

视觉图像显著性检测综述



袁 野^{1,2,3} 和晓歌¹ 朱定坤⁴ 王富利⁴ 谢浩然⁵ 汪 俊¹ 魏明强^{1,2,3}
郭延文³

1 南京航空航天大学计算机科学与技术学院/人工智能学院 南京 210006

2 模式分析与机器智能工业和信息化部重点实验室 南京 210016

3 南京大学计算机软件新技术国家重点实验室 南京 210023

4 香港公开大学 香港 999077

5 香港教育大学 香港 999077

(yuenye@nuaa.edu.cn)

摘 要 当今图像数据呈爆炸式增长,如何利用计算机高效地获取、处理图片信息成为领域内重要的研究课题。在人类视觉注意机制的启发下,研究人员发现将这种机制引入机器图像处理任务中可以大大提高信息提取的效率,从而更好地节省有限的计算资源。视觉图像显著性检测即利用计算机模拟人类的视觉注意机制,对图片中各部分信息的重要程度进行计算。其在图像分割、视频压缩、目标检测、图像索引等领域得到了广泛的应用,有着重要的研究价值。文中介绍了图像显著性检测算法的研究现状,首先以信息驱动来源为切入点,对显著性检测模型进行概述,之后分析了现有几种典型的显著性检测算法,并根据是否基于学习的模型将其分为基于非学习模型、基于传统机器学习模型以及基于深度学习模型 3 类。针对第一类,文中较为详细地对基于局部对比度和基于全局对比度的显著性检测算法进行了分类比较,指出了各自的优势与不足;针对后两类,分析了机器学习算法及深度学习在显著性检测中的应用。最后对现有的显著性检测算法进行了总结比较,对该领域研究的下一步发展方向进行了展望。

关键词: 显著区域检测;视觉显著性检测;深度学习;机器学习;视觉注意机制

中图法分类号 TP391

Survey of Visual Image Saliency Detection

YUAN Ye^{1,2,3}, HE Xiao-ge¹, ZHU Ding-kun⁴, WANG Fu-lee⁴, XIE Hao-ran⁵, WANG Jun¹, WEI Ming-qiang^{1,2,3} and
GUO Yan-wen³

1 College of Computer Science and Technology/College of Artificial Intelligence, Nanjing University of Aeronautics and Astronautics,
Nanjing 210006, China

2 MIIT Key Laboratory of Pattern Analysis and Machine Intelligence, Nanjing 210016, China

3 State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210023, China

4 The Open University of Hong Kong, Hong Kong 999077, China

5 The Education University of Hong Kong, Hong Kong 999077, China

Abstract In today's society where image data are exploding, how to use computer to efficiently acquire and process image information has become an important research topic. Under the inspiration of human visual attention mechanism, researchers have found that when this mechanism is introduced into machine image processing tasks, the efficiency of information extraction can be greatly improved, thus saving more limited computing resources. Visual image saliency detection is to use computers to simulate the human visual attention mechanism to calculate the importance of the information of each part in the image, which has been

到稿日期:2019-09-02 返修日期:2019-11-15 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目:国家自然科学基金项目(61502137);中央高校基本科研业务费(NJ2019010);南京大学计算机软件新技术国家重点实验室开题课题(KFKT2018B20);香港教育大学 2018-19 院长研究基金之早期职业计划种子基金(SFG-10);香港教育大学 2018-19 院长研究基金之跨学科研究计划基金(FLASS/DRF/IDS-3)

This work was supported by the National Natural Science Foundation of China (61502137), Fundamental Research Funds for the Central Universities (NJ2019010), Grant from State Key Laboratory for Novel Software Technology at Nanjing University (KFKT2018B20), Seed Fund for Early Career Scheme of the Dean's Research Fund 2018-19 (SFG-10) of the Education University of Hong Kong and Interdisciplinary Research Scheme of the Dean's Research Fund 2018-19 (FLASS/DRF/IDS-3) of the Education University of Hong Kong.

通信作者:魏明强(mqwei@nuaa.edu.cn)

widely used in image segmentation, video compression, target detection, image indexing and other aspects, and has important research values. This paper summarizes and introduces the research situation of image saliency detection algorithms. Firstly, it takes information-driven sources as starting point to summarize the saliency detection model, and then analyzes several typical saliency detection algorithms. The models are divided into 3 categories according to whether they are based on learning models, which are based on non-learning models, based on traditional machine learning models and based on deep learning models. For the first category, the paper compares in more details the saliency detection algorithms based on local contrast and global contrast, and points out their respective advantages and disadvantages. For the latter two categories, this paper analyzes the application of machine learning algorithms and deep learning in saliency detection. Finally, this paper summarizes and compares the existing saliency detection algorithms and prospects the future development direction of the research in this aspect.

Keywords Salient region detection, Visual saliency detection, Deep learning, Machine learning, Visual attention mechanism

1 引言

随着计算机科学与技术的不断发展,数字图像数据呈爆炸式增长,如何高效地处理海量图片信息,并从中获取感兴趣的部分成为了人们所面临的巨大问题。涵盖生物学、心理学等各领域的人类视觉注意机制受到了研究人员的关注。人们在面对复杂的环境时,会下意识地识别出其中的特殊区域,并且识别过程迅速且准确,人类拥有的这种能力被称为视觉注意机制,而被识别出的特殊区域被称为显著性区域。图像显著性检测即利用计算机模拟人类的视觉注意机制,对图片中各部分信息的重要程度进行计算,进而得到与原图尺寸相同的灰度图,灰度图中各个点的像素值表征原图像素点的显著值,亮度大小与显著性值的大小呈正相关,这种灰度图也称为显著性图,通过显著性图可以快速找到人类感兴趣的显著性区域。

