

显著区域检测技术研究

梁 晔^{1,2} 于 剑² 郎丛妍² 刘宏哲¹

(北京联合大学北京市信息服务工程重点实验室 北京 100101)¹

(北京交通大学计算机与信息技术学院 北京 100044)²

摘 要 显著区域检测是计算机视觉中非常活跃的研究方向,其应用领域极为广泛。如何快速准确地找到图像的显著区域尚未形成完整的理论体系,且与具体应用密切相关,对研究人员来说仍是一个富有挑战的课题。对显著区域检测技术进行了综述。首先深入讨论了自底向上和自顶向下的显著区域检测方法,对方法进行了归类,并对典型方法进行了梳理;其次讨论了算法的评价标准和目前流行的显著性评测数据库;最后对目前存在的问题进行了总结,给出了未来的研究方向。

关键词 显著性,视觉注意,显著区域检测,自顶向下,自底向上

中图法分类号 TP391

文献标识码 A

DOI 10.11896/j.issn.1002-137X.2016.3.005

Research on Salient Region Extraction Technology

LIANG Ye^{1,2} YU Jian² LANG Cong-yan² LIU Hong-zhe¹

(Beijing Key Laboratory of Information Service Engineering, Beijing Union University, Beijing 100101, China)¹

(School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China)²

Abstract Salient region detection technology is a very active research area and is applied extensively. How to find salient region of the image quickly and accurately has not yet formed a complete theoretical system. In addition, salient region detection technology is closely related to application. So this technology is still a challenging research topic. A survey on salient region detection technology was given in the paper. Firstly, bottom-up and top-down salient region detection approaches were discussed in detail, including technique classification and typical techniques. Secondly, evaluation criteria and open saliency evaluation databases were discussed. At last, the main problems and challenges were highlighted based on analysis of current research.

Keywords Saliency, Visual attention, Salient region detection, Top-down, Bottom-up

随着计算机和网络通信技术的飞速发展,图像数据量变得越来越大。海量的图像数据给信息处理带来了巨大的挑战,如何高效地接收、分析和处理这些图像信息,已成为近年来的研究热点。人类的视觉系统具有一项特别的能力,即可以实时地解析周围的复杂环境,通过主动选择机制来减少接收到的视觉信息数量,忽略大量的不感兴趣部分,我们把图像中最能引起用户兴趣和体现图像内容的区域称为显著区域。人类视觉具有的这种快速搜索到显著区域的能力被称为视觉注意力机制。视觉注意力机制是视觉感知的重要组成部分,以认知心理学家 Triesman 和 Gelade 与神经生物学家 Koch 和 Ullman 为代表的研究者将视觉注意力机制主要分为两大类:自底向上(Bottom-up)数据驱动的预注意机制和自顶向下(Top-down)任务驱动的后注意机制。其中,自底向上的处理是在没有先验知识的情况下由底层数据驱动进行显著目标检测的方法,它属于低级的认知过程,不考虑认知任务对显著

性提取的影响,处理速度快;而自顶向下的处理过程则属于任务驱动发现显著性目标的过程,属于高级的认知过程,根据任务有意识地进行处理并提取出所需要的感兴趣区域,处理速度慢。根据视觉注意力机制的两类处理过程,把显著区域检测方法分为两类,自底向上的显著区域检测方法和自顶向下的显著区域检测方法(这类方法也包括自顶向下和自底向上的结合)。一般说来显著区域具有以下特点:(1)局部差异性,局部区域总是与周围区域具有明显的特征差异;(2)全局稀缺性,从全局范围来看,显著区域的特征出现的频率低,不容易由图像的其他区域复合得到;(3)高层语义特征,人在观察中经常注意到的对象(如人脸、汽车等)经常是图像中的显著区域。两类显著区域检测方法均围绕显著区域的特点并结合颜色、方向、梯度、边缘、边界、纹理等图像特征进行显著区域的提取。

视觉显著性特征提取作为一项独立的技术,可以帮助我

到稿日期:2015-03-10 返修日期:2015-06-15 本文受国家自然科学基金项目(61271369,61372148),北京市信息服务工程重点实验室开放课题(Zk20201402),北京市自然科学基金项目(4152016),北京市教育委员会科技计划面上项目(KM201511417008),北京联合大学人才强校计划人才资助项目(Rk100201510)资助。

梁 晔(1978—),女,博士生,讲师,主要研究领域为图像处理,E-mail:liangye@buu.edu.cn(通信作者);于 剑(1969—),男,教授,博士生导师,主要研究领域为机器学习和图像处理;郎丛妍(1978—),女,教授,博士生导师,主要研究领域为机器视觉和显著性;刘宏哲(1971—),女,博士,副教授,主要研究领域为语义计算和图像处理。

