ISSN 1004-9037,CODEN SCYCE4
Journal of Data Acquisition and Processing Vol. 38,No. 1,Jan. 2023,pp. 21－50
DOI:10. 16337/j. 1004-9037. 2023. 01. 002

http://sjcj. nuaa. edu. cn
E-mail:sjcj@nuaa. edu. cn
Tel/Fax:＋86-025-84892742

基于深度学习的显著性目标检测综述

孙　涵，刘译善，林昱涵

（南京航空航天大学计算机科学与技术学院/人工智能学院/软件学院，南京 211106）

摘　要：显著性目标检测通过模仿人的视觉感知系统，寻找最吸引视觉注意的目标，已被广泛应用于图像理解、语义分割、目标跟踪等计算机视觉任务中。随着深度学习技术的快速发展，显著性目标检测研究取得了巨大突破。本文总结了近5年相关工作，全面回顾了3类不同模态的显著性目标检测任务，包括基于RGB图像、基于RGB-D/T（Depth/Thermal）图像以及基于光场图像的显著性目标检测。首先分析了3类研究分支的任务特点，并概述了研究难点；然后就各分支的研究技术路线和优缺点进行阐述和分析，并简单介绍了3类研究分支常用的数据集和主流的评价指标。最后，对基于深度学习的显著性目标检测领域未来研究方向进行了探讨。

关键词：深度学习；RGB图像显著性目标检测；RGB-D/T图像显著性目标检测；光场图像显著性目标检测

中图分类号：TP391　　**文献标志码：**A

Deep Learning Based Salient Object Detection: A Survey

SUN Han, LIU Yishan, LIN Yuhan

(College of Computer Science and Technology/College of Artificial Intelligence/College of Software, Nanjing University of Aeronautics & Astronautics, Nanjing 211106, China)

Abstract: Salient object detection has been widely used in computer vision tasks such as image understanding, semantic segmentation, and object tracking by simulating the human visual system to find the most attractive targets for visual attention. With the rapid development of deep learning technology, salient object detection research has made great breakthroughs. This paper presents a comprehensive and systematic survey of salient object detection based on RGB images, RGB-D/T (Depth/Thermal) images, and light field images in the past five years. Firstly, the task characteristics and research difficulties of the three research branches are analyzed. Then the research technical route of each branch is expounded and the advantages and disadvantages are analyzed. At the same time, the mainstream datasets and common performance evaluation indexes of three kinds of research branches are introduced. Finally, possible future research trends are prospected.

Key words: deep learning; RGB salient object detection; RGB-D/T salient object detection; light field salient object detection

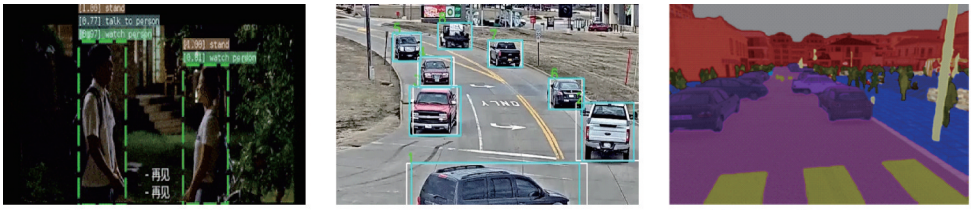
收稿日期：2022-06-01；**修订日期：**2022-10-02

引　言

人类视觉注意力检测研究起源于认知心理学和神经科学，包括人眼关注点检测和显著性目标检测。人眼关注点检测作为引入计算机视觉的早期人类视觉注意力机制研究工作，通过数据建模的方式模拟人类视觉注意系统的机能，对人眼在场景中某一个位置停留的可能性进行预测。随着计算机视觉领域的不断发展，强调对场景中显著目标整体的准确预测并且获取清晰的显著目标边界，由此产生了显著性目标检测分支，为目标级别的视觉任务提供更直接、更有效的信息，其研究历史相对较短，是一个纯计算机视觉任务。

随着信息技术的快速发展，手机、相机、笔记本电脑等多种智能终端设备的应用使得海量的图像、视频信息获取变得更加便捷，但信息总量在呈现指数级增长趋势的同时也产生了大量的冗余数据。因此，如何从海量信息中选择出有效部分显得尤为关键。显著性目标检测主要有以下两点优势：第一，给图像视频中更重要的部分分配有限的计算资源；第二，所得到的检测结果能够符合人的认知。

作为视觉注意力机制在目标分割任务上的延拓，并作为计算机视觉任务中非常重要的预处理步骤之一，显著性目标检测在立体匹配[1]、图像理解[2]、动作识别[3]、视频检测和分割[4]、语义分割[5]、医学图像分割[6]、目标跟踪[7]、行人重识别[8]、伪装目标检测[9]以及图像检索[10]等领域中发挥着非常重要的作用，如图1所示。由此可见，显著性目标检测有着广泛的应用价值和重要的研究意义。



(a) Action recognition　(b) Object recognition　(c) Semantic segmentation

图1　显著性目标检测应用领域

Fig.1　Salient object detection application field

显著性目标检测方法最早在1998年被Itti等[11]提出。此后一段时间内，基于手工提取图像特征分析的显著性目标检测虽然得到了一定的发展，但检测算法不仅耗时费力，而且面对复杂背景、光照变化等挑战性因素难以取得理想效果。2015年以来深度学习技术[12]极大地推动了显著性目标检测领域的相关研究工作，并提高了显著目标检测的性能，此后显著性目标检测的研究主要基于深度学习展开。为了在不同场景下获得更好的检测效果，目前基于深度学习的显著性目标检测研究方向大致可以分为：RGB图像显著性目标检测、RGB-D/T(Depth/Thermal)图像显著性目标检测、视频显著性目标检测、协同显著性目标检测以及光场图像显著性目标检测多个方向。

关于显著性目标检测虽然已有多篇优秀的综述论文发表[13-18]，本文则主要围绕基于深度学习的RGB图像、RGB-D/T图像、光场图像3个不同模态的单帧图像显著性目标检测分支展开研究，对当前工作进展以及存在的挑战进行详细分析和总结。

1　基于深度学习的显著性目标检测分类及难点分析

1.1　基于数据源的显著性目标检测分类

显著性目标检测包括两种机制，一是图像本身对人产生吸引的从下而上机制，二是在人意识控制下对图像主动关注的从上往下机制。由于目前的研究对于人的大脑结构与功能的了解还很肤浅，因此

主要围绕从下而上的机制来展开显著性目标检测的一系列研究,侧重于检测场景中吸引最多注意力的目标,然后逐像素地提取目标的轮廓。经过多年发展,显著性目标检测在图像分类、视频压缩及语义分割等领域中有着重要的应用价值。本文结合单帧图像数据源的不同,主要围绕以下3个显著性目标检测分支展开。

(1)RGB图像显著性目标检测,如图2(a)所示。RGB图像易获取的特性使得基于RGB图像的显著性目标检测研究是最多的,也是最早的。早期的研究方法主要利用图像的纹理、颜色、形状等底层特征获取显著信息,后来通过例如稀疏编码、卷积神经网络或者循环神经网络等特征学习方式检测显著目标。虽然这些方法在显著性目标检测任务上取得了较好的效果,但在面对复杂场景时仍具有局限性。

(2)RGB-D图像显著性目标检测,如图2(b)所示。研究人员发现深度信息的引入能够弥补RGB图像缺失的深度信息,有助于从杂乱背景、光照变化等挑战性情况中检测出显著性目标,自此显著性目标检测的方法产生了基于深度信息的显著性目标检测新分支,即RGB-D图像显著性目标检测。早期的RGB-D图像显著性目标检测模型倾向于手工提取特征,然后融合RGB图像和深度图像。随着Microsoft Kinect等深度传感器的广泛使用,深度图像更易获取,进一步推动了RGB-D图像显著性目标检测领域的研究。该方向的检测任务能够在前景和背景相似的复杂场景下,利用RGB图像所包含的底层特征以及深度图像所包含的空间信息区分显著目标和背景,进而提高检测结果图的质量。RGB-T图像显著性目标检测,如图2(c)所示。在纹理相似、背景暗光及复杂场景下,RGB图像不能为模型训练提供更多更具有区分度的信息,且常常会导致预测结果不准确,或者没有办法识别出目标。当前有很多研究将深度信息引入显著性目标检测任务,但深度信息在目标和镜头垂直或是同一个目标在深度图上的差别很大时便会失去作用。近年来随着热红外技术的普及,研究人员发现热红外信息对于照明条件差、照明不均匀产生的目标模糊问题非常有效,且对天气条件不敏感,适合处理在全黑环境、大雾天气、杂乱背景等恶劣条件下拍摄的场景,例如在城市街景的语义分割任务中就取得了很好的效果。因此,研究人员将热像仪生成的红外图像作为重要的信息补充,进而提高显著性目标检测效果。

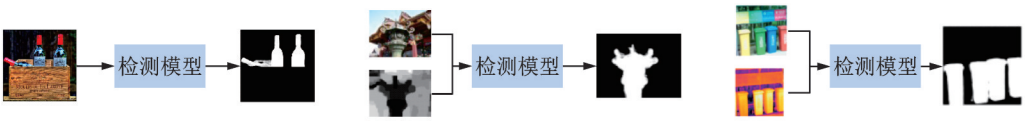


图2　RGB、RGB-D、RGB-T图像显著性目标检测方法流程示意图

Fig.2　Schematic diagram of processes of RGB, RGB-D, RGB-T image salient object detection methods