图像显著性检测已经在视觉及图像相关的各个领域得到了应用,如图像分割^[1-2]、图像视频压缩^[3]、目标检测^[4]、图像索引^[5]等。根据显著性检测算法的建模方法是否基于学习,可将其分为3类:非学习型、传统机器学习型以及深度学习型。本文对几种显著性检测算法模型进行梳理和介绍,以帮助研究人员更加清楚地了解现状及未来的发展方向。本文首先对几种典型的显著性检测算法进行分类介绍,其次对现有的一些显著性检测算法进行梳理和概括,最后对目前显著性检测研究遇到的技术难点进行简单介绍,并对未来研究进行展望。

2 基于非学习模型的显著性检测

按照信息驱动的来源,显著性检测模型可以分为两类:自顶向下模型和自底向上模型^[6]。前者对应于人类视觉系统中,在已有知识的前提下,或是在任务的驱动下,对外界环境自上向下的检索过程。虽然它能够在全局上获得较好的结果,但在某些局部细节上会有模糊现象产生。后者对应于人类视觉系统中在图像特征及数据的驱动下,对外界环境自下而上的快速感知过程,其在局部细节上会有更加优良的效果。大多非学习的模型使用自底向上的方法。

在对非学习模型进行显著性检测研究时,研究人员通常使用颜色、亮度、纹理等低级特征或水平线、中心点等中级特征。在底层视觉显著性计算时,对比度起着重要的作用,现有

的显著性检测算法大多是通过计算图像中各个子区域与周围相邻区域的对比度,来得出该区域的显著性。依据相邻区域范围的大小,可将算法划分为基于局部对比度和基于全局对比度两种。

2.1 基于局部对比度的显著性检测算法

基于局部对比度的显著性检测算法是将图像中各个子区域的相邻区域定义为与其距离较小的局部区域,通过求解它们之间的对比度来得到该子区域的显著性值。

受到生物学中视觉注意机制的启发,Itti等^[7]提出一种自底向上的显著性检测 Itti 模型(IT),其成为该领域的开山之作。该模型首先通过高斯金字塔对图像进行高斯模糊和降采样,形成包括原图尺度在内的9个尺度的图像,并提取图像亮度、颜色、方向特征,然后根据特征计算多尺度中心-周边差异,得到每种特征对应的多个特征图。具体过程如下。

(1)提取亮度特征

$$I = \frac{r+g+b}{3} \quad (1)$$

其中, I 表示亮度; r, g, b 分别为红、绿、蓝通道的色彩值。根据高斯金字塔可以计算得到9个层次上的灰度图像,再根据中心与周边尺度做差(见式(2)),得到6张亮度特征图。

$$I_{(c,s)} = |I(c) \ominus I(s)| \quad (2)$$

(2)提取颜色特征

因为颜色在亮度很低时难以感知,所以归一化时亮度 I 仅选取大于整幅图像中最高亮度值1/10的亮度,并采用4个颜色通道 R, G, B, Y 分别表示红、绿、蓝、黄4个颜色通道:

$$\begin{aligned} R &= r - (g+b)/2 \\ G &= g - (r+b)/2 \\ B &= b - (r+g)/2 \\ Y &= (r+g)/2 - |r-g|/2 - b \end{aligned} \quad (3)$$

根据双色对立理论,红与绿、蓝与黄颜色对互相抑制,因此根据式(4)可以得到12张颜色特征图。

$$\begin{aligned} RG(c,s) &= |(R(c) - G(c)) \ominus (G(s) - R(s))| \\ BY(c,s) &= |(B(c) - Y(c)) \ominus (Y(s) - B(s))| \end{aligned} \quad (4)$$

(3)提取方向特征

局部方向信息根据 Gabor 滤波器从亮度 I 中获取,分别从 $0^\circ, 45^\circ, 90^\circ, 135^\circ$ 4个方向计算方向显著图,然后计算4个角度上的差异,得到24张方向特征图,具体公式为:

$$O(c,s,\theta) = |O(c,\theta) \ominus O(s,\theta)| \quad (5)$$

其中, I 为亮度, Θ 表示中心周边差异, c 表示中心尺度, s 表示周边尺度, RG 表示红绿颜色差, BY 表示蓝黄颜色差。模型最后将各个特征图所对应的显著性图进行归一化, 融合得到最终的显著性图。该模型在特定的环境下对图像有良好的效果, 但由于是对图像不断地进行分解处理, 因此显著性图像只能模糊地确定显著目标的位置, 在轮廓和边缘区域上会有缺失, 分辨率也会降低。

该方法较为严格地模仿了生物结构模型, 因此计算复杂度较高。为了解决这一问题并提高检测质量, Ma 等^[8]提出一种基于局部对比度的显著区域检测算法(MZ), 该算法对图像进行均值滤波和降采样, 然后计算各个像素点与其相邻 N 个像素点在 LUV 颜色空间中的欧氏距离之和, 归一化后得到对应像素点的显著值, 并利用模仿人类心理行为的模糊增长方法从显著图中提取显著性区域。

$$\mu_A = \begin{cases} 1, & x \geq a \\ \frac{x-u}{a-u}, & u < x < a \\ 0, & x \leq u \end{cases} \quad (6)$$

$$\mu_U = \begin{cases} 0, & x \geq a \\ \frac{x-a}{u-a}, & u < x < a \\ 1, & x \leq u \end{cases} \quad (7)$$

式(6)和式(7)将显著性图像的像素分为两个互斥的模糊集: 视觉注意区 B_A 与非注意区 B_U 。 μ_A 与 μ_U 分别定义了各个像素点对两个模糊集的隶属度, x 为灰度自变量, a 与 u 为确定函数形状的参数。由于这种方法只能在单一尺度上进行, 因此显著性对象的边缘区域显著值较高, 内部区域显著值较低, 体现在显著性图像上即轮廓边缘处高亮而内部区域较暗, 显著性对象无法得到均匀的显示。