们更好地分析和理解数字图像。提取的显著物体可以为诸多应用提供原始对象,其应用领域极为广泛,如对象识别^[1,31]、对象分割^[2,3]、基于内容的图像检索^[4]、内容已知的图像缩放^[5-7]和图像压缩^[8]等。如何快速准确地找到图像的显著区域尚未形成完整的理论体系,且与具体应用有着密切的关系,对研究人员来说仍是一个富有挑战的课题。本文将对显著区域的检测技术进行综述,以便研究人员更好地了解现状和开展下一步的工作。

本文第1节讨论自下而上的显著区域检测方法,对自下而上的显著区域检测方法进行归类,并对典型方法进行梳理;第2节讨论自上而下的显著区域检测方法,对自上而下的显著区域检测方法进行归类,并对典型方法进行梳理;第3节讨论算法的评价标准;第4节讨论目前流行的显著性评测数据库;最后对目前存在的问题进行总结,给出未来的研究方向。

1 自底向上的显著区域检测方法

自底向上的显著区域检测方法是完全由底层数据驱动的,它没有先验知识指导,不考虑认知任务对显著性提取的影响。本节首先讨论自底向上的显著区域检测方法分类,然后对典型的自底向上的显著区域检测方法进行分析和梳理。

1.1 自底向上的显著区域检测方法分类

研究人员从不同视角提出了很多显著区域检测方法,根据不同视角可以进行不同的分类。

按照是否基于生物学模型可以将显著区域检测方法分成3类:基于生物学模型的方法、纯计算型的方法和混合型的方法。最经典的自底向上的视觉计算模型是 Koch 和 Ullman^[9]在1985年提出来的,该模型首次并行地提取多特征的显著图,并将结果融合为最终的显著图。Itti 等人^[10]在 Koch 和 Ullman 模型的基础上并行地提取多尺度、多特征的显著图,模拟了人类视觉的搜索过程,并将多个显著特征图融合为最终的显著图。这个模型成为以后众多研究工作的基础。

随着研究的进展,基于生物学模型的显著性检测方法过于复杂,研究人员转向以对比度计算为主的方法,进而产生了纯计算模型。纯计算模型通过数学方法对显著性特征提取及其度量进行计算。对比度是计算显著性的核心,按照对比度的范围可以分为基于局部对比度的检测方法和基于全局对比度的检测方法。基于局部对比度的方法容易在轮廓附近产生高显著值,不能高亮地显示整个区域;而基于全局对比度的方法能够给相似的图像区域分配相近的显著值,进而能一致高亮显示整个区域,均匀突出整个目标,但计算速度一般较慢,因为需要比较的次数多。

按照处理的图像信号是否在空域,可以将检测方法分为空域模型和频域模型。空域模型是最直接的研究方式,不需要对图像进行变换,直接在图像本身的二维空间进行特征提取和处理,目前大多数方法都属于空域模型。基于频域的检测算法是对图像进行变换后,在变换域进行处理。目前的方法主要集中在傅立叶变换域中的幅度谱和相位谱。相对于空域模型,基于频域的方法运算量小、速度快。

参与计算对比度的单元粒度对于对比度的计算结果会有很大的不同,根据参与计算对比度的单元粒度可以将检测方法分为基于像素粒度的检测方法、基于统计结果粒度的检测方法和基于聚类区域粒度的检测方法。基于像素粒度的检测方法在计算时比较像素和像素之间的差异;基于统计结果粒

度的检测方法在计算时比较图像的基于某种特征的统计结果的差异;基于聚类区域粒度的检测方法首先通过分割获得子区域块,然后比较子区域块的差异。

机器学习是通过数据改进计算机算法性能的研究,已经在图像检索、图像分类方面有众多的应用。相比之下,目前机器学习方法在显著性提取方面没有那么多应用,但机器学习方法在显著性提取方面的应用也变得越来越重要,因此是否应用机器学习方法也可以作为分类的标准。基于机器学习的模型使用手工标注中的显著区域的局部或全局的特征训练显著性模型,将学习到的结果用于指导显著性提取,其本质思想是从学习中找到视觉特征到显著区域的映射,用于预测图像中的显著对象。

另外,研究人员也发现单纯地依赖图像本身的特征(如颜色、形状、纹理等)计算显著区域是不够的,因此更多的研究人员开始利用图像外部的信息辅助计算图像显著区域。这些外部信息可能是大量的相似图片集、图像的深度信息或者图像外部的文字信息及标注。因此,从这个角度来说,可以分为基于内部线索的显著区域提取和基于外部线索的显著区域提取。

值得强调的是,一种显著区域检测方法可能属于上述归纳的多个类别。

1.2 典型的自底向上的显著区域检测方法

研究人员已经提出了各种各样的自底向上的显著区域检测方法,本文以典型的显著区域检测方法为例,对这些方法的设计思路进行分析和梳理。所选方法基于不同的理论基础,覆盖范围广,各具特点,是目前引用率很高的主流方法。