(3)光场图像显著性目标检测,如图3所示。使用光场图像进行显著性目标检测的思想最早在文献[19]中提出。使用专门设计的相机(例如Lytro)拍摄的光场图像,本质上是一个由观察场景的相机网格拍摄的图像阵列。光场图像数据为显著性目标检测提供了2个有效支持:一是允许合成一叠聚焦在不同深度的焦点堆栈图像;二是提供了丰富的多模态信息,包括光线的位置、方向和几何等空间信息和结构信息。此外,由于RGB-D图像显著性目标检测模型所需深度信息可以使用一定的技术手段从光场数据中获取,因此RGB-D图像显著性目标检测可以被视为退化的光场图像显著性目标检测。

1.2　基于深度学习的显著性目标检测难点分析

在计算机视觉中,显著性目标检测通常包含以下2个阶段:一是检测最为显著的目标,二是精确分割显著目标所在的区域。虽然当前显著性目标检测领域的研究取得了较大进展,但针对以上2个阶段,

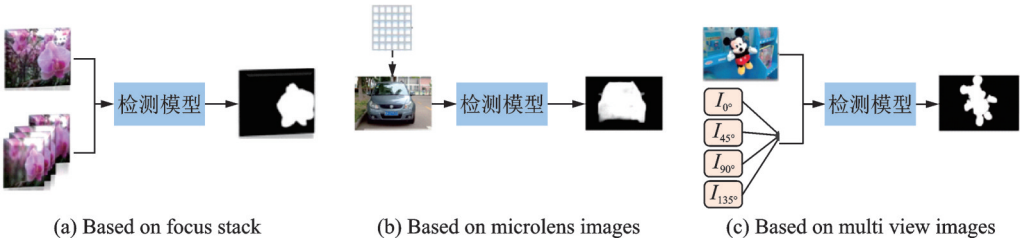


图3 光场显著性目标检测不同方法流程图

Fig.3 Flowchart of different methods for light field salient object detection

面对复杂背景、光照变化等挑战性因素，显著性目标检测研究仍有探索空间，目前的主流方法仍然存在显著目标边界模糊、数据集图像数量不够或深度等图像质量不高以及多目标显著性检测不完整等问题。以下将针对几个分支分别说明当前显著性目标检测所面临的难点。

1.2.1 RGB图像显著性目标检测难点

（1）目标尺度变化大

无论是自然场景图像还是遥感图像中的显著性目标检测，显著性目标的尺度变化都极大地影响着模型性能。由于基于深度学习的显著性目标检测任务模型大多是建立在主流的图像分类骨干网络上，缺乏对像素级别精确预测的基础。对于不同大小的目标来说，大尺度目标需要在分辨率更小的深层特征图中才能被捕捉到，而小尺度目标的特征则可能会由于像素的减少在分辨率降低的过程中被丢失；相反的，小尺度目标虽然可以在分辨率较大的浅层特征图中保留相关的有效信息，却又缺乏足够的语义信息来指导网络准确定位目标的位置。因此，如何设计合适的网络结构来应对多尺度目标的检测需求是长期以来RGB图像显著性目标检测任务的突破点之一。

（2）相似背景和复杂背景对检测的干扰

当显著目标的纹理及颜色特征明显区别于其背景时，很容易获得精确的检测效果。然而，现实场景中往往存在着大量前景和背景相似的情况，这将导致模型无法较好地辨别显著性目标的准确位置，从而导致大量的假阳性和假阴性预测，阻碍了显著性目标检测任务模型的性能提升。而复杂背景情况下的显著性目标检测则对模型的性能提出了更大的考验。常见的复杂背景如阴影、倒影等信息常常会误导模型将其误检为前景目标的一部分，从而降低检测质量。现有的大多数方法采用注意力机制来增强模型的鉴别能力，达到过滤背景信息和高亮前景目标的目的，但仍有很大的空间值得不断改进。

（3）获得的显著图边界模糊

为了实现高质量的显著性目标检测，精确的边缘检测是关键基础，在显著性目标检测的其他分支中均面临这一挑战。当前的研究大多集中在区域精度上而不在边界质量上，导致显著目标检测结果的边界问题不佳，如图4所示[20]，其中：第1列为RGB图，第2列为真值图，第3～5列为对应模型检测图；第1、3行为显著图，第2、4行红色阴影区域为所检测的显著区域，便于对比查看模型的检测效果。以前的显著性目标检测方法通过一个步骤同时捕捉图像的语义信息和边界细节，但这两个问题

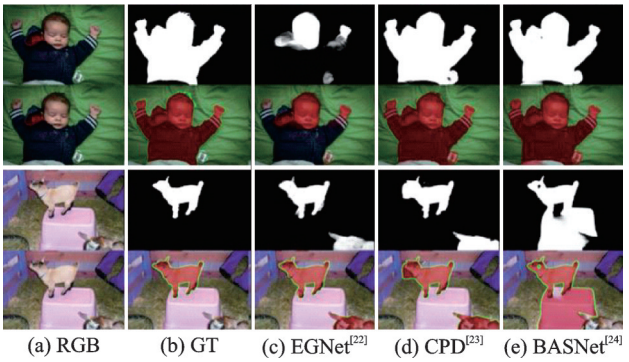


图4 不同显著性目标检测模型边缘检测情况对比图

Fig.4 Edge detection comparison of different salient object detection models

本质不同,导致处理高清图像时结果较差。语义信息的捕捉需要较大感受野,边界信息的捕捉需要低级结构信息,即较小感受野。如果直接应用低分辨率模型的话,此时大感受野会带来更大的计算开销,因此往往需要先将图像进行下采样,而这样就会导致低级结构信息的丢失。早期的检测方法是简单使用全连接层,但是却破坏了数据的空间结构信息。后来使用全连接网络(Fully connected network,FCN)[21]缓解了这一问题,但是细节信息容易丢失。而在显著性目标检测方法中常用的损失函数交叉熵损失在判别边界像素点时,通常置信度都比较低,导致边界模糊。除此之外,现有的模型缺乏强大边界检测能力的另一个原因在于传统的显著性标签掩码作为训练标签时平等和独立地处理显著目标内的所有像素,因此它们缺乏像素间关系的信息,使边缘难以检测。基于光场图像的检测模型同样很少考虑边缘像素的检测质量,虽然光场数据能够衍生出多种可处理的图像表征形式,但是却缺少专门描述目标边界的方法,因此基于边界感知的检测模型具有很大的挑战性和探索空间。

1.2.2　RGB-D/T图像显著性目标检测难点

近年来,互补信息,即深度或热红外信息的使用已经显示出其在显著性目标检测任务上的优势,深度信息能够从低对比度、复杂背景等场景中准确地检测出显著目标,热红外信息能够获取透明度较高、深色等情况下显著目标的深度信息。两个分支当前所面临的挑战有相近之处。

(1) 在真实和复杂场景下的检测效果不佳

对于RGB-D图像显著性目标检测来说,现有数据集存在设计偏差,多为环境相对简单且至少包含一个显著目标的图像,或是在相对简单环境中多个目标通常只有人的图像,这些图像只能帮助分析模型的整体性能,缺乏反映真实场景中所面临挑战的各种属性,使用其训练出来的模型在真实场景中的应用效果难以令人满意。除此之外,部分数据集还存在图像数量较少、显著目标标注质量较低的情况。例如数据集MSRA-A[25]和MSRA-B[25]中的显著目标基本是以标定框的形式进行标注,数据集ASD[26]和MSRA10K[27]在每帧图像中大多只包含一个显著目标,数据集SED2[28]在单帧图像中包含2个显著目标但仅有100幅图像。除此之外,神经网络虽然在RGB-D显著性目标检测上的应用取得了非常显著的成果,但较大的计算成本以及相对复杂的模型结构导致相关方法难以应用到移动设备上,例如自动驾驶、机器人、手机等,如若使用类似于MobileNet[29]和ShuffleNets[30-31]等轻量网络,特征表示能力远不如深层网络,导致检测精确度受到限制,模型在移动设备上的使用效果不尽如人意。

同样的,对于RGB-T图像显著性目标检测来说,目前仅有VT821[32]、VT1000[33]、VT5000[34]三个数据集,其中首个数据集VT821于2017年被提出,因此真实场景下数据集的构建对于RGB-T显著性目标检测来说仍然是一个很大的挑战,场景的多复杂性、高多样性、高分辨率等以及数据集的规模都是下一步的研究方向。

(2) 跨模态信息不能有效融合

对于RGB-D模型来说,深度信息和RGB信息的有效融合至关重要。目前已有的方法通常将RGB信息和深度信息视为独立的信息,分别为其特征提取设计单独的网络,不能有效地进行特征提取和融合[35],难以捕捉两种模态的相互作用,且庞大的网络结构需要大量的参数和训练数据,当前高质量的深度图仍然是稀缺的。除此之外,现有方法常常没有带监督的解码器来指导学习,这可能会导致无法获得最佳深度特征,且很少有RGB-D的显著性目标检测模型明确利用了模态的特异性[36]。

与RGB-D图像显著性目标检测相同,RGB-T显著性目标检测问题目前只是独立解决,大部分都是直接从主干中提取并融合原始特征,这类方法很容易受到低质量模态数据和冗余跨模态特征的限制。当前的方法通过简单的连接或逐元素求和操作来融合多模态特征,而没有考虑来自不同模态特征的重要性,未能很好地探索RGB图像和深度图像或热红外图像之间的互补信息/特征。这种操作允许包含冗余或非显著的特征,从而使得融合过程不能够很好地互补。