Achanta 等^[9]提出了一种更为简单的新方法(AC), 针对不同的像素点, 计算出以其为中心点的区域在 Lab 颜色空间中的平均特征向量, 以及该区域的邻域在 Lab 颜色空间中的平均特征向量, 并将它们的距离之差作为该点的显著性值。

$$C_{i,j} = D \left[\left(\frac{1}{N_1} \sum_{p=1}^{N_1} v_p \right), \left(\frac{1}{N_2} \sum_{q=1}^{N_2} v_q \right) \right] \quad (8)$$

其中, $C_{i,j}$ 为图像中 (i, j) 位置像素点的显著值; N_1, N_2 分别为中心点区域与其邻域的像素数; v_p, v_q 为像素对应的特征向量。当特征向量不相关时, 距离距离 D 为欧氏距离; 当特征向量相关时, 距离 D 为马氏距离。多尺度的检测则通过调整窗口的大小来实现。最后, 采用平均阈值法和 K 均值算法对图像进行分割, 得到最终的显著图。此算法较之前的方法, 模型更简单、计算速度更快、效果更好, 且保持了与原始图像相同的分辨率。

Goferman 等^[10]通过分析比较现有的几种方法, 认为生成的显著性图像不仅要包括提取出的显著性对象, 同时要包括包含语义部分的背景区域, 由此提出了一种新的模型(CA)。其根据视觉注意机制提出算法应遵循 4 项原则: 1) 应考虑图像的局部低级特征, 如对比度和颜色等; 2) 应通过抑制频繁出现的特征, 在全局上突出具有独特性的特征; 3) 显著性对象的像素点应当具有集中聚集性; 4) 应考虑高级特征。在

4 项原则的基础上, 有:

$$d(p_i, p_j) = \frac{d_{\text{color}}(p_i, p_j)}{1 + c \cdot d_{\text{position}}(p_i, p_j)} \quad (9)$$

其中, p_i, p_j 分别表示以像素点 i 或像素点 j 为中心的局部区域块; $d_{\text{color}}(p_i, p_j)$ 表示 p_i, p_j 在 Lab 颜色空间中的欧氏距离; $d_{\text{position}}(p_i, p_j)$ 表示 p_i, p_j 在 Lab 物理空间上的欧氏距离; 本算法中 $c=3$ 。当像素点 i 对于任意像素点 j 的 $d(p_i, p_j)$ 值都很高时, 像素点 i 显著。然后, 通过平均各个尺度下的显著值, 增强显著区域与非显著区域的对比度, 从而提高显著性图像的质量。最后加入上下文信息修正, 进一步提高显著性对象附近的显著值, 得出最终的显著性图像。

如图 1 所示, 以上几种基于局部对比度的显著性检测算法虽然能够检测出显著性对象在图像中的位置, 但由于算法侧重于各个像素点所在的局部区域, 显著性对象的边缘往往有着较高的显著值, 而对象内部的显著值较低。在显著图上体现为显著性对象边缘亮度高, 内部亮度低, 显著性对象无法得到均匀的突出。当输入图像的显著性对象的内部特征较为一致时, 这种缺陷尤为明显。因此, 为了解决这一问题, 研究人员提出了关注整幅图像的基于全局对比度的显著性检测算法, 其在实验中取得了较为良好的效果。

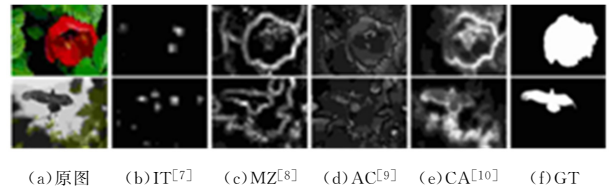


图 1 基于局部对比度显著性检测算法的显著性图像^[29]

Fig. 1 Saliency images based on local contrast saliency detection algorithms^[29]

2.2 基于全局对比度的显著性检测算法

基于全局对比度的显著性检测算法是将图像中各个子区域的相邻区域定义为整幅图像, 通过求解图像中每个子区域与整幅图像的对比度, 来得到图像中每个子区域的显著性值。

Achanta 等^[11]最先提出了基于全局对比度的显著性检测方法(FT)。他们发现, 图像在频率域中的高频部分表示图像的纹理等细节信息, 低频部分表示图像的边缘及轮廓的整体信息。利用这一特性, 他们将图像从空间域转化到频率域。原本在空间中的卷积操作对应在频率域上就是进行滤波处理, 于是将高斯差分公式作为带通滤波器提取显著性区域。高斯差分公式为:

$$DoG(x, y) = \frac{1}{2\pi} \left[\frac{1}{\sigma_1^2} e^{-\frac{(x^2+y^2)}{2\sigma_1^2}} - \frac{1}{\sigma_2^2} e^{-\frac{(x^2+y^2)}{2\sigma_2^2}} \right] - G(x, y, \sigma_1) - G(x, y, \sigma_2) \quad (10)$$

其中, σ_1 与 σ_2 为高斯标准差 ($\sigma_1 > \sigma_2$)。当设定两个高斯标准差成一定比例时, 可得联合公式:

$$\sum_{n=0}^{N-1} G(x, y, \rho^{n+1}\sigma) - G(x, y, \rho^n\sigma) = G(x, y, \sigma\rho^N) - G(x, y, \sigma) \quad (11)$$

其中, $\rho = \sigma_1 / \sigma_2$, 选取比例为 1.6:1 时, 每次高斯差分都会保留一定范围的频率。将所有高斯差分结合起来就是将所有得到