1.2.1 完全生物学启发的视觉注意模型

IT 注意模型是完全生物学启发的视觉注意模型。1998年 Itti、Koch 等人在 Koch 和 Ullman 的工作^[9]基础上提出一种模拟生物体视觉注意机制的选择性注意模型^[10],完全生物学启发是目前最有影响力的模型。此模型首先在多个尺度下对输入的图像通过线性滤波器提取颜色、亮度和方向特征,经过中央周边差操作计算初级的特征图;接着使用有效的特征合并策略将特征图合并,采用胜者为王的竞争策略找到图中最显著的区域,并通过禁止返回机制使注意力不再返回到已经关注过的区域,继续找到以显著性降序排序的其他目标。IT 方法计算复杂,资源消耗大。这种算法是对生物结构的模拟,但现阶段的研究还不能够对人类视觉系统的实现原理进行完全准确的剖析,因此此方法还有很大的不确定性。另外,此模型是基于局部对比度的,且运算结果严重依赖于参数的选取,得到的显著区域轮廓模糊。

1.2.2 基于频域分析的显著性检测模型

在基于频域分析的显著性检测算法中最具代表性的是 SR 模型^[11],即谱残余模型。通常一幅图像中包含两部分信息:目标信息和冗余信息,谱残余模型正是基于此原理。该模型通过对图像的亮度特征进行二维傅里叶变换得到频域的幅度谱和相位谱。对幅度值取对数后,经过低通滤波器得到的信号为冗余信息,从原图中除去这部分信息就得到了图像的显著区域。谱残余方法既不抽取特征也不计算对比度,更不依赖于先验知识,而是对图像的对数谱进行处理。SR 方法的优点是计算简单、运算速度快,结果不依赖于参数的选择;缺点是从频域转换到空域的显著图会丢失图像中很多细节信息,由于没有保留足够的高频信息,导致边界不清,轮廓模糊,不能高亮地显示整个区域,抗噪能力差。

2009年 Achanta 等提出了一种基于图像频域分析的显著性检测方法:频率调谐的显著区域检测方法(FT方法)^[12]。此方法用每个像素和整幅图像的平均色的色差来定义像素的显著度,这种方法也可以看作是全局对比度的检测方法。采用这种算法得到的显著图的分辨率与原图相同,容易实现,运算量小,计算速度快。它因为保留了比SR算法更多的高频信息,从而显著目标的边界清晰。其缺点是显著区域与背景的对比不是很明显,并且只考虑了一阶平均颜色,这对于复杂自然图像的处理是不够的。

1.2.3 基于局部对比度的纯计算模型

MZ方法^[13]是一种纯计算的基于局部对比度的方法,通过计算每个像素和其邻域的对比度来计算其显著性。MZ方法从显著点、显著区域和显著视图3个角度来描述图像的显著性。MZ的缺点是物体的边缘部分被过分强调,物体内部没有得到一致高亮的显示。

1.2.4 生物学结构和纯计算相结合的模式

GB方法^[14]是将生物学结构和纯计算结合起来的一种方法。GB方法和IT方法采用相同的特征,但是二者在特征显著性的计算以及显著特征图的融合上是不一样的。假设图像中的每个像素都和其他所有像素连接,这样就形成了一个全连接的有向图。在生物学结构中,每个神经元都与周围的神经元连接,构成一个通讯网络,而GB方法和这种方法非常类似,模拟了生物学结构。GB方法采用马尔可夫链来计算特征的显著性。图中边的权重定义为转移概率,权重的计算与两个顶点的相似度和距离成反比。转移概率的大小反映出位置的显著性的大小。通过随机游走的方法搜寻被访问次数少的结点,计算马尔可夫链的平衡状态得到最终的显著图。马尔可夫链的平衡态的计算复杂度非常高,十分费时,所以GB方法的速度不快,一般情况下需要降低图的分辨率,显著图的分辨率也会降低。GB方法是基于全局对比度的,效果虽然比IT方法好,但是仍然存在显著图模糊的缺点。

1.2.5 基于统计结果粒度的检测模型

HC方法^[15]首先按照颜色相似性将像素分成数目较少的类,然后基于全局对比度进行计算。HC方法通过两种技术来加速直方图计算,即①颜色量化:将RGB颜色通道量化为12个级别,大大减少颜色的数量;②合并出现频率低的颜色:出现频率低的颜色即使去掉也不会对图像产生很大的影响。由于采用了简单的颜色量化方法,导致存在相似的颜色可能被归到不同的类别中的问题。为了减少量化的误差和噪音,HC方法对每种颜色的显著性通过相似颜色的显著性进行了平滑。HC方法的计算速度较快,效果也非常不错。