对显著性目标检测来说，重要的是抓住这两种不同模态数据之间的相互关系，并利用它们之间的互补性，然而当前有效的融合方法仍未得到充分探索。在没有充分考虑跨模态数据不一致性和多尺度检测鲁棒性的情况下，直接融合跨模态特征可能会产生次优结果[37]。除此之外，大多数基于多模态的检测模型几乎没有考虑多尺度深度特征，尽管多尺度深度特征已经被证明对于传统的显著目标检测任务是有效的。

（3）低质量互补信息对显著性目标检测结果有较大影响

对于RGB-D/T图像显著性目标检测方法来说，深度信息和热红外信息等互补信息的引入虽然能够弥补RGB图像缺失的空间信息，但常用数据集以及真实场景中互补信息的质量是不稳定的，而低质量互补信息的引入会影响最终显著性目标检测的效果。截至目前已经有不少研究人员通过各类方法进行尝试，但仍难以做好平衡，以RGB-D图像显著性目标检测方法为例，以下3种情况均会影响最终显著图的质量：一是尝试改善深度图前景和背景之间的对比度来修复深度图质量时，如果不能有效增强前景，则会产生显著目标不完全检测的结果，如图5第1行所示[38]；二是当深度图被识别为低质量时选择直接丢弃，深度信息的引入则失去互补作用，最终的检测结果会受到前景和背景对比度较低的RGB图像影响，如图5第2行所示；三是采用知识提取技术使RGB数据流能够学习深度信息，虽然避免了测试阶段低质量深度图的影响，但当提取的深度信息与测试中的高质量深度图不一致时，深度信息的引入也会失去互补作用，深度模型的性能将会受到影响，如图5第3行所示。由此可见，平衡、利用好深度信息和热红外信息在跨模态特征融合阶段的作用是RGB-D/T图像显著性目标检测方法不可忽视的挑战。

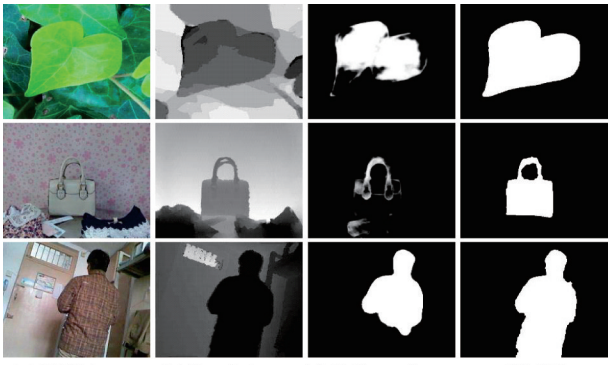


图5 低质量互补信息影响RGB-D图像显著性目标检测结果

Fig.5 Effect of low-quality complementary information on RGB-D salient object detection results

1.2.3 光场图像显著性目标检测难点

光场图像显著性目标检测任务的不同之处主要在于研究如何利用光场数据如深度信息、聚焦线索以及角度变化来实现显著性目标检测。原始的4D光场数据难以被直接应用处理，人们往往会对光场数据进行预处理，从而生成微透镜图像、子孔径图像、多视角图像、焦点堆栈图像和全聚焦图等数据。由于研究时间还不够长，基于深度学习的光场图像显著性目标检测仍颇具挑战性，下面对其主要难点进行分析。

（1）光场数据视角基线狭窄，影响最终显著图质量

如今光场数据的获取主要通过Lytro光场相机采集，有视角基线狭窄、视差范围小的缺点，而狭窄的视角基线不仅会导致视角信息的冗余，还会影响深度图的质量，进而让显著性目标检测变得困难。除此之外，视角基线狭窄情况下像素在单张光场图像内无显著变化，导致光场图像中相对深度信息的获取受限。因此，在这一情况下保障光场数据恢复深度信息的有效性和准确性是检测模型在设计过程中的难点之一。与此同时，提高数据质量，针对视角基线狭窄、视差范围小等问题进行改进也是下一步研究方向。

（2）受通用基准等条件限制，当前经验成果较少

尽管光场数据内丰富的线索和信息有助于算法更好地识别目标并提高显著性目标检测任务性能，

但人们对许多数据形式,如多视角图像、微透镜图像、高分辨率图像等研究还很少。大多数现有方法都集中在对焦点堆栈图像的利用上并取得了理想的效果,这是一个很好的趋势,说明光场图像显著性目标检测任务可以被很好地解决。然而,一个未解决的问题是,既然光场数据的不同数据形式可以提供对场景的不同描述,如何充分挖掘其他受关注较少的数据特征来建立更多的光场图像显著性目标检测任务是值得进一步研究的方向,包括对于光场数据中隐含深度信息的利用,对于弱监督、无监督学习方式在该分支的应用也是未来需要进一步突破的难点。

2　基于深度学习的显著性目标检测研究思路

在已有的研究中,往往认为一个显著性目标检测模型能取得较好的效果至少应该满足以下3个标准:一是好的检测能力,尽量少地遗漏真正的显著区域或错误地将背景标记为显著区域;二是高分辨率,显著图应具有较高的分辨率或全分辨率,以准确定位显著目标并保留原始图像信息;三是高计算效率,作为其他任务的前置阶段,能够快速检测到显著区域。基于不同数据源的显著性目标检测方法研究思路也主要围绕以上3个方面展开,对检测模型的性能进行不断优化和提升,如图6所示。

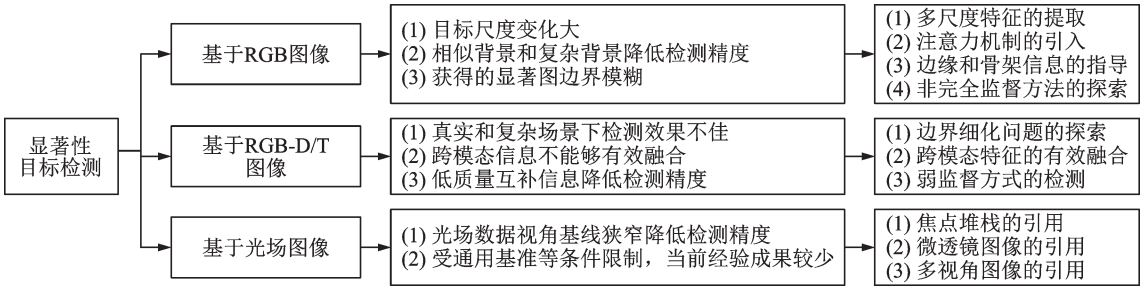


图6　基于深度学习的显著性目标检测难点及研究思路简图

Fig.6　Difficulties and research ideas of salient object detection based on deep learning

2.1　RGB图像显著性目标检测研究思路

2.1.1　多尺度特征的提取

计算机视觉中的语义分割旨在从像素的角度分割出图像中的不同目标,而经典的语义分割方法大都是基于FCN[21]实现。多尺度特征的学习在FCN模型中便已崭露头角,网络中生成的多组分辨率的特征图代表了不同特征尺度的信息。而后在U-Net模型[39]中,多尺度特征图在上采样解码阶段中被充分融合,成为如今许多深度学习网络的基础架构形式,如图7所示,其中:Conv表示卷积,ReLU表示激活函数,Copy and crop表示拼接,Max pool表示最大池化。还有一种多尺度特征学习方式则是在同一模块内提取本层的多尺度信息,这也是许多文献里最常用的解决思路,例如Chen等在DeepLab模型[40]中提出的ASPP模块和Zhao等在PSPNet模型[41]中提出的PPM模块,结构示意如图8所示,其中:Conv $i\times i$表示核大小为i的卷积层,Rate=i表示卷积核的空洞率大小,Concat表示特征图在通道上的级联操作,Pool($i\times i$)表示核大小为i的池化层,Upsample表示特征图的上采样操作。图8(a)使当前层特征图通过并联的形式采用不同空洞率的空洞卷积层来捕获多尺度信息,图8(b)采用不同倍率的池化操作将特征图转换到不同的尺寸后再融合,从而学习多尺度上下文信息。抛开显著性目标检测特有的任务特点来说,由于都是对像素进行类别预测,显著性目标检测任务实则可以归属到语义分割的讨论中。正因如此,大多数语义分割领域的网络架构和多尺度特征学习思路都被继承到了显著性目标检测模型中,并取得了不错的性能。