的频率收集起来,从而使显著性区域被全覆盖。接着将各个像素点在滤波后的像素值与全局平均像素值的欧氏距离定义为图像的显著值,从而得到显著图。然而这种方法的本质是检测图像中颜色特征稀疏性大的区域,因此当图像中显著性区域偏大或非显著区域有少量且突出的颜色时,结果往往不理想。

Zhai 等^[12]提出一种以亮度为特征的全局对比度显著性检测算法(LC)。为了降低由于 Lab 颜色空间颜色数量过多所造成的巨大计算代价,该算法仅使用亮度特征来减少颜色数量,从而使特征值量化在 $[0,255]$,再利用一维直方图将像素点按特征归类,计算出各类像素点与其余像素点的特征距离,将该距离作为各个像素点的特征值,最后得到显著图。

上述方法仅仅考虑了亮度作为显著性特征,Cheng 等^[13]在其启发下提出了以颜色为显著性特征的两种显著性检测算法。一种是基于直方图对比度的检测算法(HC),每个像素点的显著值定义为该像素点和图像中其他像素在 Lab 空间中的颜色距离,因此相同颜色具有相同的显著值。

$$S(I_k)=S(c_l)=\sum_{j=1}^n f_j D(C_l,C_j) \tag{12}$$

其中, c_l 为对应像素点 I_k 的颜色; f_j 为颜色 c_l 在整幅图像中出现的频率; $D(C_l,C_j)$ 为两种颜色在 Lab 颜色空间中的距离。然后使用基于直方图加速的方法,将每个颜色通道由 256 量化至 12,去除出现频率较小的颜色,确保剩余的颜色可以覆盖图像 95%的像素,其余 5%的像素对应的颜色用直方图中相邻颜色代替,从而大大降低了时间复杂度。为了减少量化误差,该算法又对颜色空间进行了平滑,最终可以得到较好的实验结果。另一种是基于局部对比度的方法(RC),将图像分割为若干区域后,把空间距离作为权重加入到之前的过程中。

$$S(r_k)=\sum_{r_k \neq r_i} \exp(-D_S(r_k,r_i)/\sigma_s^2)\omega(r_i)D_r(r_k,\gamma_i) \tag{13}$$

其中, $\omega(r_i)$ 为区域 r_i 的权值; $D_r(r_k,\gamma_i)$ 为两个区域在颜色空间中的距离; $D_S(r_k,r_i)$ 为两个区域在物理空间中的距离; σ_s 控制空间权值的强度,由此得出各个区域的显著值 $S(r_k)$,最终得到的显著性图像更为准确且质量更高。

Perazzi 等^[14]在上述方法的基础上对颜色和空间的权值进行了进一步的扩充和定义,提出了一种以高斯滤波定义对对比度的显著性检测模型。首先对输入图像进行区域分割,然后分别计算出以空间距离为加权的区域颜色唯一性和以颜色相似度为加权的颜色空间分布,具体公式如下:

$$U_i=\sum_{j=1}^N \|c_i-c_j\|^2 \cdot \omega(p_i,p_j) \tag{14}$$

其中, U_i 为颜色唯一性, p_i 与 c_i 分别为区域 i 在空间及 Lab 颜色空间中的位置; $\omega(p_i,p_j)$ 为局部或全局对比度的控制项,权值大小与空间距离呈负相关。

$$D_i=\sum_{j=1}^N \|p_j-\mu_i\|^2 \cdot \omega(c_i,c_j) \tag{15}$$

其中, D_i 为颜色空间分布, μ_i 为经过颜色 c_i 加权后 p_i 的平均空间位置; $\omega(c_i,c_j)$ 为两种颜色的相似度,权值大小与颜色相似度呈正相关。

$$S_i=U_i \cdot \exp(-k \cdot D_i) \tag{16}$$

然后,以颜色空间分布的负指数形式来加权区域颜色唯

一性,从而得到超像素区域的显著值。最后,将各个像素点所在的超像素区域与周围的超像素区域进行高斯线性加权,得到的值即为各个像素点的显著值。

Shi 等^[15]提出了一种基于颜色和结构对比度并施加空间先验的显著性检测算法。通过计算局部与全局图像的颜色对比度和结构对比度,对每个像素点计算显著值,然后施加空间加权先验,来保证最终的高显著值像素点紧凑且居中,最后利用高效的边缘图像感知表示和滤波技术生成空间相干且边缘保留的显著性图像。

上述几种方法在 SOD 数据集和 ASD 数据集下的准确率和召回率评分如表 1 和表 2 所列。如图 2 所示,基于全局对比度的显著性检测方法较基于局部对比度生成的显著图,能更加均匀、清晰地突出显著性对象。但当图像中显著性对象区域过大时,由于算法大多是将当前像素与其余像素的特征距离之和作为当前像素的显著值,因此会出现显著性目标区域的显著值反而较小的情况,从而将背景的非显著区域标注为显著区域,出现错误的检测结果。它们对应的 PR 曲线如图 3 所示。随着机器学习的不断进步和大规模的应用,研究人员寻求利用更复杂的学习模型来解决这个问题,进而提出了基于机器学习的显著性检测模型,并且取得了更为优秀的结果。

表 1 检测算法基于 SOD 数据集的对比

Table 1 Comparison of detection algorithms based on SOD

Algorithm	Accuracy	Recall rate
IT	0.52	0.24
MZ	0.48	0.37
AC	0.51	0.32
CA	0.58	0.39
FT	0.45	0.26
LC	0.43	0.32
HC	0.52	0.33
RC	0.71	0.36

表 2 检测算法基于 ASD 数据集的对比

Table 2 Comparison of detection algorithms based on ASD

Algorithm	Accuracy	Recall rate
IT	0.58	0.22
MZ	0.42	0.41
AC	0.58	0.49
CA	0.62	0.57
FT	0.73	0.60
LC	0.56	0.55
HC	0.76	0.75
RC	0.83	0.60