1.2.6 基于聚类区域粒度的检测模型

RC方法^[15]是一种全局对比度的方法。与基于像素粒度的显著度计算方法不同,RC方法是一种基于聚类区域粒度的显著度计算方法。该方法首先通过基于图的分割方法得到不同的子区域,每个子区域认为是同质的,同时考虑区域块大小、颜色和空间因素,以子区域块为单位计算全局对比度。由于RC方法中有一个步骤是分割,因此显著图的质量会受到分割结果的影响。RC方法中采用的分割方法不能很好地贴合区域的边缘,所以可以采用更好的分割方法提高显著图的质量。RC方法能够产生全分辨率的显著图,比HC方法速度慢,但效果更好。自从文献^[15]提出全局对比度之后,研究人员大多从全局对比度出发,从不同的视角使用不同的显著性

度量方法得到不同的显著区域检测方法^[42-44]。2013年及以后出现的显著区域检测论文中基于机器学习的方法和加入高层先验的方法越来越多。

1.2.7 基于深度信息的检测模型

人类生活在真实的三维世界中,深度信息对于理解周围的世界非常重要,在视觉注意中具有重要的作用。文献^[45]通过实验证明了深度信息对于视点预测的重要性。近年来,研究人员开始研究如何利用深度信息进行显著对象的提取^[46-48]。文献^[46]阐述了如何利用立体图像中的领域知识来计算显著图。文献^[48]针对目前RGBD数据库规模大小的问题,建立了包含5000幅深度信息的图像库;并针对已有的融合深度信息的方法过于简单的缺点,提出了利用底层特征对比度、中层区域特征和高层先验多阶段的深度信息和外观特征进行融合。

1.2.8 典型的基于机器学习的检测模型

文献^[19]将显著对象的二值标注看作是显著对象从背景中分割出来的问题,采用条件随机场对显著区域检测问题进行建模,是一种有监督的学习方法。文中采用了多种特征,包括多尺度对比度、中央-周边直方图和颜色空间分布,并从局部、区域和全局的角度去描述显著对象。利用条件随机场进行学习,有效地将特征结合起来用于显著对象的检测,并将显著对象的分割也整合到随机场模型中。文献^[20]将显著图的计算看成了一个回归问题,首先将图像进行从细到粗的基于图的多层分割,对于每个子区域,都通过有监督的学习方法学习区域特征到显著值的映射,通过多层的显著值融合,得到最终的显著图,这种方法最耗时的部分在于多层分割的计算。不同的显著性计算方法存在性能互补的现象,基于此,文献^[21]提出了一种数据驱动的显著性聚集方法。作者观察到不同的显著性方法所做的显著性分析是不一样的,并且每种方法在一幅图片上表现出来的性能也是不同的。作者通过CRF模型对各种显著性方法对一幅图像所作出的贡献进行学习,最终得到融合后的显著图。文献^[17]提出了一种软图像抽象的方法用于图像的显著区域提取。与以往的通过超像素进行图像抽象的方法不同,这种软图像抽象的方法首先对颜色进行聚类,然后通过GMM高斯混合模型进行表示,高斯混合模型能够将图像分解为许多同质的区域。原始的高斯混合模型忽略了图像中的空间联系,因此需要通过空间一致性建立两个GMM通道之间的联系,这样,空间一致的通道会得到更紧凑的聚类。软图像抽象的方法避免了超像素的硬边界划分,能够高亮地显示整个显著区域。

1.2.9 基于相似图像的检测模型

网上可以获得大量的相似图像,这种便利性使得基于相似图像的显著区域检测得以实施,这是一种典型的利用图像外部信息进行显著区域提取的方法^[21,49-51]。文献^[21]提出了一种数据驱动的显著性聚集方法,即通过CRF模型对检索到的相似图像进行训练,得到各种显著性方法对一幅图像所作出的贡献,最终对显著图进行融合。文献^[49]首先建立一个图像索引库,索引图像库里的图像都进行过标注,提取这些图像的局部图像块的特征并将其聚类为视觉字典,这样就可以得到图像区域是否显著的高层表示。对于新的图像,在图像索引库中搜索近邻图像,通过训练好的分类模型对新图像的区域进行显著和非显著的分类,最后将分类结果传播给像素得到显著图。文献^[50]将显著性的计算建模为一个抽样问

题,图像块如果由自身图像和相似图像抽样出来的概率很低,则称之为显著。文献[51]认为相似的图像能够很好地对背景进行估计,可以通过近邻图像对图像的显著区域进行估计。

2 自顶向下的显著区域检测方法

自顶向下的显著区域检测方法(这类方法也包括自顶向下和自底向上的结合方法)包含了高级的认知过程,处理过程属于任务驱动。

2.1 自顶向下的显著区域检测方法概述

目前大多数研究都集中在自底向上的计算模型,自顶向下的模型研究较少。因为自顶向下的模型与特定的任务相关,处理起来更加复杂多变,没有通用的模型可以描述。自底向上的模型完全是数据驱动的,而自顶向下是在底层特征刺激的基础上,将人的先验、知识和期望等高层语义知识融入到检测模型中,对检测过程进行指导,能更好地定位到目标对象,缩减搜索目标。

