DOI：10.19789/j.1004-9398.2024.06.005
文献引用：刘铁，陈楠，张瀚丹，等．视觉显著物体检测综述：方法、挑战及未来[J]．首都师范大学学报（自然科学版），2024，45（6）：36-48. LIU T，CHEN N，ZHANG H D，et al. Review of salient object detection：methods，challenges and directions[J]. Journal of Capital Normal University（Natural Science Edition），2024，45（6）：36-48.

视觉显著物体检测综述：方法、挑战及未来*

刘　铁**，陈　楠，张瀚丹，尚媛园，丁　辉，邵珠宏
（首都师范大学信息工程学院，北京　100048）

摘要：视觉显著物体检测作为计算机视觉领域的关键研究方向，也是学术研究的热点之一。本文系统性地梳理了该领域的研究方法、面临的挑战和未来的发展方向。首先，概述了视觉显著物体检测的发展脉络，以及其在计算机视觉领域的广泛应用；其次，对视觉显著物体检测方法进行了详尽的回顾，涵盖了基于显著性特征以及深度学习框架下的检测方法；再次，深入探讨了基于传统卷积神经网络和全卷积神经网络的显著物体检测方法，以及基于注意力机制的显著物体检测方法，并对视觉显著物体检测领域常用的数据集和评价指标进行了介绍；从次，针对当前视觉显著物体检测面临的挑战，如现有数据集的局限、复杂场景下的检测准确度等，文章进行了总结分析；最后，展望了视觉显著物体检测的未来发展方向。通过本文的综述，旨在为从事视觉显著物体检测的研究者提供全面而深入的参考，以促进该领域的进一步发展。

关键词：显著物体检测；计算机视觉；深度学习；注意力机制
中图分类号：TP391.4　　**文献标志码**：A

Review of salient object detection：methods，challenges and directions*

LIU Tie**，CHEN Nan，ZHANG Handan，SHANG Yuanyuan，DING Hui，SHAO Zhuhong
（Information Engineering College，Capital Normal University，Beijing　100048）

Abstract：Salient object detection，as a key research direction in the field of computer vision，is also one of the hotspots of academic research. In this paper，we systematically sort out the research methods，challenges and future development directions in this field. First，the development of salient object detection is summarized，as well as its wide application in the field of computer vision. Second，a detailed review of salient object detection methods is provided，covering detection methods based on saliency features as well as those under the deep learning framework. Third，salient object detection methods based on traditional convolutional neural networks and full convolutional neural networks，as well as salient object detection methods based on the attention mechanism，are discussed in depth，and commonly used datasets and evaluation metrics in the field of salient object detection are introduced. Again，the article summarizes and analyzes the current challenges of salient object detection，such as the limitations of existing datasets and the detection accuracy in complex scenes. Finally，it looks forward to the future development direction of salient object detection. Through this review，this article

收稿日期：2024-04-11
*北京市自然科技基金面上项目（4242034）
**通信作者：tliu@cnu.edu.cn

aims to provide a comprehensive and in-depth reference for researchers engaged in salient object detection in order to promote the further development of this field.

Keywords:salient object detection;computer vision;deep learning;attention mechanism

CLC:TP391.4　　**DC**:A

0　引　言

选择性视觉注意力机制指人类和其他动物能够在任何场景中快速地找到最醒目或是最引起注意的事物,进而排除无关和干扰信息。在计算机视觉处理中,借鉴选择性视觉注意力机制,可以有效减少视觉计算需要处理的数据量大小,从而加快整体处理速度。视觉显著物体检测(salient object detection,SOD)自2007年被提出以来,已发展成为计算机视觉领域的重要研究方向之一[1]。其主要任务是检测图像中视觉显著物体的区域并对其进行像素级标识[2],如图1所示,第1行为原始图像,第2行为这些图像对应的显著物体的真值(ground truth,GT),使用二值图像表示真值,白色部分为检测到的显著物体,而黑色部分为背景。

视觉显著物体检测为许多计算机视觉应用提供了重要的信息和支持,广泛应用于各种视觉任务的预处理过程,例如目标检测[3-4]、目标识别[5]、图像分割[6-7]、目标跟踪[8]、关键点定位[9]、视觉跟踪[10]、图像缩略图[11]和图像重定向[12]等。除了计算机视觉领域,显著物体检测在其他领域中也有许多应用,例如自动驾驶、航天航空和生物学相关领域等。在自动驾驶领域,研究者们通过视觉显著物体检测算法为无人驾驶汽车提供外部环境信息和路况信息[13];在航天航空领域中,研究者利用显著物体检测算法对空中加油目标识别和定位[14];在生物学相关领域,也有使用显著物体检测算法对相近生物的纹理特征进行提取,从而进行生物分类的方法[15]。此外,视觉显著物体检测还被广泛应用于医学图像分割[16]和文物识别[17]等领域,其发展兼具学术意义与应用价值。

随着显著物体检测任务的发展和应用,SOD已经从三原色(red green blue,RGB)图像检测单个显著物体扩展到基于RGB-D的显著物体检测[18-19]、协同显著物体检测[20-21]和视频显著物体检测[22-23]等领域。由于基于RGB的显著物体检测方法仍然存在不足,在一些复杂场景中仍然难以准确识别和定位显著物体,一些研究者开始引入深度信息来解决复杂场景下的显著物体检测问题,提出了多种基于RGB-D的显著物体检测方法,随着卷积神经网络和Transformer的不断发展,相继被引入到基于RGB-D的显著物体检测中,并取得不错的效果。协同显著物体检测是显著物体检测领域中一个快速发展的延伸,不同于单幅图像显著物体检测,其目标为检测出一组图像中共同出现的显著目标,深度学习的快速发展也不断推动着这一领域的发展。近年来,随着短视频文化的迅速发展,视频数据分析及处理也成为了如今计算机视觉领域的研究热点之一,视频显著物体检测旨在识别和分割视频中最具视觉冲击力的物体,对动态场景中不同时刻的同一显著目标进行跟踪和检测,视频显著物体检测中各帧内容高度相关,类似于协同显著物体检测。

自视觉显著物体检测被提出以来,已发展了许多的方法分支。大致可以分为传统的显著物体检测方法、基于机器学习的方法、基于深度学习的方法、结合注意力机制的方法以及多模态融合的方法