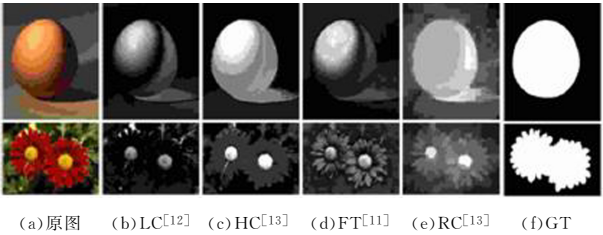


图 2 基于全局对比度显著性检测算法的显著性图像^[29]

Fig. 2 Saliency images based on global contrast saliency detection algorithms^[29]

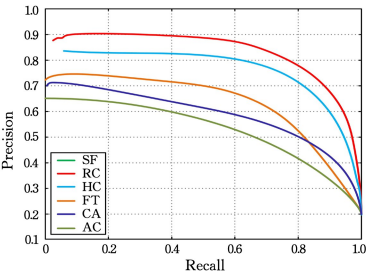


图3 显著性检测方法对应的召回率-准确率曲线图^[14]
(电子版为彩色)

Fig. 3 Precision and Recall rates for all algorithms

3 基于传统机器学习模型的显著性检测

Liu等^[16]最先将机器学习算法用于解决显著性检测问题,提出可以提取原始图像的多尺度对比度、中心环绕直方图以及颜色空间分布等显著特征,以更好地对目标区域进行检测。其首先构建了一个包含20000多个标记图像的大型图像数据库,将其用于训练和测试。之后,为了解决显著性对象识别的问题,将提取的特征通过条件随机场(Conditional Random Field,CRF)有效地结合起来。

$$E(A|I)=\sum_{x,k=1}^K\lambda_kF_k(a_x,I)+\sum_{x,x'}S(a_x,a_{x'},I)\tag{17}$$

其中, I 为给定图像; F_k 为对应的显著特征函数,用于描述一个像素点是否属于显著性对象; S 为配对特征,为两个相邻像素点的空间关系; λ_k 为第 k 个特征的权重; x,x' 是两个相邻的像素。最后结合分割技术检测出未知形状的显著性目标。该模型首次将图像显著性检测问题定义为二分类问题,将属于显著性对象的像素点标注为1,不属于显著性对象的像素点标注为0,具有开创性的意义。

Rahtu等^[17]根据基于滑动窗口和带通滤波的方法,提出一种基于显著性度量和随机条件场模型的显著性分割模型,它能够使输出图像与输入图像保持相同的分辨率。通过计算滑动窗口中心与周围的特征差异,来得到对应区域的显著性值,公式如下:

$$S_o(x)=\frac{h_{K,o}(x)p_0}{h_{K,o}(x)p_0+h_{B,o}(x)(1-p_0)}\tag{18}$$

其中, p_0 为给定0到1之间的值; K 为滑动窗口的内部; B 为滑动窗口的邻域; h 为在对应区域内(K 或 B)像素点 x 对应特征出现的概率。然后建立随机条件场模型,通过求解能量方程来实现目标的分割。该模型在静态图像和动态视频中都能获得较好的分割效果。

Jiang等^[18]提出将图像显著性图计算问题归结为监督学习回归问题。首先对输入图像进行多尺度分割;其次使用随机森林回归算法进行区域显著值的计算,将区域特征值映射到各个区域的显著值上;最后将生成的各个尺度上的显著图像进行融合,从而得到最终的显著性图像。该模型主要的贡献在于:1)它将区域对比(Regional Contrast)、区域属性(Regional Property)和区域背景描述符(Regional Backgroundness Descriptors)整合在一起,从而形成显著图;2)引入一个新的区域特征向量——对应于对象描述符的背景描述符(Backgroundness Descriptors):

$$x_k^b(R_i)=D_k(v^{R_i},v^B)\tag{19}$$

其中, $x_k^b(R_i)$ 为第 k 个特征上区域 R_i 的背景值; v 表示对应区域内的特征向量,用于表征背景特征。

Mai等^[19]发现不同的显著性检测方法具有互补性,因此针对不同的图像选取不同的计算方法将产生更好的结果,故提出了一种基于数据驱动的显著性检测方法。该方法利用条件随机场(Conditional Random Field,CRF)对多个特征图进行整合并得到对应的显著性图。首先利用Gist全局描述子找到对应相似的图像;然后利用条件随机场对多张图像采用对应的多种算法,同时对输出的显著性图像进行整合得到对应显著性图。经过实验发现,与其他显著性检测算法相比,这种模型可以得到更好的结果。

Lu等^[20]提出一种以自适应学习融合显著性图像的模型。首先将图像边缘作为背景,通过计算全局对比度来构建弱显著图;然后以弱显著性图像为训练样本,对强显著性模型进行训练,再利用强分类器对图像中的显著性和非显著性超像素进行分类,得到单尺度上的显著性图像;最后将弱显著性图像与强显著性图像进行融合,生成最终的显著性图像。

近年来,随着深度学习理论及应用的迅速发展,许多国内外研究人员开始将深度学习模型应用于图像的显著性检测。由于深度卷积神经网络能够在大量的数据样本下学习得到高级特征,因此可以避免传统显著性检测方法在复杂环境中由于低级特征的限制而无法达到预期的效果,其在一些复杂的场景中也可以取得良好的效果。

4 基于深度学习模型的显著性检测

在深度学习中,数据集的选取和质量对于训练所得的模型起着至关重要的作用。不同的数据集有不同的图像数量和分辨率,选取高质量、大容量的数据集可以使最终的显著性检测模型在应用中获得更好的效果。此外,由于数据集的差异性,同一种显著性检测模型在不同的数据集上往往有着不同的效果。因此,使用多个数据集来测试才能更好地反映该模型的性能。目前常用的数据集有ECSSD,DUT-OMRON,SOD,HKU-IS等,它们的图像数量及物体个数如表3所列。