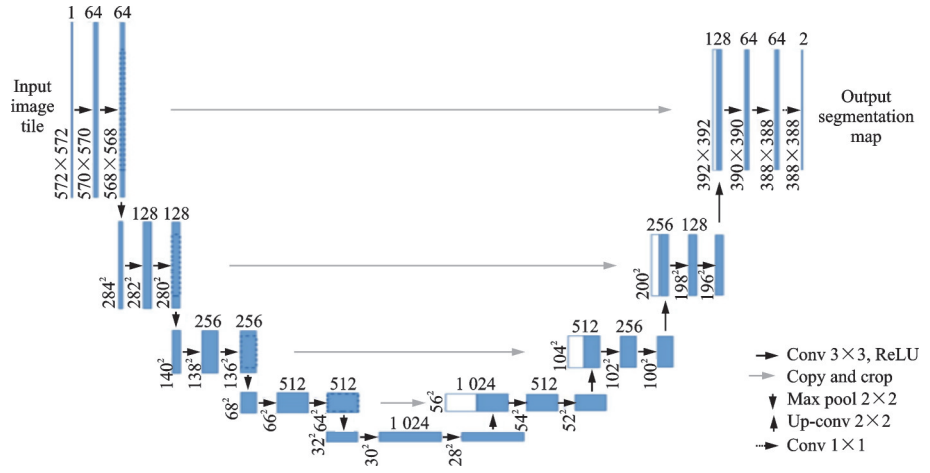


图7 U-Net网络结构示意图

Fig.7 Schematic diagram of U-Net network structure

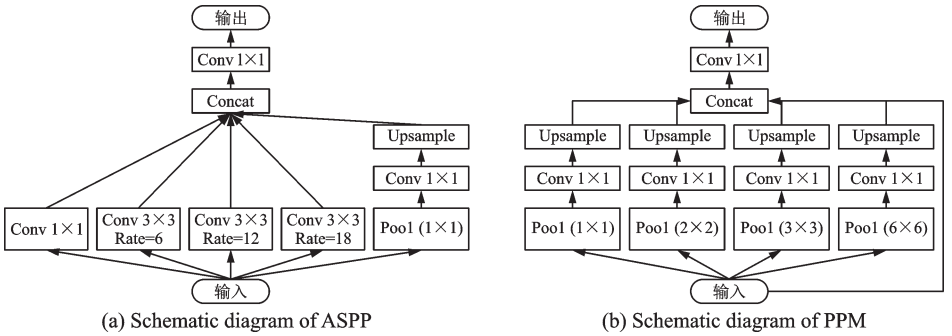


图8 不同多尺度特征学习模块

Fig.8 Different multi-scale feature learning modules

显著性检测的应用场景往往是在现实物理世界中,从相关数据集的图像特征中可以发现显著目标具有非常大的尺度变化,故而针对多尺度特征的提取和学习成为了显著性目标检测模型中被探讨最为频繁的内容。以文献[33]方法为代表的研究从层间多尺度特征学习着手,使用聚合交互模块和自我交互模块来解决多尺度特征学习的问题,前者可以通过相互学习的方式有效地利用相邻层的特征,而后者则使网络可以自适应地从本层特征图中提取多尺度信息,并更好地应对尺度变化;而与文献[42]类似的方案则关注于挖掘层内的多尺度信息,利用U型网络模块来学习多尺度信息,将U型结构嵌套到网络的基本组件中,每个U型网络模块通过多次卷积和池化操作生成了反映不同尺度信息的特征图。探索更丰富的特征尺度信息需要更多的卷积操作,从而带来计算量猛增的问题。因此,一些轻量级多尺度特征提取方案受到广泛关注。Shen等[43]提出了一种基于全压缩多尺度模块的轻量级模型,该模型整合3个由普通卷积、空洞卷积和池化层组成的多尺度分支生成的多尺度特征,取得了可匹敌大规模网络性能的成绩;而Liu等[34]不仅沿用了空洞卷积的策略,还引入了注意力机制的思想,在该方法中提出了一种立体注意力多尺度模块,它采用空洞深度扩张卷积在保持低计算复杂度的同时获得不同尺度的感受野,并基于由空间注意力和通道注意力组成的立体注意力自适应地融合多尺度特征。

而后的工作不仅着眼于多尺度特征的提取,而且对多尺度特征的融合问题也进行了研究。事实上,绝大多数显著性目标检测网络都是基于特征金字塔网络搭建的,输入图像经过一系列的卷积特征

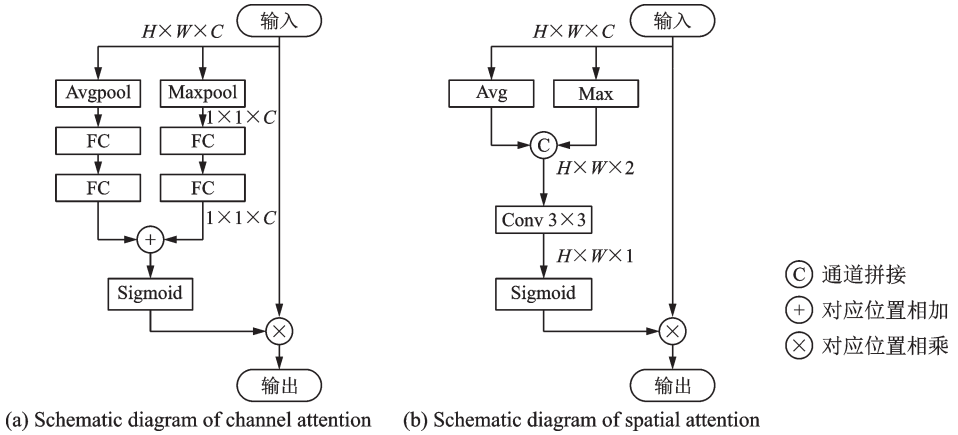
提取和空间池化之后,将会得到若干组具有不同分辨率的特征图。这些特征图反映了其在不同尺度上的特征表征,高效的多尺度特征融合方式有助于结合不同尺度的特征信息,以达到“查缺补漏、自我优化”的效果。

特征融合的研究重点是如何让不同的尺度特征叠加后能更有效地反映显著性目标的信息。Wei等[44]设计的F^{3}Net模型包含一种交叉特征融合机制,使用像素乘法提取深层和浅层特征的公共部分,再将公共特征分别与原始特征相结合生成融合后的浅层和深层特征;Chen等[45]则提出了一种新的聚合特征的方法,和F^{3}Net模型一样,它也使用乘法运算充分结合低层特征、高层特征和全局上下文特征,不同的是,它没有将乘法运算后的结果和原始特征结合,而是将3种特征两两相乘后的输出进行级联融合,从而达到增强显著目标响应的同时抑制背景噪声的目的;Liu等[46]采用另一种以最深层特征引导浅层特征融合的思想,它并不是对每个相邻尺度特征图进行融合优化,而是利用PPM模块在最深层特征图中提取全局语义信息,而后将该全局信息引导至每个自顶向下的特征融合阶段,从而避免高层语义信息被稀释,保证显著性目标的位置信息不被背景噪声所吞噬。上述方法均采用了对称的U型网络编解码器,即网络的输出集成了主干网络中的所有特征图;而Wu等[23]认为浅层的特征图对性能的贡献较小可以直接舍弃,在模型中设计了仅集成较深层特征的部分解码器,该解码器首先利用3个高层特征生成初始显著图,再将其反馈到优化层继续进行第2阶段的特征聚合。同时,由于解码器丢弃了浅层的较大分辨率特征,从而实现了推理加速。

2.1.2　注意力机制的引入

卷积神经网络总是无差别地处理每一个特征像素或是特征通道,但这不利于网络学习和任务相关的特征信息。不少研究者使用注意力机制推理出一些权重信息,引导深度网络将其学习重心转移到显著的特征区域或是富有语义信息的通道上,从而推进模型挖掘有效的显著性目标特征。

注意力机制在显著性目标检测上的应用大体上能够分为通道注意力和空间注意力两大类。常见的通道注意力和空间注意力的结构图如图9所示,其中:FC表示全连接层;Avgpool和Maxpool分别表示空间上的平均池化和最大池化,即输出为一维向量,每一维特征等于对应特征图的全局平均值或全局最大值;Avg和Max分别表示通道上的取平均值和取最大值,即输出为单通道特征图,每个位置上的特征等于原始输入中对应位置在所有通道上的平均值或最大值。对这两类注意力机制的使用,Zhao等[47]提供了一种非常经典的思路,根据特征图自有的特质,在浅层特征图上使用空间注意力以过滤背景信息并高亮目标细节,而在高级特征图上使用通道注意力以选择合适的尺度和感受野。



(a) Schematic diagram of channel attention　(b) Schematic diagram of spatial attention

图9　不同注意力机制

Fig.9　Different attention mechanisms

此外,很多方法更青睐于设计融合多种维度的高效注意力模块。Li等[48]将上述经典的空间和通道注意力进行了集成,设计了一种新的注意力模块引导网络同时从通道语义和空间细节上对显著区域进行高亮;而Lin等[49]尝试利用更多维度的注意力方式并取得了理想的效果,提出了一种新颖的多内容互补模块,该模块利用注意力机制的思想探索前景特征、边缘特征、背景特征和全局图像级特征之间的互补性,以此增强遥感图像中不同尺度的显著性区域;Cong等[50]则将对注意力机制的理解放在了多尺度层面,通过不同大小的卷积核生成代表不同感受野尺度的注意力图,之后将这些注意力信息施加在网络浅层,从而使得浅层特征能够更加突出目标的细节信息并解决显著性目标的尺度变化问题。

研究发现,注意力机制除了可以用于指导特征图自身的优化,也能够对网络中的其他层特征图产生有益的指导。Zhang等[51]利用特征提取及下采样的过程中,浅层特征能够帮助深层特征在失去分辨率的同时保留有效细节信息的特点,设计了一种密集的注意力流网络,使浅层注意力流向深层从而指导深层特征图的生成;Chen等[52]则与之恰恰相反,以自上而下的方式指导侧输出的残差学习,并对每层的侧输出使用反向注意力,借助侧输出来达到当前预测结果覆盖的目的,网络最终可以探索丢失的目标部件和细节。注意力信息不仅可以提供给其他层的特征图,结合相邻层的特征服务本层的注意力生成也是一个很好的启发,Tang等[53]就同时将高层语义特征和当前层特征图作为输入,分别考虑了二者的通道信息和位置信息,所提出的跨层注意力模块结合非局部模型捕获长范围依赖的特点,实现了非常有效的性能提升。

2.1.3 边缘和骨架信息的指导

RGB图像显著性检测模型基本上都是基于像素级标注的二值化标签进行损失优化指导的,因此大多数模型的特征图能够普遍反映出目标的大致位置,但是这对于复杂结构目标的检测定位和细节恢复就显得力不从心。随后,一系列相关工作的研究成果证明了边缘信息和骨架信息能够有效地帮助网络提升在目标边缘和结构完整性方面的检测效果。