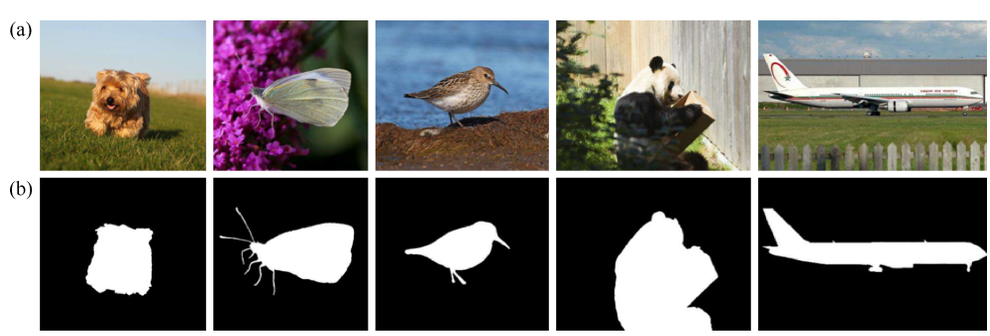


图1　视觉显著物体检测

(a)图像;(b)基准真值

等分支。最早的显著物体检测方法主要基于传统的计算机视觉方法,如图像特征提取、区域分割、对比度计算等,这些方法通常依赖于手工设计的特征和规则,准确性和鲁棒性有限;随着机器学习的发展,出现了一些基于机器学习的显著物体检测方法,如支持向量机(support vector machine,SVM)、随机森林(random forest,RF)等,这些方法可利用大量标注数据自动学习特征和模式,相比传统方法有了一定的提升;近年来,随着深度学习的兴起,基于深度学习的显著物体检测方法取得了巨大进展,这些方法利用深度神经网络可以学习到更加丰富和高级的特征表示,从而在显著物体检测任务中取得更好的效果;为了更加精确地检测,注意力机制在显著物体检测中得到了广泛应用,可以帮助网络集中注意力在图像中的显著区域,提高检测的准确性,包括空间注意力、通道注意力等;除了单一模态的图像,还有一些方法将多模态信息融合到显著物体检测中,以提高检测的准确性和鲁棒性。

本文简要介绍了视觉显著物体检测应用及方法的发展历史,回顾和总结传统显著物体检测、基于深度学习的显著物体检测以及基于注意力机制的显著物体检测的方法,介绍了视觉显著物体检测的实验数据集与主要评价标准,并分析现有显著物体检测方法存在的问题、面临的挑战以及未来可能的发展方向。

1 视觉显著物体检测方法

1.1 传统显著物体检测

传统的自底向上检测方法主要依赖低层次的视觉特征,如颜色、纹理和亮度等,通过颜色对比以及图像中心与边缘的差异生成显著图。Itti等[24]首次将认知心理学和神经科学领域中的视觉显著性概念引入计算机视觉;Liu等[1]将显著物体检测视为二值分割问题,结合局部、区域和全局特征表示显著物体,并使用这些特征训练条件随机场,对图像中的像素块进行显著性分类。此后,显著物体检测技术逐渐从依赖生物模型转向基于底层视觉特征的计算模型,大幅提升了检测精度。

传统显著物体检测算法根据其原理可大致分为基于局部对比的方法[25-26]、基于全局对比的方法[27-28]和基于背景先验的方法[29-30]。基于局部对比的显著物体检测算法通过使用一个或多个低级特征来确定元素相对于周围环境的对比度来区分显著区域。与基于局部对比的算法不同,基于全局对比的方法侧重于分析整个图像的全局结构和内容,局部对比度算子因其仅考虑有限的邻近区域,所以往往不能有效地识别较大范围的显著物体。因此,基于全局对比的方法需要采取优化技术进一步增强显著线索,如Cheng等[27]提出了一种生成大规模感知段的方法,采用直方图量化技术对高斯混合模型(gaussian mixture model,GMM)的外观线索进行采样,进而对模型组件进行聚类,以便识别出具有相似语义关系的图像区域,通过这2种软抽象技术,全局独特性和颜色空间分布的线索能够一致地强调整个显著对象的区域;Wei等[31]首次提出基于背景先验的显著物体检测方法,这种方法结合了图像边界先验和连通性先验,利用颜色先验[29]、中心先验[30]和形状先验[32]等知识,通过确定图像中的背景范围并抑制背景,从而划分显著区域。

1.2 基于深度学习的显著物体检测

随着深度学习模型在众多视觉任务中取得成功,基于深度学习的视觉显著物体检测逐渐成为主流方法。这些方法根据其网络结构通常分为传统卷积神经网络(convolutional neural network,CNN)和完全卷积神经网络(fully convolutional network,FCN)。基于CNN的显著物体检测通常采用一系列卷积层和池化层来提取图像的高级特征,其输入图像必须为统一尺寸,然后通过全连接层进行分类或回归,输出一个整体的显著性得分;基于FCN的显著物体检测将全连接层替换为全卷积层,从而可以处理任意尺寸的输入图像,保留空间信息,并生成相同尺寸的显著图,适用于像素级别的任务。

1.2.1 基于CNN的显著物体检测

基于CNN的显著物体检测方法通过将图像输入到卷积神经网络进行特征提取,捕获图像中的语义信息和视觉特征,然后结合先验信息如颜色对比度、边缘等,利用显著物体检测模型生成显著图,进而标识出图像中的显著物体区域。Wang等[33]结合了深度神经网络的局部估计和全局搜索,该方法使用深度神经网络进行局部显著性检测,并通过探索对象概念对局部显著图进行优化,全局搜索阶段结合局部显著图、全局对比度和几何信息描述对象候选区域;Zhao等[34]提出了一种多语境深度学习框架,联合建模全局和局部语境,利用了上采样和下采样技术,引入上下文信息;Li和Yu[35]提出利用

CNN提取多尺度特征学习高质量视觉显著物体检测模型的方法,在CNN的基础上添加全连接层用于实现3个不同尺度上的特征提取,引入空间一致性改进方法,最后将不同级别图像分割计算得到的多个显著性图进行聚合,得到了比单个分割更好的显著性图;He等[36]提出了基于超像素的卷积神经网络方法SuperCNN,该方法能够学习分层对比度特征,恢复超像素之间的上下文信息,降低密集标记图的预测数量,以及利用多尺度网络结构独立于区域大小进行显著性检测。

1.2.2　基于FCN的显著物体检测

基于FCN的显著物体检测方法采用完全卷积网络结构,通过将卷积神经网络中的全连接层替换为适用于密集预测的卷积层和上采样层,实现对输入图像进行端到端的像素级别显著物体预测,从而能够提取图像的高级语义特征,还能够保留输入图像的空间信息,更有效地捕获显著目标的空间分布和形状。与CNN方法相比,基于FCN的方法无需对图像进行固定大小的裁剪或缩放,能够直接处理任意大小的输入图像,并且能够在像素级别上进行密集的显著性预测,因此在显著物体检测任务上通常具有更好的性能和泛化能力。Liu和Han[37]提出了一个名为DHSNet的端到端显著物体检测网络,利用全局结构信息生成粗略显著图,再通过层次结构的循环神经网络HRCNN逐步细化显著图,整合局部上下文信息;Feng等[38]设计了注意力反馈模块,进一步采用边界增强损失函数来学习清晰的边界轮廓;Zeng等[39]提出了一个高分辨率显著物体检测数据集,通过全局语义网络、局部细化网络和全局局部融合网络结合全局语义信息和局部高分辨率细节;Wu等[40]提出了一种方法去改善显著物体检测的不完整性和边界不准确性问题,利用前景轮廓检测和边缘检测的监督来训练显著物体检测网络,结合显著物体检测和前景轮廓检测任务生成统一突出的显著性图,再通过前景轮廓和边缘检测任务相互指导,实现更精确的前景轮廓预测和较少局部噪声的边缘预测;Wang等[41]提出了一种迭代协作方式的显著物体检测方法,整合自顶向下和自下而上2个过程,通过网络层的递归机制实现逐步优化,有效地促使显著性信息在自下而上、自顶向下及层内流动;Su等[42]提出了一种图像连续扩展的边界感知网络以提高显著物体的内部一致性和边界清晰度。