表3 常用数据集的比较

Table 3 Comparison of common data sets		
Data set	Number of images	Number of objects
ECSSD	1 000	Multiple
DUT-OMRON	5 168	Multiple
SOD	300	Multiple
ASD	1 000	single
MSRA10K	1 000	single
PASCAL-S	850	Multiple
HKU-IS	4 447	Multiple

研究人员发现,将深度神经网络应用在视觉识别任务上时,可以取得非常良好的效果。受此启发,Li等^[21]首次将深度神经网络应用于显著性检测,提出了一种基于多尺度特征的深度卷积神经网络显著性检测模型。首先构建3个子网络共享的一个输出层和两个完全连接的隐藏层,然后通过矩形区域对图像进行3种不同尺度的特征提取,分别为被选中的区域、其相邻局部区域以及整个图像。然后将提取的3种不

同尺度的特征输入到网络中,使用一组训练图像和其标记的显著性图像对神经网络进行训练,并对输入的测试数据进行显著性图像输出。最后将生成的各个尺度上的显著图像进行融合,从而得到最终的显著性图像。

为了降低计算的时间复杂度,Zhao 等^[22]提出一种基于多语境的深度学习显著性检测模型,考虑到了全局上下文信息和局部上下文信息。如图 4 所示,首先对输入图像进行分割,将分割得到的各个区域作为矩形框的中心;其次设置两个矩形框,一种是尺寸为整个图像大小的矩形框,另一种是尺寸为图像 1/3 大小的矩形框;然后进行归一化处理,并分别将其应用于基于全局上下文和基于局部上下文的卷积神经网络模型中;最后利用回归器对二者的输出进行融合,送入网络模型

中得到了最终的显著性图像。与上一种方法相比,该算法的检测时间得到了大幅度缩短。

Wang 等^[23]提出一种将局部估计与全局搜索相结合的显著性检测算法。在局部估计阶段,基于邻域内的低级特征,使用由 3 个卷积层和 3 个全连接层组成的深度神经网络(DNN-L)来检测局部显著性,然后结合高级目标搜索对局部显著性图像做进一步细化。在全局搜索阶段,将局部显著性图像与全局对比度和几何信息一起用作对全局特征的描述,提取每个目标候选区域的 72 维特征来训练由 6 个全连接层组成的另一个深度神经网络(DNN-G),从而在全局层面对显著值做出预测。最后,根据各个显著对象区域的显著值,生成最终的显著性图像。

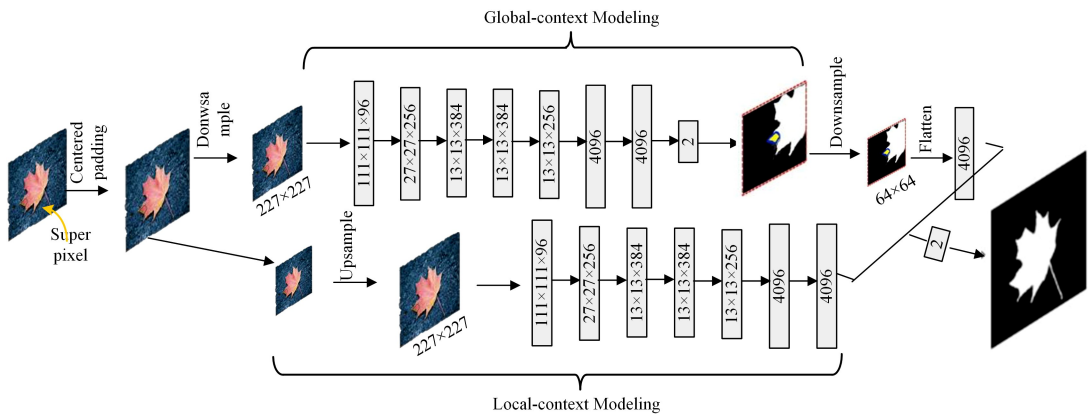


图 4 多语境的深度学习显著性检测模型^[22]

Fig. 4 Multi-context deep learning significance detection model^[22]

以上几种基于深度学习模型的显著性检测算法大多仅采用 CNN 提取图像中的高级特征,放弃了之前被研究证明有效的各种先验信息;其次由于卷积神经网络模型的层数较多,最终得到的显著性图像往往具有边界不清晰、模糊的特点。

为了解决显著性检测中显著图模糊的问题,Li 等^[24]提出一种端到端的深度对比度网络结构。该模型只需要在输入图像上运行一次,就能够得到与其有相同分辨率的显著图。该网络模型由两部分组成,分别是像素级全卷积流(Pixel-level fully Convolutional Stream)以及区域级空间池化流(Segment-wise Spatial Pooling Stream)。前者直接从输入图像得到像素级显著图,能够推断出显著性对象的语义属性;后者能有效地提取分段特征,对区域边界有更精确的描述。最后使用条件随机场(CRF)将两个显著性图像有效地结合起来,得到最终的显著性图像。

Liu 等^[25]提出一种深层卷积网(DHSnet),首先对全局的各种显著性特征进行自动学习,对输入图像进行全局上的大致预测,生成初始的显著性图像;然后通过提出的分层递归卷积神经网络(Hierarchical Recurrent Convolutional Neural Network),融合局部上下文信息,并逐步对显著性图像细节进行细化。该模型呈现出从全局到局部、从整体粗分到逐步精细的特点,最终得到了较好的检测效果。

为了将先验信息应用于显著性检测模型, Lee 等^[26]提出了一种将低级特征和高级特征相结合的显著性检测算法。通过编码低级距离图(Encoded Low Level Distance Map)来编码图像中的低级特征,从而对高级特征起到补充作用。高级特征和低级特征通过多级全连接神经网络进行组合,得到最终的显著性图像。