在基于边缘信息的研究工作中,Zhao等[22]提出了经典的联合边缘信息指导显著性检测的方法,其网络结构图如图10所示,它设计了一种边缘引导策略,利用局部边缘信息和全局位置信息生成了高质量的显著边缘信息,而后又将该边缘特征与不同尺度的显著目标特征图相结合,从而获得显著目标的精细边界。图10中棕色粗线表示尺度之间的信息流,NLSEM为非局部显著边缘特征提取模块,PS-FEM为渐进显著目标特征提取模块,O2OGM为一对一引导模块,FF为特征融合模块。此后Ke等[54]延续前人的思想,在网络的推理过程中通过递归的卷积运算不断结合轮廓信息和显著性信息,在多阶段的损失监督和信息交互基础上,增强了两种特征的检测质量;而Yang等[55]则另辟蹊径,在边缘特征的生成和监督方面提出了一种全新的策略,使用连接性掩码和显著性掩码作为标签以有效建模像素间关系和目标显著性,同时提出了一个双边投票模块来增强输出的连接图和一种有效利用边缘特定特征的边缘特征增强方法。

骨架信息同样有助于提升网络对显著性目标的检测性能。Wu等[56]提出的DCN模型将整个网络分为2个阶段,其中分解网络迭代地利用跨任务聚合和跨层聚合模块同时进行显著性、边缘和骨架图的预测,而在合成网络中,使用边缘和骨架信息学习分别定位显著目标的边界和内部,进而用于填补缺陷和抑制噪声,最终设计的模型能够达到完整分割显著目标的效果;类似地,Liu等[57]也将骨架、边缘和显著性三者融合在同一个网络中,不同于DCN,它实现了一个多任务网络框架,其设计的选择性融合模块允许每个任务根据其自身的特征从共享的主干网络不同层级中动态选择有用的相关特征,由于不同任务均源于主干网络的特征提取结果,进而驱使网络在特征学习的同时考虑显著性、边缘和骨架信息的挖掘。

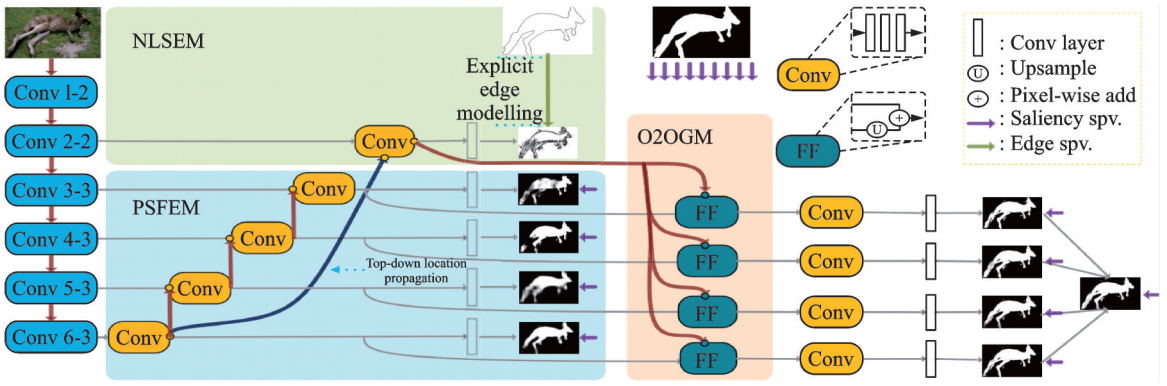


图 10　EGNet模型整体架构

Fig.10　Schematic diagram of EGNet

2.1.4　非完全监督方法的探索

尽管基于完全监督的显著性目标检测模型是最主要的技术路线,但是由于其依赖像素级精细的注释标签,导致数据集的构建耗费大量的时间和人力,缺陷明显。如今为了打破这一阻碍,提出了许多非完全监督的显著性目标检测模型,大致上分为两类:弱监督方法和无监督方法。

基于弱监督技术的方法研究重点主要是在伪标签的生成和应用上。例如,Gao等[58]提出NSS模型,采用点注释的方法来解决显著性目标检测任务,首先为了推断显著图标签,利用自适应的漫水填充算法来生成伪标签进行第一轮的训练,而后提出了一种非显著性抑制策略来优化第一轮的显著图并利用其进行第二轮的训练;而Piao等[59]则基于类激活图生成多个伪标签,通过指导滤波器来有效提取伪标签中准确显著性信息,整合后的显著性信息在多路指导损失函数的监督下传递给最终的显著性目标检测网络。

对于无监督方法,Zhang等[60]提出LSPENM模型使用显著性预测模块,基于当前噪声估计和噪声显著图学习潜在显著图,而后采取传统无监督方法的噪声建模模块则基于更新的显著性预测和噪声显著图更新不同显著图中的噪声估计,最终的损失函数由显著性预测模块损失函数和噪声模块损失函数构成;Yan等[61]利用“假标签”进行监督训练,从合成但干净的标签中学习显著性,进而节省人工标注,同时为了缓解合成场景和真实场景之间的巨大外观差异,提出了一种新的无监督域自适应显著性目标检测方法,通过不确定性感知自我训练在两个域之间进行调整。

2.1.5　小结

RGB图像显著性目标检测是该领域最经典的任务,从基于传统图像处理方法的Itti等模型[11]到基于深度学习的MINet[33]、EGNet[22]等模型,针对RGB图像的显著性目标检测任务形成了上述4种技术路线。多尺度特征学习技术赋予神经网络检测多尺度目标的能力,以此处理真实世界中尺度变化多样的显著性目标;注意力机制则能够迫使网络建模语义信息、细节信息等特征相关性,从而保证目标分割的完整性和准确性;多信息辅助路线引入除显著性真值图外其他有效的边缘、骨架信息,进一步引导了更加精细的检测效果;而非完全监督方法打破了费时费力的数据标注的瓶颈限制,现有的多种方法已将非完全监督模型的性能推向了具有竞争力的水平。文中所介绍的4种研究思路是近年来最常用的模型改进和性能提升技术,正如表1中所述方法在其训练集对应的测试集上平均绝对误差(Mean absolute error,MAE)的表现所示,这些方法已在主流的公开数据集上取得了较为理想的精度。未来建议在以上研究成果的经验指导下,更加关注模型的轻量化和实时性设计,从而促进算法模型在现实场景下的快速转化。

表1 RGB图像显著性目标检测的不同研究思路方法对比

Table 1 Comparison of different RGB salient object detection methods

类别	名称	训练集	骨干网络	MAE	年份	出处
多尺度特征	PoolNet[46]	DUTS-TR	ResNet-50	0.036	2019	CVPR
	CPD[23]	DUTS-TR	ResNet-50	0.043	2019	CVPR
	MINet[33]	DUTS-TR	ResNet-50	0.037	2020	CVPR
	GCPANet[45]	DUTS-TR	ResNet-50	0.038	2020	AAAI
	F^3Net[44]	DUTS-TR	ResNet-50	0.035	2020	AAAI
	MEUNet[42]	DUTS-TR	ResNet-50	0.031	2021	PRICAI
	SAMNet[34]	DUTS-TR	SAM	0.058	2021	TIP
	FSMINet[43]	EORSSD	FSM	0.008	2022	GRSL
注意力机制	RANet[50]	DUTS-TR	VGG-16	0.036	2018	ECCV
	PFAN[47]	DUTS-TR	VGG-16	0.041	2019	CVPR
	DAFNet[51]	EORSSD	Res2Net-50	0.005	2020	TIP
	CLASS[53]	DUTS-TR	ResNet-50	0.034	2020	ACCV
	PurNet[48]	DUTS-TR	ResNet-50	0.039	2021	TIP
	RRNet[50]	EORSSD	ResNet-50	0.008	2021	TGRS
边缘和骨架信息的指导	EGNet[22]	DUTS-TR	ResNet-50	0.039	2019	ICCV
	DFI[57]	DUTS-TR	ResNet-50	0.038	2020	TIP
	DCN[56]	DUTS-TR	ResNet-50	0.035	2021	TIP
	RCSBNet[54]	DUTS-TR	ResNet-50	0.034	2022	WACV
	BiconNet[55]	DUTS-TR	MINet	0.035	2022	PR
非完全监督	LSPENM[60]	DUTS-TR	DeepLab	0.086	2018	CVPR
	MFNet[59]	DUTS-TR	DenseNet	0.076	2021	ICCV
	UPL[61]	DUTS-TR	LDF	0.050	2022	AAAI
	NSS[58]	DUTS-TR	ViT	0.045	2022	AAAI

2.2 RGB-D/T图像显著性目标检测的研究思路

2.2.1 边界细化问题的探索

良好的显著性目标检测都需要突出并完全突出显著区域，主要关注外部边界的清晰度和内部类的一致性。为达到理想的检测效果，主要从以下两个方面展开相关工作。

(1)在网络结构上进行优化策略的设计。Jin等[38]充分利用高级特征粗略定位显著区域、低级特征进一步细化显著区域边界的特性，所提出的模型为2个阶段跨模态特征融合网络结构，通过RGB图生成估计深度图，自适应地融合原始深度图和估计深度图，以此实现完整性显著性目标检测，提高了对不稳定深度图的鲁棒性；而Liu等[62]考虑到目前显著性目标检测方法所采用的卷积神经网络(Convolutional neural networks, CNN)框架受到滑动窗口提取特征的局限性，在学习长程依赖关系方面受到限制，第一次从序列到序列的角度重新思考显著性目标检测，将图像块作为输入，基于纯Transformer在图像块之间传递全局信息，提出了一种新颖的可用于RGB和RGB-D的检测模型，即VST(Visual saliency transformer)模型，无需使用卷积操作和双线性上采样，取得了较好的效果，如图11所示，其中：Encoder为编码器，Conventor为转换器，Decoder为解码器，Cross modality transformer为跨模态转换器模块，Patch-task attention为注意力机制，T2T为预训练阶段采用的编码器，RT2T为反向T2T。