1.3　基于注意力机制的显著物体检测

在人眼视觉系统中,通常综合多个区域或物体特征形成对眼前场景的整体感知,受这种生理机制的启发,学者们在计算机视觉算法研究中提出了注意力机制,即结合多个区域间的有效信息、排除无关区域干扰,通过为不同信息分配优先级来提高系统性能[43]。

在显著物体检测中,注意力机制通常被用于学习有效的特征。You等[44]开发了一种结合视觉特征和语义概念的语义注意力模型,应用于图像标题生成任务;Yang等[45]提出了用于图像问答的堆叠注意网络;Chen等[46]和Li等[47]通过注意力模型来自适应地挑选多尺度特征;Kuen等[48]使用递归神经网络,在每一个步骤中生成图像区域的注意力图,从而逐步优化显著性图;Hu等[49]构建了一种用于通道注意力的“压缩-激励”模块;Chu等[50]通过层次化的方式学习了全局注意力和局部注意力,并将其应用于姿势估计;Xu等[51]将空间位置信息和空间协方差特征显式地表达到自注意力机制中,以学习空间敏感的全局上下文;Wang等[52]设计了残差注意力机制模块接收来自混合特征级联聚合模块的相邻阶段特征图,减少空间细节的损失以及由于改变物体形状、比例和位置对结果造成的影响;Ngo等[53]介绍了一种名为ETANet的高效三重注意力网络用于显著物体检测,该方法通过交叉空间通道注意力模块、轻量级感受野模块和自注意力模块以及膨胀的交叉融合模块,有效提取低层和高层特征,实现准确的目标检测,同时参数量较少;Li等[54]提出的网络通过混合注意力机制模块、前馈反馈单元模块以及改进的骨干网络来提取特征、融合多尺度特征,提高了检测性能;Pang等[55]通过设计空间和通道双视角的注意力机制,增强特征的表示能力,同时引入卷积网络弥补注意力机制的不足,充分证明了注意力机制在多模态任务中的有效性。

1.4　显著物体检测模型的学习范式

传统的显著物体检测方法依赖于手工特征或人类经验,但这些方法在复杂场景下往往无法获得高品质的显著性图。基于深度学习的显著物体检测模型已经得到了广泛研究,极大地提升了显著性检测的性能。然而,这些全监督模型的训练高度依赖大量标记数据,通常需要像素级标注,而大规模的逐像素标注过程非常耗时和费力。在显著物体检测任务中,学习范式起到了至关重要的作用。学

习范式是指模型在学习过程中采用的训练和优化方法,决定了模型如何从数据中学习,如何进行特征提取、优化参数以达到最佳性能。显著物体检测模型的学习范式主要包括监督学习、弱监督学习、无监督学习和增量学习4种类型。

为了实现标注效率和模型性能之间的平衡,研究人员提出了几种弱监督和无监督的方法,这些方法能够从稀疏标记数据中学习显著性,从而减少对大规模标注数据的依赖。Li等[56]提出了名为交替显著性图优化的弱监督方法,通过图像级监督学习来更新由无监督显著性方法生成的初始显著性图,通过条件随机场模型对初始显著性图、预测显著性图和类别激活图进行建模,纠正内部标注的模糊性;Zeng等[57]提出了利用多样弱监督源训练显著物体检测模型的统一框架,如类别标签、标题和未标记数据,通过分类网络和标题生成网络生成像素级伪标签,并设计注意力转移损失和注意力一致性损失来增强模型的性能;Nguyen等[58]提出了无监督对象显著物体检测方法,通过自我监督技术在多次迭代中细化由不同手工方法生成的嘈杂伪标签,并使用这些细化标签训练显著物体检测网络;Zhang等[59]提出了通过从"弱"和"嘈杂"无监督手工显著性方法生成的多个噪声标注中学习的无监督显著物体检测方法,通过端到端深度学习框架用于无监督显著性检测,包含潜在显著物体预测模块和噪声建模模块,协同工作并共同优化;Wang等[60]提出了仅使用图像级标签的弱监督方法,通过前景推理网络和全卷积网络联合训练,以捕捉前景区域并优化显著性图,从而减轻标注成本并提升检测性能。

2 视觉显著物体检测数据集和评价标准

2.1 视觉显著物体检测数据集

伴随着显著物体检测模型的不断创新,相应地涌现了大批专门的数据集以支持这些模型的训练和评估(表1[1,60-69])。众多公开可用的数据集和系统的评价标准体系是显著物体检测快速发展的重要原因之一。显著物体检测数据集的发展经历了从最初的简单矩形框标注到更为精确的像素级标注的持续演进。随着技术进步和研究深入,人们逐渐意识到简单的矩形框标注无法完全捕捉目标的边界和细节,因此逐渐转向使用像素级标注,以提供更准确和详细的目标位置信息。这种精确的标注方法使得显著物体检测算法能够更准确地定位和分割目标,从而提高算法的性能和应用场景的适用性。早期的数据集往往存在选择偏差,多数图像仅包含简单背景和单一显著物体。此外,数据集中的图像数量也相对有限。近年来,为提升数据集质量,研究人员开始采集具有多个目标且背景相对复杂且杂乱的图像。这种更新的数据采集方法有助于更全面地涵盖各种真实场景,并促进显著物体检测算法在更复杂情境下性能的提升。数据集相关信息如表1所示。

2.2 视觉显著物体检测评价标准

显著物体检测算法效果主要通过计算算法生成的显著图像与标注的真值图像之间的误差进行评价,具体公式为

$$\sum_{i=1}^{n} E(f(I_i), L_i), \tag{1}$$

式中:i表示n张图像中任意一张;L_i为该图像真值;$f(I_i)$表示显著物体检测算法生成显著图像的值,$f(I_i)\in[0,1]$;E为评价标准。为了对不同显著物体检测方法进行比较,确立一致的评价准则和评估指标至关重要。下文对显著物体检测的主要评价指标进行介绍。

(1)PR曲线,反映了查准率(precision)和召回率(recall)之间的关系。查准率表示检测出的正样本数与所有检测出的正样本数的比值,而召回率表示正确检测出的正样本数占所有正样本数的比例。PR曲线越靠近右上角,模型性能越好。这表明在较高的召回率下仍能保持较高的精确率。

(2)F值,是查准率和召回率的加权调和平均数。较高的F值表示显著物体检测模型的效果更加有效。通常,F值计算公式为

$$F = \frac{(1+\beta^2)P \times R}{\beta^2 \times P \times R}, \tag{2}$$

式中:β是一个用于调整查准率和召回率之间相对重要性的参数,当$\beta>1$时,更加重视召回率,当$\beta<1$时,更加重视查准率;P为查准率;R为召回率。