Hou 等^[27]提出了在整体上嵌套边缘检测算法(Holistically-nested Edge Detection)的基础上,引入跳过层的短连接(Short Connections)的显著性检测算法。通过从较深层到较浅层的一系列短连接,可以将高级特征应用到较浅层,从而更好地对显著性区域进行定位。针对生成的显著性图像过于平滑、边界不清晰的问题,采用全连接的条件随机场(CRF)进行优化,最终生成高质量的显著性图像。

Zhang 等^[28]提出了一种用于显著性物体检测的双向消息传递模型,该模型通过集成多级的特征来实现显著性检测。如图 5 所示,首先通过采用多尺度上下文感知特征提取模块(Multi-scale Context-aware Feature Extraction Module, MCFEM)来提取多级特征,并获取上下文信息;其次在多级特征之间设置双向结构来传递信息。通过这种结构,较深层中的高级特征可以被传递到较浅层,而较浅层中的低级特征以相反方向传递到较深层中,高级与低级特征能够合并到每个级别上。与此同时,利用门函数(Gate Function)控制信息

传递速率,在传递有用特征的同时放弃多余的特征,然后通过堆叠具有不同感受野的扩张卷积层来预测生成多个显著性

图,最后将预测生成的显著图像进行融合,从而得到最终的显著性图像。

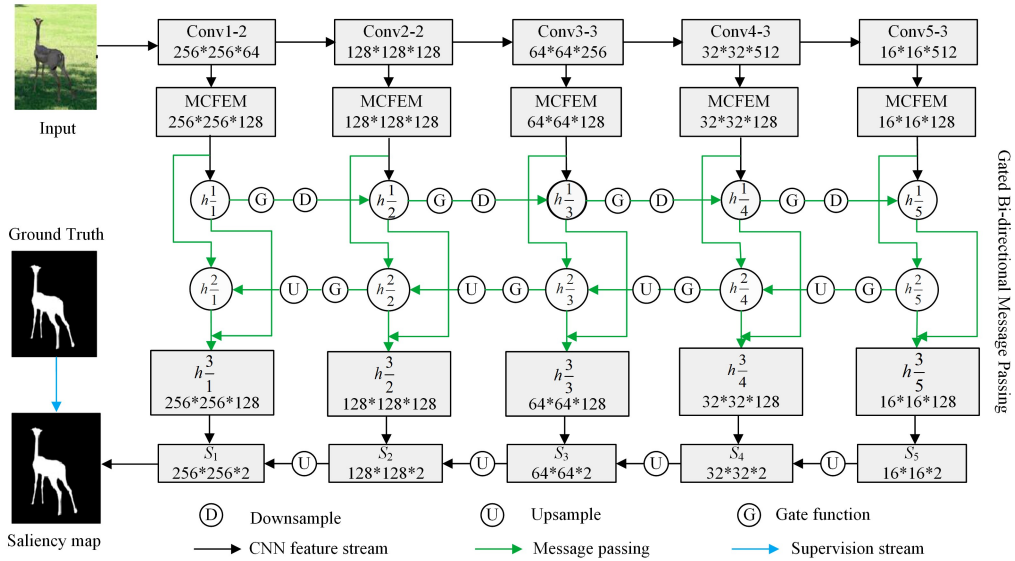


图5 双向消息传递显著性检测模型 [28](电子版为彩色)

Fig. 5 Two-way information-transferring modeling

5 发展动态分析

通过视觉图像显著性检测的一系列研究,本文针对上文介绍的3类方法的发展动态进行以下分析。

(1)由于视觉显著性检测在学术上并没有准确普遍的定义,因此显著性检测所利用的特征往往来源于前人的经验或是研究者自身的理解。根据这些特征检测生成的显著性图像往往具有局限性,常用的特征就是颜色、亮度、距离、方向等低级或中级特征。若能对图像中的特征进行进一步扩展与探索,显著性检测会得到更加良好的结果,算法对显著性对象识别的准确性与鲁棒性也会得到很大程度的提升。

(2)当面对复杂场景时,人眼通过自下而上和自上而下相结合的视觉注意机制,可以自动快速地定位显著性对象区域。但现有的显著性检测算法大多只模仿一种机制,或是依据低级、中级特征自下而上检测显著性对象,或是通过机器学习提取高级特征,自上而下地生成显著性图像。然而,传统的显著性检测方法在复杂环境中由于低级特征的限制存在无法达到预期效果的问题;仅使用高级特征的深度学习模型难以对显著性对象进行定位,因此将高级特征与低级特征相结合进行检测是当前研究的趋势。

(3)基于局部对比度的显著性检测算法由于侧重于各个像素点所在的局部区域,因此显著性对象的边缘往往有着较高的显著值,而对对象内部的显著值较低。为了使显著性对象能被均匀、清晰地突出,将局部特征与全局特征进行融合是更好的选择。此外,输入图像往往含有小尺度和高对比度的区域,因此很难提取出高质量的单一尺度的显著性对象区域,故基于多尺度检测的图像融合技术得到了越来越多的应用。

(4)采用卷积神经网络进行显著性检测时,往往会先将图像分割成若干块区域,这样虽然可以降低检测的时间复杂度,但得到的显著图往往缺乏空间连续性,如何在加强空间连续

性的同时保证不损坏显著性图像的质量,亟待研究人员的解决。

结束语 本文对现有的显著性检测算法进行了综述。通过对国内外显著性检测研究的分析可以看出,深度学习模型在复杂的环境中也可以取得良好的结果,在该领域得到了越来越多的应用。然而,研究人员在通过不断完善显著性检测模型使其能够更加准确、高质量地检测出显著性对象的同时,还需要考虑如何降低时间复杂度和提高计算速度。总之,高效且高质地对显著性图像进行提取,是优质的显著性检测算法不变的追求。因此,对图像显著性检测的进一步探索和研究对计算机视觉的发展具有重要意义。

参考文献

[1] HAN J, NGAN K, LI M, et al. Unsupervised extraction of visual attention objects in color images[J]. IEEE TCSV, 2006, 16 (1): 409-416.

