, 1:582–585.
- [37] Ojala T, Pietikäinen M, Harwood D. A Comparative Study of Texture Measures with Classification Based on Feature Distributions[J]. *Pattern Recognition*, 1996, 29:51–59.
- [38] Wang X, Han X, Yan S. An HOG-LBP Human Detector with Partial Occlusion Handling[J]. *ICCV*, 2009, 30(2):32–39.
- [39] Vincent L, Soille P. Watersheds in digital spaces: An efficient algorithm based on immersion simulations[J]. *TPAMI*, 1991, 13(6):583–598.
- [40] Comaniciu D, Meer P. Mean shift: a robust approach toward feature space analysis[J]. *TPAMI*, 2002, 24(5):603–619.
- [41] Vedaldi A, Soatto S. Quick shift and kernel methods for mode seeking[C]. *ECCV*, 2008:705–718.

- [42]Levinshtein A, Stere A, Kutulakos K, Fleet D., Dickinson S, Siddiqi K. Turbopixels: Fast superpixels using geometric flows[J]. TPAMI, 2009, 31(12): 2290-2297.
- [43]Achanta R, Shaji A, Smith K, Lucchi A, Fua P, Susstrunk S. SLIC Superpixels Compared to State-of-the-Art Superpixel Methods[C], TPAMI, 2012, 34(11): 2274-2282.
- [44]Bergh M, Boix X, Roig G, Capitani B, Gool L. SEEDS: Superpixels Extracted via Energy-Driven Sampling[J], IJCV, 2013, 111(3): 298-314.
- [45]Shi J, Malik J. Normalized cuts and image segmentation[J]. TPAMI, 2000, 22(8): 888-905.
- [46]Felzenszwalb P, Huttenlocher D. Efficient graph-based image segmentation[J]. IJCV, 2004, 59(2): 167-181.
- [47]Moore A, Prince S, Warrell J, Mohammed U, Jones G. Superpixel Lattices[C]. CVPR, 2011: 1-8.
- [48]Veksler O, Boykov Y, Mehrani P. Superpixels and supervoxels in an energy optimization framework[C]. ECCV, 2010, 6315: 211-224.
- [49]Liu M, Tuzel O, Ramalingam S, Chellappa R. Entropy Rate Superpixel Segmentation[C]. CVPR, 2011, 32(14): 2097-2104.
- [50]Zhang Y, Hartley R, Mashford J, Burn S. Superpixels via Pseudo-Boolean Optimization[C]. ICCV, 2011, 23(5): 1387-1394.
- [51]Rother C, Kolmogorov V, Blake A. GrabCut: Interactive foreground extraction using iterated graph cuts[J]. ACM TOG, 2004, 23(3): 309-314.
- [52]Ren X, Malik J. Learning a classification model for segmentation[C]. ICCV, 2003, 1: 10-17.
- [53]Achanta R, Shaji A, Smith K, Lucchi A, Fua P. Slic superpixels compared to state-of-the-art superpixel methods[J]. TPAMI, 2012, 34(11): 2274-2282.
- [54]Wright W. A re-determination of the trichromatic coefficients of the spectral colors[J]. Transaction of the Optical Society, 2002, 30(4): 141-164.
- [55]Guild J. The colorimetric properties of the spectrum[J]. Philosophical Transactions of the Royal Society of London, 1932, 230: 149-187.
- [56]Borji A, Itti L. Exploiting local and global patch rarities for saliency detection[C]. CVPR, 2012: 478-485.
- [57]Judd T, Ehinger K, Durand F, Torralba A. Learning to predict where humans look[C]. ICCV, 2009, 30(2): 2106-2113.
- [58]Zhai Y, Shah M. Visual attention detection in video sequences using spatiotemporal cues[J]. Proceedings of ACM Multimedia, 2006, 177(1): 478-482.
- [59]Avidan S, Shamir A. Seam carving for content-aware image resizing[J]. ACM TOG, 2007, 26(3): 10.
- [60]Qin Y, Lu H, Xu Y, Wang H. Saliency detection via cellular automata[C]. CVPR, 2015: 110-119.
- [61]X Li, H Lu, L Zhang, X Ruan, Saliency detection via dense and sparse reconstruction[C]. ICCV, 2013: 2976-2983.

致 谢

时光荏苒，岁月如梭，写到这里，只有这八个字久久的回荡在我的脑海里。

三年过去了，恍惚变了个人，三年的研究生活即充满快乐又充满失望，在紧张忙碌中度过的三年。三年的时间里，我学到了很多的专业知识，也增长了人生阅历，对于我来说，我想这三年是非常值得怀念的三年。

首先，我要感谢我的父母。是你们把我含辛茹苦的养大，十年如一日，从牙牙学语到远走他乡，每一程都有你们的陪伴，都有你们带给我的温暖，此刻，我只想说，我是幸福的，也是幸运的，幸福的是我有这么支持我的父母，幸运的是我生在了一个这么支持我的大家庭里面，这一切的一切使得我可以有更少的包袱去生活，去闯荡未知的世界，因为我知道，即使我失败了，我还有个这么支持我的家。

其次，我衷心感谢我的导师雷海军老师。雷老师在生活、学习和科研上都给了我悉心的指导，更在待人处事上为我树立了榜样。雷老师提出了要从事显著性这个方向的研究工作，并把方文周师兄的研究成果共享给大家参考。在我的论文的完成的过程中，雷老师都花了很多的时间给予我帮助，并提出了一些建议。雷老师时常的嘘寒问暖令我倍感温馨。在此我想向他表示由衷的感谢和祝福。

再次，衷心的感谢杨焯老师给我在生活和论文方面的指导。杨焯老师几次指出了我论文中存在的问题，耐心地和我讨论文中存在的不足并提出了详细的改进方案，同时也首次确立了本论文中基于动态背景的主题。正是杨老师的热心帮助，才使得我的论文的主题更明晰，更具有说服力。

最后，衷心的感谢罗秋明老师对本文在写作方面的指导，罗老师提供了宝贵的建议使得本文的主题更突出，更有说服力。

感谢实验室的老师同学们三年来的指导和帮助，三人行必有我师，广泛的交流使得我更懂得包容，更懂得去欣赏别人的长处，同时也使得我更了解自己的短处。

最后，感谢深圳大学为我提供这么好的学习、生活和科研的环境，深圳大学是一所包容、开放的大学，我为我选择深圳大学而感到骄傲和自豪。

攻读硕士学位期间的研究成果

- [1] Fangjun Wen, Haijun Lei, Qiumin Luo, and Xuan Yang, Iterative Saliency via Dynamic Image Partitioning, Signal Processing Letters, 2016.(拟投)