(3)平均绝对误差(mean absolute error,MAE),反映了模型检测结果与实际真值之间的误差,MAE越低,模型检测越精确,其公式为

$$E_{\mathrm{MA}} = \frac{1}{w \times h}\sum_{i=1}^{w}\sum_{j=1}^{h}\left| S_{ij} - G_{ij}\right| F_{\beta}^{\omega}, \tag{3}$$

式中:w为检测图像的宽度;h为高度;S为显著值,由检测模型计算得出;G为标注图像的真值。

(4)ROC(receiver operating characteristic curve),

表1　显著物体检测数据集

数据集名称	年份	规模	数据集描述
MSRA-A[1]	2007	20 840	图像包含一个矩形框来标注的显著物体,由3人共同标注作为真值,其中MSRA-B作为子集包含5 000张图像,由9个人同时标注以解决人为标注偏差的问题
SED[61]	2007	200	单个显著物体的子集SED1和两个物体的子集SED2,每个子集包含100张图像并具有像素级显著物体的真值
ASD[62]	2009	1 000	MSRA-A的一个子集,具有像素级标注
SOD[63]	2010	300	许多图像的多个显著物体与背景或与图像边界形成低色彩对比,具有像素级标注
MSRA-10K[64]	2011	10 000	从MSRA-A中选取的10 000张图像,进行像素级标注,MSRA10K由于其规模大,被广泛用于训练深层SOD模型
ECSSD[65]	2013	1 000	ECSSD数据集是由CSSD数据集扩张而来,每张图像都包含相对复杂的背景和结构化信息,图像具有像素级真值
DUT-OMRON[66]	2013	5 168	图像背景相对复杂、内容多样,不仅提供了高品质的像素级真值,还提供了5名观测者的眼动数据和目标的边界框。每张图像上都有一个到多个显著物体,是检验显著物体检测算法具有极高挑战性的一个数据集
PASCAL-S[67]	2014	850	该数据集主要根据人眼注视点进行标注,与其他显著性目标检测数据集的图像区别较大,相应的其难度也较大
HKU-IS[68]	2015	4 447	该数据集包含大量位于复杂场景下的多显著物体图像、包含不同空间分布情况下多个物体之间连接断开的图像,同时具有像素级标注
DUTS[60]	2017	15 572	图像采集自ImageNet Det训练集和验证集,测试集包含5 019张图像,图像选自ImageNet Det测试集和SUN数据集,该数据集包含多种多样的图像场景以及物体类别
SOC[69]	2018	6 000	包含80个常见类别,一半的图像包含显著物体,而其余的则没有;每个包含显著物体的图像都用实例级的真值掩码、对象类别和具有挑战性的因素进行注释

表示在不同分类阈值下真阳性率(true positive rate,TPR)与假阳性率(false positive rate,FPR)之间的权衡关系。在ROC曲线中,一般设置FPR为横轴,设置TPR为纵轴。ROC曲线越靠近左上角,模型检测效果越好。

(5)AUC(area under the curve of ROC),为ROC曲线下的面积。ROC是模型的二维性能表现图,AUC是一个用于量化模型性能的指标。AUC值越接近1,显著物体检测模型的性能越好。

(6)F_β^ω-measure,是一种新的评估方法[70],通过扩展4个基本量P_T、N_T、P_F和N_F来处理非二值化地图,并按检测错误的位置和邻域给予不同权重。文中定义了加权精确度(weighted precision)和加权召回度(weighted recall),前者是衡量准确性的指标,后者是衡量完整性的指标,其公式为

$$P^\omega = \frac{P_\mathrm{T}^\omega}{P_\mathrm{T}^\omega + P_\mathrm{F}^\omega}, \tag{4}$$

$$R^\omega = \frac{P_\mathrm{T}^\omega}{P_\mathrm{T}^\omega + N_\mathrm{F}^\omega}。 \tag{5}$$

定义F_β^ω-measure计算公式为

$$F_\beta^\omega = (1+\beta^2)\frac{P^\omega \cdot R^\omega}{\beta^2 \cdot P^\omega + R^\omega}。 \tag{6}$$

(7)S值,综合考虑了区域感知和对象感知的结构相似性[71]。该评估方法主要包括区域结构相似性度量和对象结构相似性度量,旨在更全面地评估前景物体,具体公式为

$$S_m = \lambda \times S_O + (1-\alpha) \times S_r, \tag{7}$$

式中:α为权重参数,取值范围为[0,1],通常设置为0.5;S_O表示结构相似性度量值;S_r为区域结构相似性度量值。

3　视觉显著物体检测面临的挑战

视觉显著物体检测是一个活跃且不断发展的领域,但其仍然面临着一些问题与挑战,如对于上下文信息的提取、对于特征的有效提取与聚合、损失函数的设计、数据集相关问题、实时性能等。视觉显著物体检测方法发展的同时伴随着数据集的建立,数据集在显著物体检测任务中发挥了重要的作用,用于模型训练的数据集质量将会影响模型的性能。随着显著物体检测方法的不断发展,对数据集的规模和质量的要求逐渐增高,而现有的显著物体检测数据仍存在规模小、数据集中的图像质量参差不齐等问题。具有较少偏差的大型可用数据集对于开发设计显著物体检测模型至关重要,本章节将主要针对数据集相关问题进行分析介绍。

3.1 数据集规模

显著物体检测数据集从最初的矩形框标注发展为像素级精确标注,但是由于精确标注的极大代价限制了数据集规模的增长,目前显著物体检测任务中像素级标注的单个数据集规模仍然没有超过最早由矩形框标注的MSRA-A数据集的规模。同时精确标注数据集还面临标注理解不一致和数据偏差等挑战。

3.2 数据偏差

Torralba等[72]表明当前计算机视觉领域的数据集中存在普遍的数据偏差问题,其中包括选择偏差、捕获偏差和负集偏差,这些问题也可以被引申至更具体的显著物体检测领域。

3.2.1 选择偏差

选择偏差的存在是由于数据在采集过程中未能考虑尽可能全面的数据类型,或对某类型的数据存在选择偏好,这使得数据集中存在大量同类型图片或品质相似的图像。同时文献[73-74]中的实验证明,目前常用的显著物体检测数据集的数据中普遍存在强烈的颜色对比度,这对模型的训练是十分不利的。复杂图像(颜色对比度较低、背景复杂、光线昏暗)的缺失和数据种类的匮乏会使模型难以通过数据学习到更多的知识,训练获得的模型将难以处理难度较大的数据。在一些真实场景的测试集下,受到复杂环境下光照、颜色相近、纹理相似、物体占据的区域、视角、遮挡以及背景干扰等影响,图像场景将会更加复杂。如果视觉显著物体检测模型没有学习过这些场景的图像特征,模型的有效性及鲁棒性会受到影响并在处理这些图像时出现误检。本文选取在精确标注数据集上训练好的SOC[69]、BASNet[75]和Tracer[76] 3个显著物体检测模型,在精确标注数据以外的数据上进行了测试。图2给出的部分结果,可知显著物体检测模型在左边3列简单图像数据上获得了比较一致的结果,但是在右边3列稍微复杂的图像上出现了预测不一致的结果。