[2] KO B, NAM J. Object-of-interest image segmentation based on human attention and semantic region clustering[J]. Jopt Soc Am, 2006, 23(10): 409-414.

[3] HADIZADEH H, BAJIC I V. Saliency-aware video compression [J]. IEEE Transactions on Image Processing, 2014, 23(1): 19-33.

[4] LIU F, SHEN T S, LOU S L, et al. Deep network saliency detection based on global model and local optimization[J]. Acta Optica Sinica, 2017, 37(12): 1215005.

[5] ZHENG L, WANG S J, LIU Z Q, et al. Fast imageretrieval: Query pruning and early termination[J]. IEEE Transactions on Mul-timedia, 2015, 17(5): 648-659.

[6] DESIMONE R, DUNCAN J. Neural mechanisms of selective visual attention [J]. Annual Review of Neuroscience, 1995, 18(1): 193-222.

[7] ITTI L, KOCH C, NIEBUR. A model of saliency-based visualat-

- tention for rapid scene analysis[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20(11): 1254-1259.
- [8] MA Y F, ZHANG H J. Contrast-based image attention analysis-by using fuzzy growing[C]//Proceedings of the 11th ACM Conference on Multimedia. New York: ACM, 2003: 374-381.
- [9] ACHANTA R, ESTRADA F, WILS P, et al. Salient region detection and segmentation[C]//Proceedings of the 6th International Conference on Computer Vision Systems. Berlin: Springer-Verlag, 2008: 66-75.
- [10] GOFERMAN S, ZELNIK-MAMOR L, TAL A. Context-aware saliency detection[C]//Proceedings of the 23rd International Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society, 2010: 2376-2383.
- [11] ACHANTA R, HEMAMI S, ESTRADA F, et al. Frequency-tuned salient region detection[C]//Proceedings of the 22nd Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society, 2009: 1597-1604.
- [12] ZHAI Y, SHAH M. Visual attention detection in video sequences using spatiotemporal cues[C]//Proceedings of the 14th ACM Conference on Multimedia. New York: ACM Press, 2006: 815-824.
- [13] CHENG M M, ZHANG G X, MITRA N J, et al. Global contrast-based salient region detection[C]//Proceedings of the 24th International Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society, 2011: 409-416.
- [14] PERAZZI F, KRAHENBUL P, PRITCH Y, et al. Saliency filters: Contrast based filtering for salient region detection[C]//Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition. United States: CVPR, 2012: 733-740.
- [15] SHI K, WANG K, LU J, et al. PISA: Pixelwise Image Saliency by Aggregating Complementary Appearance Contrast Measures with Spatial Priors[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Guangzhou: IEEE Computer Society, 2013: 2115-2122.
- [16] LIU T, SUN J, ZHENG N, et al. Learning to detect a salient object[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Xi'an: IEEE Computer Society, 2007: 1-8.
- [17] RAHTU E, KANNALA J, SALO M, et al. Segmenting Salient Objects from Images and Videos[C]//Proceedings of European Conference on Computer Vision. Heraklion: ECCV, 2010: 366-379.
- [18] JIANG H Z, YUAN Z J, CHENG M M, et al. Salient object detection: A discriminative regional feature integration approach [C]//Proceedings of the 26th International Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society, 2013: 2083-2090.
- [19] MAI L, NIU Y, LIU F. Saliency aggregation: a data driven approach[C]//Proceedings of 2013 IEEE Conference on Computer Vision and Pattern Recognition. Portland, OR, USA: IEEE Computer Society, 2013: 1131-1138.
- [20] LU H C, TONG N, ZHANG X N, et al. Co-Bootstrapping saliency[J]. IEEE Transactions on Image Processing, 2017, 26(1): 414-425.
- [21] LI G, YU Y. Visual Saliency Based on Multiscale Deep Features [C]//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA: IEEE Computer Society, 2015: 5455-5463.
- [22] ZHAO R, OUYANG W, LI H, et al. Saliency detection by multi-context deep learning[C]//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA: IEEE Computer Society, 2015: 1265-1274.
- [23] WANG L, LU H, RUAN X, et al. Deep networks for saliency detection via local estimation and global search[C]//Proceedings of the 28th International Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society, 2015: 3183-3192.
- [24] LI G, YU Y. Deep contrast learning for salient object detection [C]//Proceedings of the 29th International Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society, 2016: 478-487.
- [25] LIU N, HAN J. Dhsnet: Deep hierarchical saliency network for salient object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV: IEEE Computer Society, 2016: 678-686.
- [26] LEE G, TAI Y W, KIM J. Deep saliency with encoded low level distance map and high level features[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA: IEEE Computer Society, 2016: 660-668.
- [27] HOU Q, CHENG M M, HU X, et al. Deeply supervised salient object detection with short connections[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI: IEEE Computer Society, 2017: 5300-5309.
- [28] ZHANG L, DAI J, LU H C, et al. A Bi-directional Message Passing Model for Salient Object Detections[C]//Proceedings of The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Salt Lake City, UT, USA: IEEE Computer Society, 2018: 1741-1750.
- [29] WU J Y. Review of Bottom-up Salient Object Detection[J]. Computer Science, 2019, 46(3): 49-52.



YUAN Ye, born in 1998, postgraduate. His main research interests include computer vision and so on.



WEI Ming-qiang, born in 1985, Ph.D., associate professor. His main research interest is computer graphics with an emphasis on smart geometry processing.