基于高层语义线索的模型结合了自底向上和自顶向下的优点,充分利用了底层显著特征,又融入了经验、先验的高层语义信息,在计算显著图时考虑了各种先验:颜色、人脸、中心等一般是人眼注意力集中的区域。语义线索在一定程度上改善了显著模型的性能,但同时也限定了模型的应用范围。

2.2 典型的自顶向下的显著区域检测方法

文献[22]提出了一种新的图像表示方法,将其表示为一个低秩矩阵(非显著区域)加上稀疏噪声(显著区域),再利用RPCA技术进行低秩矩阵恢复,得到的噪声就是显著区域。文献[22]认为纯粹的低秩特征在显著区域的提取中并不能得到满意的结果,高层的知识在任务依赖的显著区域提取中非常重要,作者采用了中心位置先验、语义先验(人脸识别)和颜色先验(暖色调更显著)进行高层指导,作为目标函数的先验项来改善检测性能。文献[23]与文献[22]类似,也采用了自底向上和自顶向下相融合的方式,高层的先验有人、人脸和汽车等。文献[24]中除了采用常用的底层特征之外,还采用了水平线中层特征和人、人脸和中心先验这样的高层特征,并通过线性SVM训练特征的权重。

上面的工作^[22-24]使用几种特定类别的对象检测子进行显著区域的检测,虽然能够改善性能,但有一定的局限性。近期,已经有研究工作^[25,39]关注于研究通用的、与类别无关的对象检测子。文献[26]提出了应用这种通用对象检测子进行显著区域的检测,采用自顶向下和自底向上相结合的方法,首先通过通用的对象检测子获得高层的显著性先验,然后应用多示例学习^[27]启发的改进高斯MRF来加强显著区域的一致性,强调了潜在前景像素的影响,获得一致高亮的显著区域。文献[28]使用UFO特征,即独特性、焦点(中心先验)和对象性,进行显著区域的检测,UFO特征中的O代表通用对象检测子。文献[29]讨论了对象性和显著性的相互关系,认为二者可以通过优化相互促进。对象性检测结果既反映了对象存在的可能性有多大,同时也反映了对象可能存在的位置,这对作为先验用于显著性的检测是非常有帮助的。

文献[20]提出了一种新的区域特征:背景度特征,这个概念可以看作是对象性(objectness)相对的,背景度通过有监督的学习得到,可以看作是一种背景先验。文献[18]不是从对比度出发,而是从背景和前景的角度出发进行显著区域的检测。利用图像边框大多是背景的先验信息,通过流形排序

的方法将此先验扩散并增加,得到前景的比较可靠的估计;随后将前景的信息用类似的流形排序的方法加强,得到最终的显著物体。这种方法综合利用图像中背景、前景的先验位置分布及连通性,取得了不错的结果。

3 算法评价标准

3.1 显著图质量的评价

在显著区域检测中,输入是一幅图像,输出是图像的显著图。显著图的质量是评价显著性检测算法好坏的重要标准,一般从以下几个方面评价显著图的质量^[6,15,30]:

(1)最为显著的物体是否突出。显著图能否突显最显著的物体,且这个区域与人的视觉选择保持高度一致。

(2)整个显著物体是否一致高亮。显著物体的一致高亮才能将显著物体区域和背景区域完全分离开来。

(3)物体边界是否精确完整。如果得到的显著区域有着良好的边界,则说明显著区域提取得较准确完整,漏检和误检较少。

(4)抗噪性是否强。若显著性检测算法比较鲁棒,则受图像中的噪声、复杂纹理和杂乱背景的影响较小。

(5)是否全分辨率。若显著图具有和原图像相同的分辨率,则有助于实际的应用。

3.2 可计算的评价标准

显著图质量评价标准可以通过下面4个可计算的评价标准进行度量。

(1)ROC和AUC

ROC(Receiver Operating Characteristic curve)曲线,又称为感受性曲线,是在不同的判断标准下对同一信号刺激的反映。AUC(Area Under ROC curve)指处于ROC曲线下方的面积大小,较大的AUC代表了较好的性能。ROC曲线和AUC可以用来衡量显著性目标的检测结果。

(2)Precision-Recall曲线

将显著图按不同的阈值进行二值化,可以得到不同的准确率Precision和召回率Recall,并绘制Precision-Recall曲线,是衡量显著性检测模型的常用评价方法。

(3)F-Measure

F-Measure又称为F-Score,即将准确率Precision和召回率Recall进行加权处理,其定义如下:

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

为了更多地强调准确率Precision,通常将 β^2 设为0.3。

(4)平均绝对误差(Mean Absolute Error)