3.2.2 捕获偏差

捕获偏差主要针对数据集中图像的构成问题,其中最常见的问题是中心偏差问题,即大部分数据集中的图像具有中心区域为显著区域的指向,而不是分布在整个图像中。这一问题可能会影响到模型的性能和泛化能力,使其更倾向于将注意力集中在图像中心,而忽略了其他可能的重要区域。

3.2.3 负集偏差

负集偏差是指当前现有的数据集中存在负集不够丰富甚至数据中不包含负集的问题。具体来说,大多数的数据集假设图像至少应该包含一个显著对象,因此这些数据集忽略掉了不包含任何显著对象的图像。常用的ECSSD[65]、DUT-OMRON[66]、PASCAL-S[67]、HKU-IS[68]、SOD[71]和DUTS[60]等数据集大多都基于图像中至少包含一个显著物体的假设,因此会忽略不包含任何显著物体的图像[69]。同时,标注者对某些特定图像或对象类型的个人偏

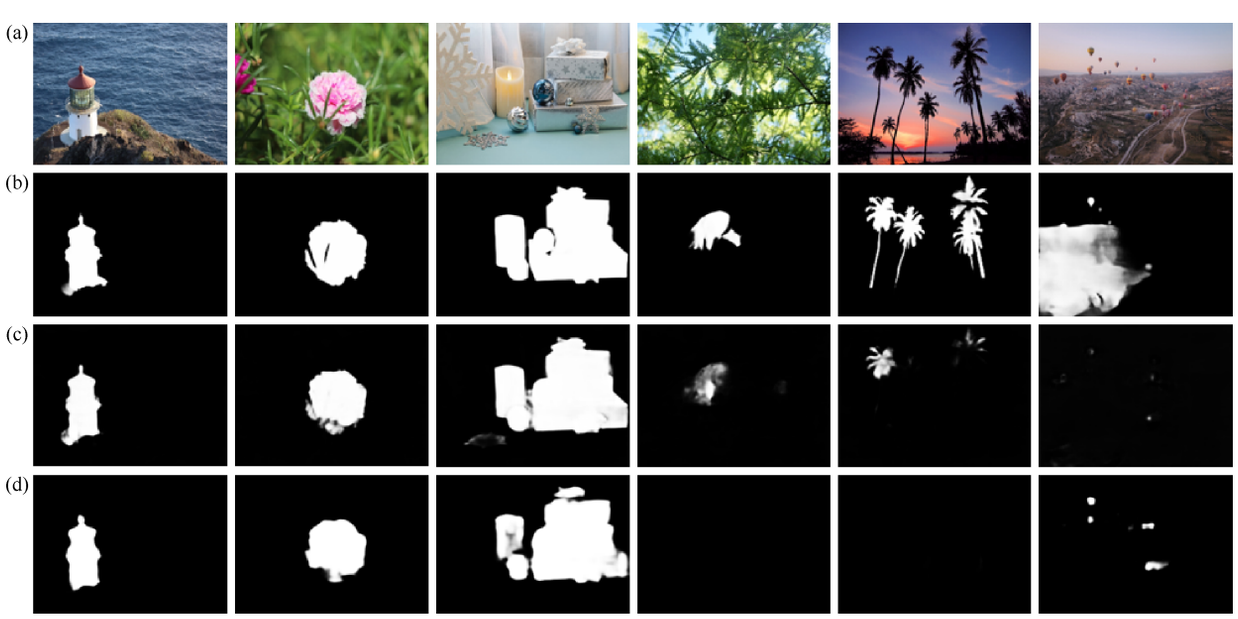


图2 显著物体检测模型结果

(a)原图;(b)BASNet结果;(c)SOC结果;(d)Tracer结果

好可能会影响数据集中的真值,这种情况也将造成负集偏差。如图3(a)中不同数据集[65-68]对自行车和蝴蝶的标注受到标注人理解的影响,同样在图3(b)中数据集[64]对多个物体或者大面积前景的标注值得斟酌。

3.3　模型“灾难性遗忘”问题

现有的视觉显著物体检测算法主要在规模有限的标注数据集上进行训练和测试,且大多数已有方法的设计都基于模型应用于闭集场景的假设,即测试集、无标签训练集数据中仅包含有标签训练集中出现过的类别。基于这种假设的研究都集中在开发针对特定数据集或者环境的专用模型,但是在闭集上表现良好的模型在开集数据上进行测试时很难达到同样的效果。在现实的应用场景之中存在着大量的未知图像,且这些图像的数据将不断增加,当前业界缺少能够持续处理这些大规模数据的有效方法。若模型想要检测这些大规模图像,最直接的方法是进行精确的人工标注和耗时的重新训练,这样不仅消耗大量的人力、物力且所获得的模型无法在持续学习的环境中工作,也就是说当模型需要逐步地学习新数据时,将会遗忘以前所学过的知识,模型在旧数据上的性能下降,造成“灾难性遗忘”的问题。

为适应开集场景下视觉显著物体检测模型发展的需求,最直接的想法是将在精确标注数据上训练的模型和学习方法迁移到开集场景中,但模型进行强监督学习面临着训练所需的数据精确标注成本较高且标注不一致的问题[2],很难获得大规模数据的真实标签进行强监督学习[75],为当前依赖于精确标注的显著物体检测模型和学习方法带来挑战。

综上所述,当前视觉显著物体检测算法主要在有限规模的标注数据集上进行训练和测试,假设测试集和训练集中的数据仅包含训练集中出现过的类别,即闭集场景。然而,在现实世界的开集场景中,存在大量未知图像,这些图像的数据不断增加,而当前的算法缺少能够持续处理这些大规模数据的有效方法。此外,用于训练模型的数据集的品质也会影响模型的性能,但目前可用的视觉显著物体检测数据集在一定程度上制约了算法的发展和应用。在开集场景中,图像场景更加复杂,如果模型没有学习过这些场景的图像特征,其有效性和鲁棒性会受到影响。因此,如何适应开集场景下视觉显著物体检测模型的发展需求,以及如何处理不同模型预测结果不一致的图像,成为了亟待解决的问题。

4　视觉显著物体检测的未来与发展

视觉显著物体检测任务自提出以来已经发展成为一个独立的研究分支。复杂场景下的显著物体检测可以作为其他视觉认知任务的基础,例如在视觉系统检测出显著物体后再进行后续的目标识别能够缩小后续任务的搜索范围。针对上一章节提出的视觉显著物体检测面临的问题与挑战,本章节提出了一些未来可能的应对策略与发展方向。

(1)开集场景下显著物体检测数据集的构建。随着互联网技术的快速发展和广泛应用,可获得的图像数量在不断激增,社会各部门也积累了大量数据,这使得开集场景下数据集的构建变得尤为重要,而在开集场景下的视觉显著物体检测也将具有较广的应用场景。开集场景中将包含室内和室外等更多复杂场景的数据,同时这些数据对图像中显著物体的存在与否及个数不作限制,即图像中可能包含多个显著物体,或者图像中并不包含显著物体,而是背景或非显著物体占据了整个场景,比如天空、草地或纹理等无定形的背景图像。