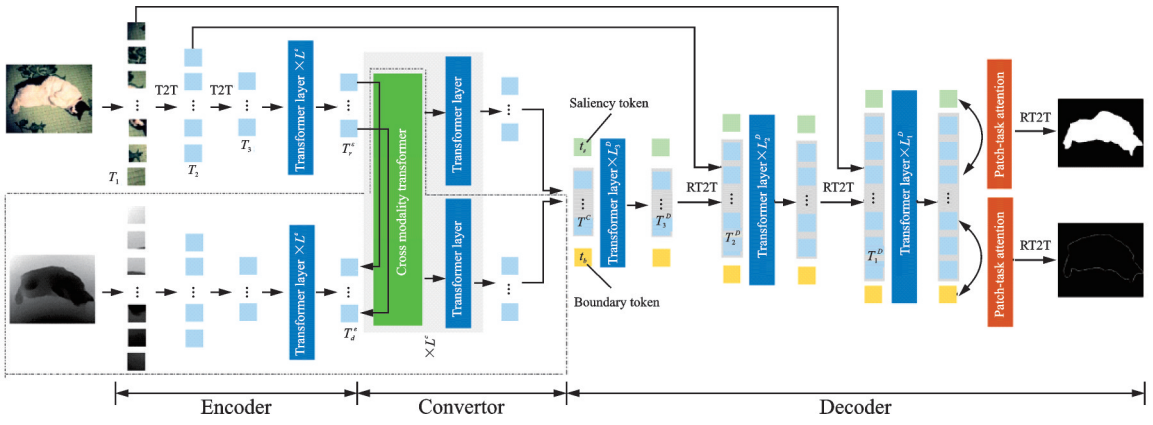


图 11　VST 模型整体架构

Fig.11　Schematic diagram of VST

(2)设计针对边界感知显著性目标检测的损失函数来控制显著目标边界的质量。低质量深度图的困难样本往往会导致生成的显著图边界模糊,这也是显著性目标检测的挑战之一,因此针对性的设计损失函数能够进一步提升模型性能。Zhang 等[63]针对困难样本提出的 SSF 模型将传统的 BCE 损失和设计的补偿感知损失相结合,通过引入补偿感知损失来帮助网络挖掘包含在跨模态特征中的结构信息,以此提高网络对困难样本的置信度,确保在面对挑战性场景时模型的鲁棒性。为解决 FCN 网络结构带来的大量细节丢失,以往的方法采用较浅层特征进行恢复或采用计算成本较大的 CRF 后处理细化结构,而 Pang 等[64]采用不同的思路,设计由 BCE 损失、边界损失和区域损失组成的新混合增强损失函数来约束边缘和前、背景区域,保证边缘区域和显著对象内部的一致性,从而实现更清晰的边界。

2.2.2　跨模态特征的有效融合

对于 RGB-D 图像显著性目标检测模型来说,避免单一使用 RGB 图像或深度图像,减少低质量深度图的影响从而提高模型鲁棒性是至关重要的,而达到这一目的的举措之一就是有效融合跨模态数据所提供的互补信息,当前工作融合策略主要分为早期融合、中期融合和后期融合 3 种类别,如图 12 所示,其中:RGB 为彩色图,Depth 为深度图,Concatenation 为融合,Convolution layer 为卷积层,Saliency map 为显著图。

早期融合主要分为以下两个类别:一是将 RGB 图像和深度图像直接融合,形成四通道输入送入检测模型,称为“输入融合”,如 Liu 等[65]采用单流递归卷积神经网络,将每一级的所有显著图融合在一起

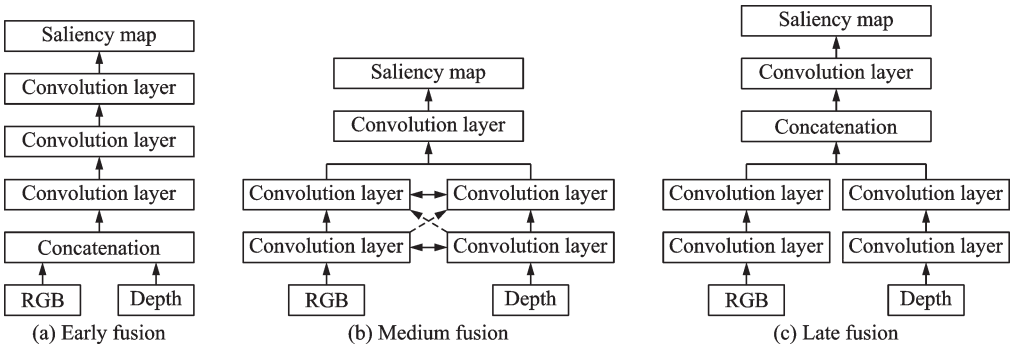


图 12　3 种跨模态特征融合策略

Fig.12　Three strategies for cross-modal feature fusion

生成最终结果。二是将RGB和深度图像的底层特征单独提取后再将其融合，输入到后续网络中，进一步实现显著性图的预测，这一类别称为“早期特征融合”，例如不同于将原始图像像素直接输入到卷积神经网络的模型，Qu等[66]通过新的卷积神经网络自动学习RGB和深度显著性特征的交互机制，将不同的底层显著性信息和分层特征进行融合，以快速准确地获取显著性区域；Fan等[67]将独立网络的方法进一步扩展，设计了简单通用的D^3Net模型，训练阶段分为处理RGB、深度和跨模态信息的3个独立子网络，测试阶段则创新地引入深度图过滤单元自动丢弃低质量深度图。

中期融合是对前期融合和后期融合的补充和完善，中期融合策略特征提取和后续融合都由较深的卷积神经网络进行处理，因此可从两种模态中学习高层语义，同时挖掘更加复杂的特征信息。除此之外，对RGB图像和深度图像进行独立的深监督也更直接，能够提取和融合具有代表性的多模态特征。因此当前的检测方法主要采用中期融合的策略，它可以分为以下两个类别：一是采用2个并行网络分别获取RGB图像和深度图像对应的显著图，这2个并行网络之间进行跨模态交互，将它们融合到特征学习网络中，2个显著图级联在一起，用于解码器网络中以获得最终的显著性预测结果。其中比较有代表性的工作有：Li等[68]为改进跨模态互补性和跨层次连续性利用不充分、不同信息源特征未区别对待的问题，独立提取RGB和深度特征后，通过跨模态特征交互模块将二者进行融合，融合后的特征经过信息转换模块处理输入解码器以获取最终的显著图；不同于以往单独从RGB和深度中独立提取特征。二是采用2个并行网络分别学习RGB图像和深度图像的特征，融合不同层中映射的RGB图像和深度图像特征，然后将它们集成到解码器网络，如残差连接（Skip connection）中以生成最终的显著图。其中比较有代表性的工作有：Fu等[35]为进一步探索跨模态信息之间的关联，采用孪生网络实现跨模态知识和参数权值共享，将深度图视为彩色图像的特例，使用共享的卷积神经网络进行RGB图像和深度图像特征的提取，使跨模态迁移成为可能；之前的研究多在跨模态特征互补上进行探索，Fan等[69]就多层次和多模态特征之间的最优融合问题进行讨论，提出一个新颖的级联细化网络，通过二分支主干策略将多尺度特征重新组合为教师特征和学生特征，通过深度增强模块从空间和通道中提取有用的深度特征信息，将RGB和深度以互补的方式融合；而Huang等[70]针对现有模型常常忽略的输入图像较差影响跨模态特征融合后辨别能力的问题进行改进，提出了新的端到端检测模型，在RGB和深度图像语义信息的引导下，生成模态权重图来对两种模态的显著性特征进行判断，同时通过双向结构来更好地捕捉跨尺度和跨模态的互补信息。

后期融合主要分为以下两个类别：一是分别采用两个并行网络分别学习RGB图像和深度图像的高层特征，然后将它们级联起来用于后续网络，以获得最终的显著性预测结果，称为“后期特征融合”，例如Han等[71]利用卷积神经网络分别学习RGB图像和深度图像的特征表示，将分别提取的特征融合后利用互补关系和联合表达增强显著性检测效果，通过全局结构损失来对模型训练进行监督。二是采用两个并行网络分别获取RGB图像和深度图像对应的显著图，然后将两个显著图级联在一起，用于解码器网络以获得最终的显著性预测结果，称为“后期结果融合”，例如Wang等[72]利用大多数突出的目标至少在一种模态下突出的特性，设计了一个双流卷积神经网络分别提取RGB和深度的特征并预测显著图，在显著性融合模块自适应地融合预测的显著图，通过边缘损失函数来对模型训练进行监督；不同于之前的方法，为了提高跨模态特征的交互性，Ding等[73]提出了一种基于RGB显著性网络、深度显著性网络和显著性融合网络的端到端深度感知显著性目标检测模型，该模型中深度显著性网络权重与RGB显著性网络某些层共享，以达到突出显著目标和抑制背景区域的目的。