平均绝对误差用于衡量二值化后的显著图与真实二值化模板图像之间的区别,其定义为:

$$\text{MAE} = \frac{1}{W \times H} \sum_{x=1}^W \sum_{y=1}^H |S(x,y) - GT(x,y)|$$

其中,W和H分别为图像的宽度和高度,S为二值化后的显著度图,GT为标准的二值化标注图像。

4 显著性评测数据库

显著性评测数据库所包含的图像中有一个或多个显著区域,被用户通过矩形框或像素级二值化的方式标注,用于显著性检测方法的性能评估。目前显著性评测数据库大体有两种:一种是专门为了显著性研究而建立的数据库,另一种是从图像分割领域延伸过来的显著性数据库。

4.1 专门为显著性研究的数据库

(1)MSRA 数据库

文献[32]公布的 MSRA 数据库共包含 130099 幅图片,大多数来自于图像论坛和图像搜索引擎,图片类型非常丰富。从下载的图片库中人工挑选了 20840 幅图片,每幅图像包含一个显著区域,对其进行了矩形框级别的标注。

(2)Achanta 数据库(也称 ASD 数据库)

文献[12]从 Liu 等人^[32]的数据集中选出了 1000 张图像,并为每个图像标出具有 1 个像素级精度的显著区域。这个数据库是引用率非常高的显著性评测数据库。

(3)TDUT-OMRON 数据集

文献[18]公布的数据库中包含了 5168 幅高质量的图片,是手工从 140000 幅图像中挑选出来的。这些图片中包含 1 个或多个显著区域,背景也相对复杂。最终标注的结果有 3 种形式:像素级精度的二值图像、矩形框级别的基准数据集和眼动基准数据集。

(4)MSRA-B1 5000 数据集

文献[20]从 MSRA-B1 数据库中挑选了 5000 幅图像,对已经标注了矩形框的 5000 幅图像手工分割出显著区域,得到二值化的像素级精度的基准数据集。

(5)ImgSal 数据集

文献[33]公布的数据库包含 235 张图片,图片来自于 Google 和其他参考文献。图片的尺寸为 480×640,包含 6 类:50 幅图像包含大的显著区域;80 幅图像包含中等的显著区域;60 幅图像包含小的显著区域;15 幅图像包含杂乱的背景;15 幅图像包含重复的干扰;15 幅图像既包含大的显著区域也包含小的显著区域。标注结果为像素级精度的基准数据集。

(6)ECSSD 数据集

MSRA-1000 数据库中的图像尽管内容有很大的变化,但还存在图像背景结构简单的缺点。针对此缺点,文献[16]构造了 ECSSD 数据集,其包含 1000 幅图像,图像语义更丰富、结构更复杂。标注结果为像素级精度的二值图像。

(7)PASCAL-S 数据集

此数据集^[41]的图像选自于 PASCAL VOC 数据集中的复杂场景,总共包含 20 个对象类。图像集由一个人标注,另一个人检查,最终得到最显著对象的像素级精度标注。这个数据集的缺点在于局限于 20 个类。

4.2 图像分割领域延伸来的数据库

(1)SOD 数据集

SOD 数据库^[34]包含了 300 幅来自于伯克利大学的分割数据库(BSD)^[35],这是第一次将分割领域的数据库用于显著性的评测^[36]。挑选出的图片更加复杂,包含多个前景对象,且多个前景对象具有不同的大小和位置。

(2)SED 数据集

SED 数据集^[37]共包含两个数据集:单一显著区域的数据子集 SED1 和两个显著区域的数据子集 SED2,每个子集包含 100 幅图片。标注结果为像素级精度的二值图像。

(3)CMU-Cornell iCoseg 数据集

iCoSeg 是一个公布的分割数据集^[38],共包含 38 组、643 幅图片,来自于 Flickr 图片网站。标注结果为包含 1 个或更多的显著区域的像素级的二值图像。

结束语 基于生物感知的图像显著特征分析结合人类心

理学、生理学和人类视觉注意机制,模拟人眼机能建立图像显著性提取模型。反过来,显著区域的提取对于检验和丰富相关的理论和模型有很好的促进作用。视觉显著性提取作为一项独立的技术,可以更好地分析和理解数字图像,应用范围十分广泛。近些年来,已经涌现了大量的显著区域的提取方法,且其取得了长足的进步,但仍然存在一些问题有待研究。

(1)特征的选择:显著性检测方法的特征多种多样,各有利弊。如果特征选择得恰当,可以起到相互补充、促进的作用,从而克服某一个特征导致的不合理显著区域。但多种特征一起参与计算,也可能起到反作用,而且也需要更加复杂的显著特征图融合算法以保证最终的显著图更加符合人的感知特点。

(2)生物学模型的完善:视觉显著性提取要遵循生物学规律并基于一定的生物学模型。在生物学上,关于人类视觉显著性的研究已经有很多成果,但是还没有完全揭密人类视觉行为的实现细节,并且一些模型的描述比较复杂,在计算机上难于仿真实现。