(2)持续从大规模未标注数据集学习经验知识。智能体的视觉感知系统需要不断学习和调整以适应复杂多变的环境,如何不断地从大规模未精确标注的图像中学习经验知识是一项极具挑战的任务,例如,视觉感知系统需要检测前方通道或者道路上是否具有障碍物,而系统很难一次收集所有

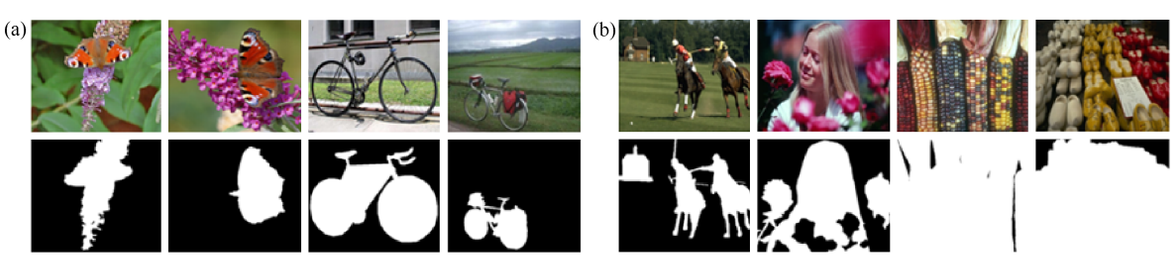


图3　像素级精确标注数据集示例

(a)数据集标注不一致[2];(b)多个物体或者大面积前景标注

的场景和样本进行训练。对自动驾驶场景而言,视觉感知模型需要在能够应对之前学习过场景的同时,在遇到不同天气状况时也能够及时适应新的环境。增量学习是指模型在不断处理连续信息流的过程中,能够学习新知识并保留整合、优化旧知识的能力。随着深度学习的不断发展,用于模型训练的数据量不断扩增,如何使神经网络能够从非平稳数据流中进行增量学习成为深度学习中的重要问题和研究趋势。随着增量学习方法的快速发展,增量学习已被广泛地用于图像分类、目标检测和语义分割等计算机视觉领域中。然而,在显著目标检测领域还少有该类研究,因此将增量学习策略引入到显著物体检测领域中去,将为显著物体检测领域研究打开新的思路、提供更多的可能性以面对开集场景的挑战。此外,当前评价显著物体检测方法的评价指标依然在精确标注数据集上采用F值或者MAE等来进行方法的比较,在增量学习的过程中,模型在学习新的数据域时会不断地迭代更新,那么中间场景下产生的这些模型,若无精确标注的真值图该如何进行评价,以及模型学习过程中的遗忘率,也是值得探究的问题。

(3)显著物体检测与现实应用场景的结合。显著物体检测任务作为计算机视觉领域的基础研究任务,在许多实际应用中都具有重要的作用,如自动驾驶、视频监控、图像检索等。然而,现有的显著物体检测方法大多基于有限规模的闭集数据集进行训练和测试,这在现实世界的开集场景中遇到了性能下降的问题。通过引入更多的实际场景数据,提高显著物体检测模型的泛化能力和鲁棒性。在自动驾驶领域,显著物体检测技术可以帮助车辆快速识别道路中的关键物体,如行人、车辆和交通标志等,从而提高驾驶的安全性和效率。在视频监控领域,显著物体检测技术可以帮助监控系统快速定位和跟踪感兴趣的目标,提高监控的效率和准确性。在图像检索领域,显著物体检测技术可以帮助用户快速找到感兴趣的目标图像,提高图像检索的效率和用户体验。因此,开集场景下显著物体检测方法的研究对于推动计算机视觉技术的应用和发展具有重要的理论和实际意义。

5 总 结

作为计算机视觉领域的研究热点之一,显著物体检测自提出以来已经逐渐发展成为一个独立的研究分支。复杂场景下的显著物体检测可以作为其他视觉认知任务的基础,例如在视觉系统检测出显著物体后再进行后续的目标识别,能够缩小后续任务的搜索范围。此外,视觉显著物体检测通常与眼动检测相结合,用于心理学、医学诊断等相关领域,其发展具有重要的现实意义。

本文回顾了显著物体检测自提出以来的大量文献,对显著物体检测数据集、方法及其评价指标进行了综述,并就其现有问题和未来发展方向进行讨论。视觉显著物体检测领域中仍然存在一些难点问题,例如数据集的偏置问题、复杂场景中显著物体检测问题、深度学习网络构建问题,以及显著物体检测模型和人类视觉机制之间的差异问题等。目前,深度学习因其较高的检测精度与适应复杂场景的能力成为了视觉显著物体检测的首选方法。展望未来,视觉显著物体检测的研究将更加注重开集场景下数据集构建、网络模型轻量化、检测精度提升以及实时性增强等方面的优化。视觉显著物体检测作为多种视觉任务的前置预处理步骤,将广泛地服务于各类实际应用,进一步促进计算机视觉技术的应用和发展。

参 考 文 献

[1] LIU T, YUAN Z, SUN J, et al. Learning to detect a salient object[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(2): 353-367.

[2] WANG W, LAI Q, FU H, et al. Salient object detection in the deep learning era: an in-depth survey [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 44(6): 3239- 3259.

[3] KUMAR N, SARDANA H K, SHOME S N. Saliency based shape extraction of objects in unconstrained underwater environment [J]. Multimedia Tools and Applications, 2019, 78 (11): 15121-15139.

[4] JIA Y, YUAN J, WANG J J, et al. A saliency-based method for early smoke detection in video sequences[J]. Fire Technology, 2016, 52(5):1271-1292.

[5] FU J L, ZHENG H L, MEI T. Look closer to see better: recurrent attention convolutional neural network for fine-grained image recognition[C]// DENG J, JIA J, SVETLANA L, et al. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA: IEEE, 2017: 4476-4484.

[6] TU Z Z, ABEL A, ZHANG L, et al. A new spatio-temporal saliency-based video object segmentation[J].