对于RGB-T显著性目标检测模型来说，直接整合RGB图像和热红外图像有时会导致比使用单一模态信息更差的结果，因此探索有效的融合方式能够较大地提升显著性目标检测的质量，其中最常见的就是选择不同的网络结构。Tu等[74]尝试将显著性目标检测作为图学习问题来处理，利用分层深度

特征来共同学习显著性,将输入的RGB图像和热红外图像共同分割成一组超像素,对于每个模态和层构建一个以超像素为节点的图,每个节点用一条边连接到它的邻接点,同时自适应地融合不同的模态信息,以便更好地描述超像素之间的内在关系,捕捉显著性线索;Zhang等[75]采用端到端网络结构的思路设计RGB-T显著目标检测框架,通过多级特征融合将多模态图像信息结合在一起,以达到充分利用多模态信息深层特征的目的。以上方法仍然没有考虑来自每个模态特征图的可靠性以及多尺度特征,为了有效整合多模态互补信息,Zhang等[76]提出了由多尺度、多模态和多层次的特征融合模块组成的深度特征融合网络——HCS模型,多尺度特征融合模块从每个模态特征中捕获丰富的上下文特征,而多模态和多级特征融合模块分别集成来自不同模态特征和不同级特征的互补特征,充分利用了热红外图像对光照和遮挡的鲁棒性。随着CNN模型在显著性目标检测上的局限性以及对抗网络在视觉检测任务上的应用研究,Zhou等[77]尝试将生成对抗学习引入到RGB-T显著性目标检测中,改进了生成对抗网络(Generative adversarial network, GAN),实现了一个轻量级的有效检测框架。

除此之外,Transformer的引入也是一个新的研究思路,Liu等[78]提出了一个新的基于Swin Transformer主干网络,用于RGB-D和RGB-T任务的显著性目标检测模型SwinNet,充分发挥了卷积神经网络的局部优势和Transformer网络的长程依赖优点,同时该模型所提出的边缘引导解码器在边缘感知模块的监督下实现层间跨模态融合,生成更清晰的轮廓。

2.2.3　*弱监督方式的检测*

最新的RGB-D显著性目标检测方法依赖于大量的像素级标注数据进行训练,这种逐像素标记的注释通常依靠人工来进行,耗时且昂贵,为了大大减少人工注释同时让模型更易于扩展,不少研究者从弱监督的角度研究RGB-D显著性目标检测。

早期的弱监督方法旨在利用廉价的注释来检测显著性,例如边界框[79]、图像标签[80-81]、图像描述[82]和涂鸦标注[83-84]。在各种弱监督注释中,使用涂鸦标注训练模型达到了注释效率和模型性能之间的良好平衡。Zhang等[83]通过辅助边缘检测模块和包含门控的结构感知损失,在基于稀疏涂鸦标注的监督下保存结构信息,进而得到良好的显著性检测效果,同时建立了涂鸦标注的显著性目标检测数据集S-DUTS以促进后续相关研究;考虑到现有基于弱监督的检测模型较为复杂,Yu等[84]提出了先进行一轮无监督端到端训练的AGGM模型,通过局部相干损失,根据图像特征和像素距离将标记传播到未标记区域,从而预测具有完整对象结构的完整显著区域,训练过程通过显著性结构一致性损失作为自我监督,确保以同一幅图像的不同尺度作为输入预测一致的显著性图;Xu等[85]则通过训练具有成对RGB-D输入和前景/背景涂鸦标注的模型,其中重新标注2个广泛使用的公开可用显著性目标检测数据集NJU2K和NLPR进行模型训练,重命名为NJU2K-S和NLPR-S以供后续研究人员使用;Liu等[79]选择了不同的研究思路,放弃图像级类别标签,选择采用显著性边界框进行显著性目标检测,不仅降低了标记成本,还提供了显著目标的准确位置线索。

目前暂时没有基于弱监督的RGB-T图像显著性目标检测模型,随着研究人员对于RGB-T图像显著性目标检测方向的关注,在当前数据集数量受限的情况下,不依赖或者较少依赖大规模标注数据集的无监督和弱监督RGB-T图像显著性目标检测也是未来一个有价值的研究方向。

2.2.4　*小结*

RGB-D/T显著性目标检测在引入深度和热红外补充信息提升检测效果的同时也面临着很多挑战,为进一步达到理想的实验效果,逐步形成了以上3种常用的研究技术路线。显著图拥有完整、清晰的边缘是一个好的显著性目标检测模型标准之一,边界细化技术路线主要从网络结构和损失函数两个方面着手,目前已经取得了不错的成效,但在多目标、低对比度等较为复杂的背景以及边界精细化的显著目标分割上仍有一定研究空间;跨模态特征融合技术路线主要就抑制低质量补充信息以及有效提

取、融合补充信息和RGB信息两个方面展开并取得了不错的效果，但基于CNN的网络架构在全局信息的学习上仍有一定局限性，而全局上下文信息和全局对比度对显著性目标检测来说是非常重要的，目前已有研究人员从序列到序列的全新角度思考显著性目标检测任务，后续也可就相关方向进一步拓展；弱监督技术能够在降低高昂标注成本的情况下完成显著性目标检测任务，但在显著性区域定位的准确性和检测完整性上还有待进一步提升。正如表2和表3中所述方法在其训练集对应的测试集上MAE的表现所示，这些方法已在主流的公开数据集上取得了较理想的精度。未来可以更加关注密集型预测任务模型对于长程依赖的信息提取，针对不同的应用场景做进一步研究。除此之外，热红外图像的特殊性使得RGB-T图像显著性目标检测在实际应用中有一定前景，但该方向研究目前相对较少，

表2 RGB-D图像显著性目标检测的不同方法对比

Table 2 Comparison of different RGB-D salient object detection methods

类别		名称	训练集	骨干网络	MAE	年份	出处
针对边界细化问题进行探索	网络结构优化策略	CDNet[38]	NJU2K & NLPR	VGG-16	0.036	2021	TIP
		VST[62]	NJUD & NLPR & DUTLF-Depth	T2T-ViT	0.024	2021	CVPR
	损失函数设计	SSF[63]	DUT-RGBD & NJUD & NLPR	VGG-16	0.033	2020	CVPR
		HDFNet[64]	ImageNet	VGG-16	0.037	2020	CVPR
提高跨模态特征融合有效性	早期融合	SSRCNN[65]	NJU2K & NLPR	VGG-16	0.051	2019	Neurocomputing
		D^3Net[67]	NJU2K & NLPR	VGG-16	0.041	2019	TNNLS
	中期融合	ICNet[68]	ImageNet	VGG-16	0.052	2020	TIP
		JL-DCF[35]	NJU2K & NLPR	VGG-16	0.043	2020	CVPR
		BBS-Net[69]	NJU2K & NLPR	ResNet-50	0.035	2020	CVPR
		Bi-MCFF[70]	NJU2K & NLPR	VGG-16	0.038	2021	PR
	后期融合	MV-CNN[71]	NJU2K & NLPR	VGG-16	0.039	2017	TCYB
		AF[72]	NJU2K & NLPR	VGG-16	0.053	2019	ACCESS
		CDSF[73]	NJU2K & NLPR	VGG-16	0.054	2019	J. Vis. Commun. Image R.
基于弱监督的方式		WSSA[83]	S-DUTS	VGG16	0.047	2020	CVPR
		AGGM[84]	S-DUTS	ResNet-50	0.038	2020	arXiv
		DENet[85]	NJU2K-S & NLPR-S	VGG16	0.044	2021	TIP
		SBBs[79]	DUTS-TR	ResNet101	0.058	2021	TIP

表3 RGB-T图像显著性目标检测的不同方法对比

Table 3 Comparison of different RGB-T salient object detection methods

类别	名称	训练集	骨干网络	MAE	年份	出处
有效的跨模态特征融合	ADFC[75]	VT821 & Grayscale-thermal dataset[86] & MSRA-B dataset	VGG-16	0.025	2020	TIP
	HCS[76]	VT821 & Grayscale-thermal dataset	VGG-16	0.021	2021	TCSVT
	APNet[77]	VT5000	GAN	0.034	2021	TETCI
	SwinNet[78]	NJU2K-S & NLPR-S	Swin Transformer	0.026	2022	TCSVT

在边界优化、完整性检测等方面仍有较大的研究空间。RGB-D/T均是多模态融合的检测模型,可以充分利用深度和热红外两种模态之间的相似性,将当前丰富的RGB-D图像显著性目标检测中的多模态融合技术应用到RGB-T领域,近期不少研究在网络架构的过程中同时考虑2种模态融合[37,78]。

2.3　光场图像显著性目标检测的研究思路

光场图像显著性目标检测模型虽然研究历史还不够深厚,但也诞生了一系列有迹可寻的技术路线。由于基于深度学习的光场图像显著性目标检测模型方法较少,本文将从其所采用的输入数据类型角度来划分该任务的主要研究思路,相应检测模型的网络结构如图13所示,其中3条并行的垂直矩形表示若干个CNN网络层包括卷积层、池化层等,用于对输入进行特征提取。