(3)显著性评测数据库:数据库的偏差问题是计算机视觉一直存在的问题,如何建立合理的显著性数据库一直是研究人员的关注点^[41]。

(4)文本语义的利用:随着社交网站的流行,网页上的文本信息越来越多。传统的单一媒体相关技术忽略了文本与图像数据在语义上的共性。由于文本数据与图像数据具有互补性的特点,如何挖掘多媒体数据之间的语义关联信息并将两者有机地融合成为研究的热点。

(5)显著对象分割对于显著对象的定位和进一步的处理非常重要,但是目前还没有专门的关于显著对象分割方法的综述报道^[35]。

(6)统一的检测平台:目前的各种显著检测算法的实验环境、平台是不一样的,实验的数据库也不尽相同,缺少一个统一的平台进行各种方法的比较。

研究人员已经提出了各种各样的显著区域的检测方法,但检测的结果仍然存在误检、漏检、检测结果模糊、不够精确、鲁棒性不高的缺点。更好的检测结果一直是研究人员追求的目标。

参 考 文 献

- [1] Rutishauser U, Walther D, Koch C, et al. Is bottom-up attention useful for object recognition? [C]//CVPR. 2004:37-44
- [2] Han J, Ngan K, Li M, et al. Unsupervised extraction of visual attention objects in color images[J]. IEEE TCSV, 2006, 16(1): 141-145
- [3] Ko B, Nam J. Object-of-interest image segmentation based on human attention and semantic region clustering[J]. Journal of the Optical Society of America A, 2006, 23(10):2462-2470
- [4] Chen T, Cheng M M, Tan P, et al. Sketch photo: Internet image montage[J]. ACM TOG, 2009, 28(5):124:1-10
- [5] Wang Y S, Tai C L, Lee T Y. Optimized scale-and-stretch for image resizing[J]. ACM Trans. Graph., 2008, 27(5):32-39
- [6] Zhang G X, Cheng M M, Hu S M, et al. A shape-preserving approach to image resizing[J]. Comput. Graph. Forum, 2009, 28(7):1897-1906
- [7] Zhang Y F, Hu S M, Martin R R. Shrinkability maps for content-aware video resizing[J]. Comput. Graph. Forum, 2008, 27(7):1797-1804