Cognitive Computation, 2016, 8(4): 629-647.
[7] WANG W G, SHEN J B, YANG R G, et al. Saliency-aware video object segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(1): 20-33.
[8] CANE T, FERRYMAN J. Saliency-based detection for maritime object tracking[C]// GREGORY D, ANTON V D H, BASTIAN L, et al. 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Las Vegas: Institute of Electrical and Electronics Engineers, 2016 :1257-1264.
[9] TINCHEV G, PENATE-SANCHEZ A, FALLON M. SKD: keypoint detection for point clouds using saliency estimation[J]. IEEE Robotics and Automation Letters, 2021, 6(2): 3785-3792.
[10] DING X Y, ZHANG X. Visual tracking with salient features and selective mechanism[J]. Acta Electronica Sinica, 2020, 48(1): 118-123.
[11] WANG W G, SHEN J B, YU Y Z, et al. Stereoscopic thumbnail creation via efficient stereo saliency detection [J]. IEEE Tans on Visualization and Computer Graphics, 2016, 23(8):2014-2027.
[12] SHAFIEYAN F, KARIMI N, MIRMAHBOUB B, et al. Image retargeting using depth assisted saliency map [J]. Signal Processing(Image Communication), 2017, 50: 34-43.
[13] LIU Y X, ZHANG Y Z, COLEMAN S, et al. A new patch selection method based on parsing and saliency detection for person re-identification [J]. Neurocomputing, 2020, 374: 86-99.
[14] 武桐言, 霍梦真, 段海滨, 等. 仿鹰眼-中脑机制的无人机空中加油显著性检测[J]. 航空学报, 2023, 44(20): 628492.
[15] KIM T K, HONG J, RYU D, et al. Identifying and extracting bark key features of 42 tree species using convolutional neural networks and class activation mapping[J]. Scientific Reports, 2022, 12(1): 4772.
[16] SAPORTA A, GUI X, AGRAWAL A, et al. Benchmarking saliency methods for chest X-ray interpretation[J]. Nature Machine Intelligence, 2022, 4(10): 867-878.
[17] WINTERBOTTOM T, LEONE A, Al M N. A deep learning approach to fight illicit trafficking of antiquities using artefact instance classification [J]. Scientific Reports, 2022, 12(1): 13468.
[18] WU Z W, ALLIBERT G, MERIAUDEAU F, et al. HiDAnet: RGB-D salient object detection via hierarchical depth awareness[J]. IEEE Transactions on Image Processing, 2023, 32: 2160-2173.
[19] SUN F, REN P, YIN B, et al. CATNet: a cascaded and aggregated Transformer network for RGB-D salient object detection[J]. IEEE Transactions on Multimedia, 2024, 26: 2249-2262.
[20] ZHENG P, FU H, FAN D P, et al. GCoNet+ : a stronger group collaborative co-salient object detector [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(9): 10929-10946.
[21] GE Y, ZHANG Q, XIANG T Z, et al. GSNNet: group semantic-guided neighbor interaction network for co-salient object detection[J]. Computer Vision and Image Understanding, 2023, 227: 103611.
[22] ZHAO X, LIANG H, LI P, et al. Motion-aware memory network for fast video salient object detection [J]. IEEE Transactions on Image Processing, 2024, 68(5): 1203-1214.
[23] LIU Z, LIU J. Part-aware attention correctness for video salient object detection[J]. Engineering Applications of Artificial Intelligence, 2023, 119: 105733.
[24] ITTI L, KOCH C, NIEBUR E. A model of saliency-based visual attention for rapid scene analysis[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 1998, 20(11): 1254-1259.
[25] MA Y F, ZHANG H J. Contrast-based image attention analysis by using fuzzy growing[C]// LAWRENCE A R, HARRICK M V, THOMAS P, et al. MULTIMEDIA'03: Proceedings of the eleventh ACM international conference on Multimedia. Berkeley: Association for Computing Machinery, 2003: 374-381.
[26] ROSIN P L. A simple method for detecting salient regions [J]. Pattern Recognition, 2009, 42(11): 2363-2371.
[27] CHENG M M, WARRELL J, LIN W Y, et al. Efficient salient region detection with soft image abstraction[C]// PIETRO P, JITENDRA M, YASER S. 2013 IEEE International Conference on Computer Vision. Sydney: Institute of Electrical and Electronics Engineers, 2013: 1529-1536.
[28] JIANG B W, ZHANG L H, LU H C, et al. Saliency detection via absorbing markov chain[C] // PIETRO P, JITENDRA M, YASER S. 2013 IEEE International Conference on Computer Vision. Sydney: Institute of Electrical and Electronics Engineers, 2013: 1665-1672.
[29] SHEN X H, WU Y. A unified approach to salient object detection via low rank matrix recovery[C]//ANTONIO T, ANDREW Z, DAVID F. 2012 IEEE Conference on Computer Vision and Pattern Recognition. Providence:

Institute of Electrical and Electronics Engineers, 2012: 853-860.

[30] JIANG H Z, WANG J D, YUAN Z J, et al. Salient object detection: a discriminative regional feature integration approach[C]//BILL T, LARRY D, CARLO T. 2013 IEEE Conference on Computer Vision and Pattern Recognition.Portland: Institute of Electrical and Electronics Engineers, 2013: 2083-2090.

[31] WEI Y C, WEN F, ZHU W J, et al. Geodesic saliency using background priors [C]//ANDREW F, SVETLANA L, PIETRO P, et al. Computer Vision-ECCV 2012. Florence: Springer, 2012: 29-42.

[32] JIANG H Z, WANG J D, YUAN Z J, et al. Automatic salient object segmentation based on context and shape prior[C]//JESSE H, STEPHEN M, EMANUELE T. British Machine Vision Conference. Dundee: BMVA Press, 2011: 1-12.

[33] WANG L J, LU H C, RUAN X, et al. Deep networks for saliency detection via local estimation and global search[C]//RAMIN Z, MANJUNATH B S, LI F F, et al. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston: Institute of Electrical and Electronics Engineers, 2015: 3183-3192.

[34] ZHAO R, OUYANG W L, LI H S, et al. Saliency detection by multi-context deep learning[C]//RAMIN Z, MANJUNATH B S, LI F F, et al. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston: Institute of Electrical and Electronics Engineers, 2015: 1265-1274.

[35] LI G B, YU Y Z. Visual saliency based on multiscale deep features[C]//RAMIN Z, MANJUNATH B S, LI F F, et al.2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston: Institute of Electrical and Electronics Engineers, 2015: 5455-5463.

[36] HE S F, LAU R W H, LIU W X, et al. SuperCNN: a superpixelwise convolutional neural network for salient object detection[J]. International Journal of Computer Vision, 2015, 115(3) : 330-344.

[37] LIU N, HAN J W. DHSNet: deep hierarchical saliency network for salient object detection[C]//TOMASO P, YANN L. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: Institute of Electrical and Electronics Engineers, 2016: 678-686.

[38] FENG M Y, LU H C, DING E. Attentive feedback network for boundary-aware salient object detection [C]//TOMASO P, YANN L, DAVID M. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach: Institute of Electrical and Electronics Engineers, 2019: 1623-1632.

[39] ZENG Y, ZHANG P P, ZHANG J M, et al. Towards high-resolution salient object detection [C]// RAMANATHAN S, WANG X G, VLADIMIR I. 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul: Institute of Electrical and Electronics Engineers, 2019: 7233-7242.

[40] WU R M, FENG M Y, GUAN W L, et al. A mutual learning method for salient object detection with intertwined multisupervision[C]//TOMASO P, YANN L, DAVID M. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach: Institute of Electrical and Electronics Engineers, 2019: 8150-8159.

[41] WANG W G, SHEN J B, CHENG M M, et al. An iterative and cooperative top-down and bottom-up inference network for salient object detection [C]// TOMASO P, YANN L, DAVID M. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach: Institute of Electrical and Electronics Engineers, 2019: 5961-5970.

[42] SU J M, LI J, ZHANG Y, et al. Selectivity or invariance: boundary-aware salient object detection [C]//RAMANATHAN S, WANG X G, VLADIMIR I. 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul: Institute of Electrical and Electronics Engineers, 2019: 3798-3807.

[43] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [J]. Advances in Neural Information Processing Systems, 2017, 30: 5998-6008.

[44] YOU Q, JIN H, WANG Z, et al. Image captioning with semantic attention [C]//TOMASO P, YANN L. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: Institute of Electrical and Electronics Engineers, 2016: 4651-4659.