- [8] Christopoulos C, Skodras A, Ebrahimi T. The JPEG2000 still image coding system; an overview[J]. *IEEE Trans. on Consumer Electronics*, 2002, 46(4): 1103-1127
- [9] Koch C, Ullman S. Shifts in selective visual attention; towards the underlying neural circuitry[J]. *Human Neurobiology*, 1985, 4: 219-227
- [10] Itti L, Koch C, Niebur E. A model of saliency-based visual attention for rapid scene analysis[J]. *IEEE TPAMI*, 1998, 20(11): 1254-1259
- [11] Hou X, Zhang L. Saliency detection: A spectral residual approach[C]//CVPR. 2007; 1-8
- [12] Achanta R, Hemami S, Estrada F, et al. Frequency-tuned salient region detection[C]//CVPR. 2009; 1597-1604
- [13] Ma Yu-fei, Zhang Hong-jiang. Contrast-based image attention analysis by using fuzzy growing[C]//Proceedings of the Eleventh ACM International Conference on Multimedia. New York, New York, USA; ACM Press, 2003; 374-381
- [14] Harel J, Koch C, Perona P. Graph-based visual saliency[J]. *Advances in Neural Information Processing Systems*, 2006, 19: 545-552
- [15] Cheng Ming-ming, Zhang Guo-xin, Mitra N J, et al. Global Contrast based Salient Region Detection[C]//IEEE International Conference on Computer Vision and Pattern Recognition. 2011; 409-416
- [16] Yan Qiong, Xu Li, Shi Jian-ping, et al. Hierarchical Saliency Detection[C]//CVPR. 2013; 1155-1162
- [17] Cheng Ming-ming, Warrell J, Lin Wen-yan. Efficient Salient Region Detection with Soft Image Abstraction[C]//ICCV. 2013; 1529-1536
- [18] Yang Chuan, Zhang Li-he, Lu Hu-chan, et al. Saliency Detection Via Graph-Based Manifold Ranking[C]//CVPR. 2013; 3166-3173
- [19] Liu Tie, Yuan Ze-jian, Sun Jian, et al. Learning to Detect a Salient Object[J]. *Pattern Analysis & Machine Intelligence*, 2010, 33(33): 353-367
- [20] Jiang Huai-zu, Wang Jing-dong, Yuan Ze-jian, et al. Salient object detection; a discriminative regional feature integration approach[C]//CVPR. 2014; 2083-2090
- [21] Mai L, Niu Y, Liu F. Saliency aggregation: A data driven approach[C]//CVPR. 2013; 1131-1138
- [22] Wu Ying. A Unified Approach to Salient Object Detection via Low Rank Matrix Recovery[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR)(Oral). 2012; 853-860
- [23] Borji A. Boosting Bottom-up and Top-down Visual Features for Saliency Estimation[C]//CVPR. 2012; 438-445
- [24] Judd T, Ehinger K, Durand F, et al. Learning to predict where humans look[C]//ICCV. 2009
- [25] Alexe B, Deselaers T, Ferrari V. What is an object? [C]//CVPR. 2010
- [26] Jia Yang-qing, Han Mei. Category-Independent Object-level Saliency Detection[C]//ICCV. 2013; 1761-1768
- [27] Maron O, Lozano-Pérez T. A framework for multiple-instance learning[J]. *NIPS*, 1998, 200(2): 570-576
- [28] Jiang Peng, Ling Hai-bin, Yu Jing-yi, et al. Salient Region Detection by UFO: Uniqueness, Focusness and Objectness[C]//ICCV. 2013; 1976-1983
- [29] Chang K Y, Liu T L, Chen H T, et al. Fusing Generic Objectness and Visual Saliency for Salient Object Detection[C]//International Conference on Computer Vision (ICCV). 2011
- [30] Goferman S, Zelnik-Manor L, Tal A. Context-aware saliency detection[C]//CVPR. 2010
- [31] Hui Bin, Tang Xu-sheng, Luo Hai-bo, et al. SDF Matched Filter Based on Gabor Wavelet Transform for Face Recognition[J]. *Information and Control*, 2008, 37(5): 633-636
- [32] Liu Tie, Sun Jian, Zheng Nan-ning, et al. Learning to Detect A Salient Object[C]//Proc. IEEE Conf. on Computer Vision and pattern Recognition (CVPR). 2007
- [33] Li Jian, Levine M, An Xiang-jing, et al. Saliency Detection Based on Frequency and Spatial Domain Analyses[C]//BMVC. 2011
- [34] <http://elderlab.yorku.ca/SOD/#download>
- [35] Martin D, Fowlkes C, Tal D, et al. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics [C] // ICCV. 2001; 416-423
- [36] Movahedi V, Elder J H. Design and perceptual validation of performance measures for salient object segmentation [C] // CVPRW. 2010; 49-56
- [37] Sharon A, Meirav G, Achi B, et al. Image segmentation by probabilistic bottom-up aggregation and cue integration[C]//CVPR. 2007; 1-8
- [38] Batra D, Kowdle A, Parikh D, et al. icoseg: Interactive co-segmentation with intelligent scribble guidance[C]//CVPR. IEEE, 2010; 3169-3176
- [39] Cheng Ming-ming, Zhang Zi-ming, Lin Wen-yan, Torr P. BING: Binarized Normed Gradients for Objectness Estimation at 300fps [C]//CVPR. 2014; 3286-3293
- [40] Borji A, Sihite D N, Itti L. Salient Object Detection A Benchmark[M]//ECCV 2012. Springer Berlin Heidelberg, 2012; 414-429
- [41] Li Yin, Hou Xiao-di, Koch C. The Secrets of Salient Object Segmentation[C]//CVPR. 2014; 280-287
- [42] Siva P, Russelly C, Xiang Tao, Lourdes Agapito. Looking Beyond the Image: Unsupervised Learning for Object Saliency and Detection[C]//CVPR. 2013; 3238-3245
- [43] Li Xi, Li Yao, Shen Chun-hua, et al. Contextual Hypergraph Modelling for Salient Object Detection[C]//ICCV. 2013; 3328-3335
- [44] Jiang Bo-wen, Zhang Li-he, Lu Hu-chuan, et al. Saliency Detection via Absorbing Markov Chain[C]//ICCV. 2013; 1665-1672
- [45] Lang C, Nguyen T V, Katti H, et al. Depth matters: Influence of depth cues on visual saliency[C]//ECCV. 2012; 101-115
- [46] Niu Y, Geng Y, Li X, et al. Leveraging stereopsis for saliency analysis[C]//CVPR. 2012; 454-461
- [47] Desingh K, Krishna K M, Rajan D, et al. Depth really matters: Improving visual salient region detection with depth[C]//BMVC. 2013
- [48] Peng Hou-wen, Li Bing, Xiong Wei-hua, et al. RGBD Salient Object Detection: A Benchmark and Algorithms [C] // ECCV. 2014; 92-109
- [49] Marchesotti L, Cifarelli C, Csurka G. A framework for visual saliency detection with applications to image thumbnailing[C]//ICCV. 2009; 2232-2239
- [50] Siva P, Russell C, Xiang T, et al. Looking beyond the image: Unsupervised learning for object saliency and detection [C] // CVPR. 2013; 3238-3245
- [51] Wang M, Konrad J, Ishwar P, et al. Image saliency: From intrinsic to extrinsic context[C]//CVPR. 2011; 417-424