[45] YANG Z, HE X, GAO J, et al. Stacked attention networks for image question answering[C]//TOMASO P, YANN L. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: Institute of Electrical and Electronics Engineers, 2016: 21-29.

[46] CHEN L C, YANG Y, WANG J, et al. Attention to scale: Scale-aware semantic image segmentation[C]// TOMASO P, YANN L. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: Institute of Electrical and Electronics Engineers, 2016: 3640-3649.

[47] LI G, XIE Y, LIN L, et al. Instance-level salient object segmentation[C]//WANG X G, DENG J, SERGE B. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu: Institute of Electrical and Electronics Engineers, 2017: 247-256.

[48] KUEN J, WANG Z, WANG G. Recurrent attentional networks for saliency detection [C]//TOMASO P, YANN L. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: Institute of Electrical and Electronics Engineers, 2016: 3668-3677.

[49] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//VISHAL M P, DPDK V S, ZHANG H. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: Institute of Electrical and Electronics Engineers, 2018: 7132-7141.

[50] CHU X, YANG W, OUYANG W, et al. Multi-context attention for human pose estimation[C]//WANG X G, DENG J, SERGE B. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu: Institute of Electrical and Electronics Engineers, 2017, pp. 5669-5678.

[51] XU X, CHEN J, ZHANG H, et al. SA-DPNet: structure-aware dual pyramid network for salient object detection[J]. Pattern Recognition, 2022, 127: 108624.

[52] WANG J, ZHAO Z, YANG S, et al. Global contextual guided residual attention network for salient object detection [J]. Applied Intelligence, 2022, 52(6): 6208-6226.

[53] NGO T T, HUH E N, HONG C S. ETANet: an efficient triple-attention network for salient object detection[C]//SHIN Y, KIM S C, SOMNUK K, et al. 2023 International Conference on Information Networking (ICOIN). Bangkok: Institute of Electrical and Electronics Engineers, 2023:271-276.

[54] LI H, HAN Y, LI P, et al. Hybrid attention mechanism and forward feedback unit for RGB-D salient object detection[J]. IEEE Access,2023, 96068-96080.

[55] PANG Y, ZHAO X, ZHANG L, et al. CAVER: cross-modal view-mixed transformer for bi-modal salient object detection [J]. IEEE Transactions on Image Processing, 2023, 892-904.

[56] LI G, XIE Y, LIN L. Weakly supervised salient object detection using image labels [C]// SHEILA A M, KILIAN Q W. Proceedings of the AAAI Conference on Artificial Intelligence. Long Beach: Association for the Advancement of Artificial Intelligence, 2018: 6067-6076.

[57] ZENG Y, ZHUGE Y, LU H, et al. Multi-source weak supervision for saliency detection [C]// TOMASO P, YANN L, DAVID M. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach: Institute of Electrical and Electronics Engineers, 2019: 6067-6076.

[58] NGUYEN D T, DAX M, MUMMADI C K, et al. DeepUSPS: deep robust unsupervised saliency prediction with self-supervision [C]//TOMASO P, YANN L, DAVID M. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach: Institute of Electrical and Electronics Engineers, 2019: 6067-6076.

[59] ZHANG J, ZHANG T, DAI Y, et al. Deep unsupervised saliency detection: a multiple noisy labeling perspective [C]//ERIC M, WILLIAM B. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: Institute of Electrical and Electronics Engineers, 2018: 9029-9038.

[60] WANG L J, LU H C, WANG Y F, et al. Learning to detect salient objects with image-level supervision[C]// WANG X G, DENG J, SERGE B. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu: Institute of Electrical and Electronics Engineers, 2017: 3796-3805.

[61] ALPERT S, GALUN M, BRANDT A, et al. Image segmentation by probabilistic bottom-up aggregation and cue integration[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 34(2): 315-327.

[62] ACHANTA R, HEMAMI S, ESTRADA F, et al. Frequency-tuned salient region detection [C]//DAVID F, PHILIP T, ANDREW Z. 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami: Institute of Electrical and Electronics Engineers, 2009: 1597-1604.

[63] MOVAHEDI V, ELDER J H. Design and perceptual validation of performance measures for salient object segmentation [C]//HUA G, STEPHEN L. 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops. San Francisco: Institute of Electrical and Electronics Engineers, 2010: 49-56.

[64] CHENG M M, MITRA N J, HUANG X, et al. Global contrast based salient region detection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(3): 569-582.

[65] SHI J, YAN Q, XU L, et al. Hierarchical image

saliency detection on extended CSSD [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 38(4): 717-729.

[66] YANG C, ZHANG L, LU H, et al. Saliency detection via graph-based manifold ranking[C]//BILL T, LARRY D, CARLO T. 2013 IEEE Conference on Computer Vision and Pattern Recognition. Portland: Institute of Electrical and Electronics Engineers, 2013: 3166-3173.

[67] LI Y, HOU X, KOCH C, et al. The secrets of salient object segmentation[C]//ROBERTO C, SEBASTIANO B, GIOVANNI M F. 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus: Institute of Electrical and Electronics Engineers, 2014: 280-287.

[68] LI G, YU Y. Visual saliency based on multiscale deep features[C]//RAMIN Z, MANJUNATH B S, LI F F, et al. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston: Institute of Electrical and Electronics Engineers, 2015: 5455-5463.

[69] FAN D, CHENG M M, ZHANG J, et al. Salient objects in clutter [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 45(2): 2344-2366.

[70] MARGOLIN R, ZELNIK-MANOR L, TAL A. How to evaluate foreground maps [C]//ROBERTO C, SEBASTIANO B, GIOVANNI M F. 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus: Institute of Electrical and Electronics Engineers, 2014: 248-255.

[71] FAN D P, CHENG M M, LIU Y, et al. Structure-measure: a new way to evaluate foreground maps[C]// SHI J B, ABHINAV G, LORENZO T, et al. 2017 IEEE International Conference on Computer Vision (ICCV). Venice: Institute of Electrical and Electronics Engineers, 2017: 4558-4567.

[72] TORRALBA A, EFROS A A. Unbiased look at dataset bias[C]//DAVID F, PHILIP T, ANDREW Z. 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Colorado Springs: Institute of Electrical and Electronics Engineers, 2011: 1521-1528.

[73] CHENG M M, MITRA N J, HUANG X, et al. Global contrast based salient region detection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(3):569-582.

[74] LI Y, HOU X, KOCH C, et al. The secrets of salient object segmentation[C]//ROBERTO C, SEBASTIANO B, GIOVANNI M F. 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus: Institute of Electrical and Electronics Engineers, 2014: 280-287.

[75] ZHOU Z H. A brief introduction to weakly supervised learning[J]. National Science Review, 2018, 5(1): 44-53.

[76] LEE M, SHIN W, HAN S. TRACER: extreme attention guided salient object tracing network [C]// KRZYSZTOF J, JEROME L, PASCAL V H, et al. Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto: Institute of Electrical and Electronics Engineers, 2022, 36(11), 12993-12994.

(责任编辑:马田田)