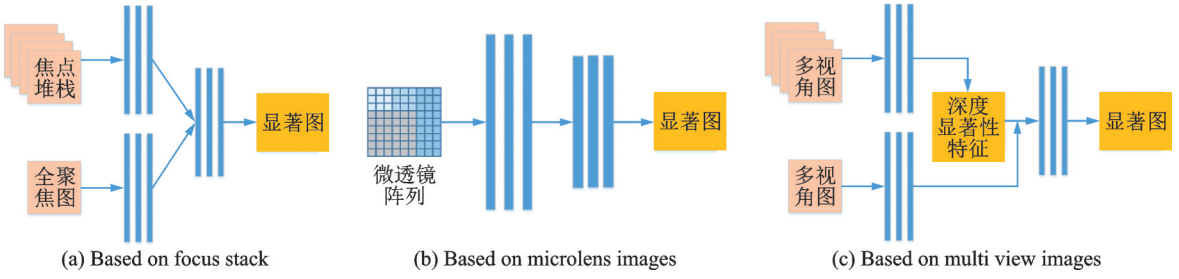


图13　3种光场显著性目标检测方法

Fig.13　Three light field salient object detection methods

2.3.1　焦点堆栈的引用

焦点堆栈指的是一系列聚焦于不同深度的图像,而全聚焦图由其对应的焦点堆栈融合而成。从图像特征上看,焦点堆栈中的各幅图像能够反映不同区域的对焦情况,即不同图像关注的区域不尽相同。目前来看,基于焦点堆栈的光场图像显著性目标检测模型是目前最常用的方法。

Zhang等[87]通过3D卷积神经网络和2D卷积神经网络分别提取焦点堆栈图像和全聚焦图像的丰富特征信息,利用通道注意力实现层间的特征融合,设计了级联的协同注意力模块实现模态间的特征交互,最后以渐进的方式融合两个模态的特征生成最终的预测图。Zhang等[88]同样是利用双流编码器分别挖掘RGB图像和焦点堆栈图像的特征,通过基于记忆的空间特征融合模块从这些光场特征中施加注意力自适应选择并使用ConvLSTM模块细化加权后的特征图,接着利用记忆机制以自上而下的方式有效集成多层光场特征。而Wang等[89]提出一种分别使用全焦图像和焦点堆栈图像作为输入的双流融合框架WLIFS,焦点堆栈图像中的每个切片通过循环注意机制自适应地进行学习,并与全焦图像中获取的特征进行融合,完成最终的预测。类似地,Zhang等[90]也以双流网络的方式同时提取全焦图像和焦点堆栈图像的特征图,采用简单但有效的细化单元来学习焦点特征,最后融合后的特征经由注意力机制和ConvLSTM模块进一步集成。可以看出,上述方法的思想是一致的,均是用独立的卷积网络对两种不同形式的输入数据进行特征提取,并在后续设计有效的特征交互集成方案。

Piao等[91]虽然也是使用两种数据作为输入,但它整体上采用的是教师-学生网络作为基础框架,核心思路主要在于知识蒸馏而不是特征融合。其中,教师网络以焦点堆栈图像为输入,轻量化的学生网络以RGB图像为输入,得益于多焦点征集模块和多焦点筛选模块,教师网络能够充分学习焦点切片中显著性特征知识,并通过双蒸馏学习策略传递给学生网络。

从上述工作的研究特点可以发现,基于焦点堆栈的方法通常需要全聚焦图像,即传统RGB图像的信息作为参考。由于不同的焦点切片展现了不同的图像特征,它能够提供帮助定位显著性目标的补充信息。故而,在利用焦点堆栈的这一技术路线上,研究人员着重探索如何有效地利用全聚焦图和焦点

堆栈经过神经网络处理后的特征并高效地融合这两类特征信息，这一思想从所提及的大多数工作，如MoLF模型[88]的双流编码器结构(图14)和ERNet[91]的教师-学生网络结构中都有所体现。图14中Mo-FIM表示功能集成模块，SFM表示空间融合模块，GPM表示全局感知模块。

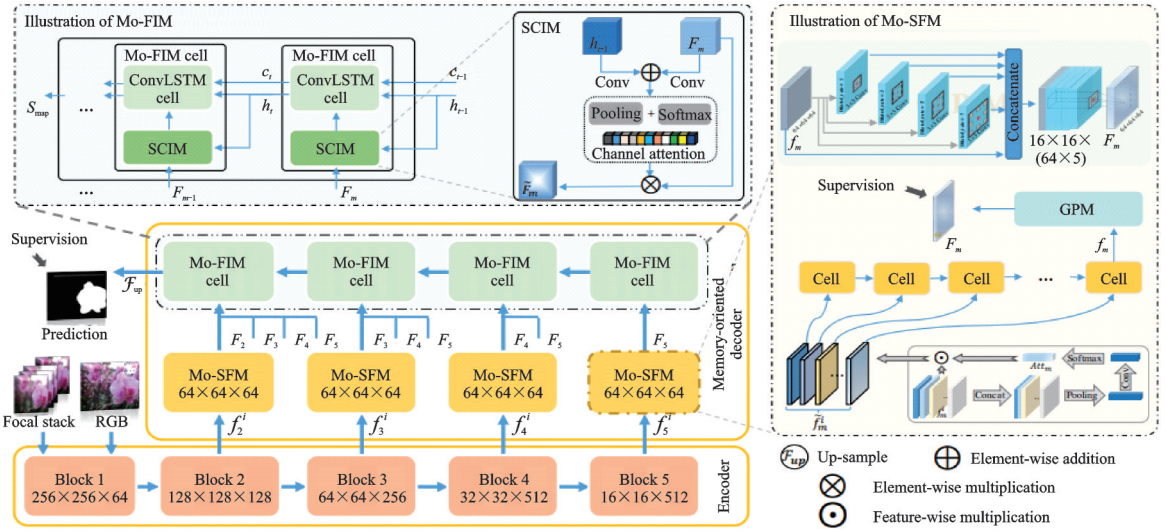


图14 MoLF模型整体架构

Fig.14 Schematic diagram of MoLF

2.3.2 微透镜图像的引用

微透镜图像包含立体空间里相应位置所有的方向信息，其携带的角度特征及其与显著或非显著线索的关系将有助于深度网络发现显著性目标。Zhang等[92]使用改进后的DeepLab-v2作为模型的骨干网络，并专门为微透镜图像阵列设计了一个角变化建模模块，该模块在每一个像素位置上以显式的卷积操作方式对角度变化进行建模。MAC模型的动机很明确，采用了卷积的方式处理局部区域上像素在多个角度变化中隐含的特征信息，将微透镜图像这一数据形式同卷积神经网络融洽地结合在一起，为后续的研究奠定了良好的基础。

2.3.3 多视角图像的引用

多视角图像能够呈现来自不同视差的目标特征，为光场图像显著性目标检测任务提供了更丰富的参考依据。Piao等[93]提出的DLSD模型将显著性目标检测分为光场合成及光场图像显著性目标检测两个子任务，首先将单视角的全焦图像通过卷积神经网络学习合成光场多视角图像，然后利用单视角图像生成的深度信息将多视角显著图变换到中央视角，最后借助注意力机制有效结合不同视角下显著性目标的协同信息。Zhang等[94]则更加简单直接地使用三维卷积强大的表示能力来建模多个视点图像之间的视差相关性，与DLSD相比，该网络架构更加简单，可以从光场的几何结构中推导出面向深度的显著性特征，同时使用空间信息、边缘信息和深度信息等多种线索之间的互补相干性搭建了一种用于光场显著目标检测的多任务协同网络，创新性地学习了其他任务中对于辅助信息的探索。

多视角图像在应用手段上常用于推理深度信息作为辅助特征，借由不同视角的视差差异，神经网络可以学习目标的深度信息，而深度信息正如2.2节中讨论的其对于显著性目标检测任务有巨大辅助作用。

2.3.4 小结

光场图像显著性目标检测领域暂未得到充分研究，其结果表现和另外两个研究分支相比，暂时还

不能令人满意,仍然有较大的发展空间。现有的基于焦点堆栈的方法常常需要额外的特征提取器来挖掘焦点切片中的显著性特征知识,这给网络带来了参数量和计算复杂度上的负担,寻求更高效的跨模态融合方法将有助于对焦点堆栈图像的充分利用。此外,多视角图像目前常被用作是深度信息的携带者,为模型提供深度特征的先验知识。这一思想同基于RGB-D的研究分支吻合,故而多视角图像这一分支上的研究可以多参考RGB-D方面的既有成果和经验。最后,微透镜图像的研究还处于初期阶段,文中MAC模型为提取角度特征设计了新颖的卷积模式,这也给后续研究带来了诸多启发,如设计适用于角度建模的注意力机制等。表4展示了上述检测方法在对应测试集上的MAE,基于焦点堆栈的方法相比其他两类方法取得更优的效果,这一方面说明了焦点堆栈对于显著性目标检测任务性能的显著提升,另一方面也提示了在微透镜和多视角方面仍存在更多可探索的空间。

表4　光场图像显著性目标检测的不同方法对比

Table 4　Comparison of different light field salient object detection methods

类别	名称	训练集	骨干网络	MAE	年份	出处
焦点堆栈	MoLF[88]	自制	VGG-19	0.089	2019	NIPS
	WLIFS[89]	DUT-LFSD	VGG-19	0.093	2019	ICCV
	ERNet[91]	DUT-LFSD & HFUT-LFSD	VGG-19	0.080	2020	AAAI
	LFNet[90]	DUT-LFSD	VGG-19	0.092	2020	TIP
	SA-Net[87]	DUT-LF & HFUT	ResNet-50	0.074	2021	BMVC
微透镜图像	MAC[92]	Lytro Illum	DeepLab	0.116	2020	TIP
多视角图像	DLSD[93]	DUTLF-MV	VGG-19	0.088	2019	IJCAI

3　数据集介绍及性能评估

3.1　数据集介绍

挖掘更为合理、更能反映真实场景的数据集一直都是研究人员尝试提升基于深度学习的显著性目标检测模型精度的一个努力方向。本小节主要介绍3个研究分支的主流数据集:RGB分支的数据集DUTS[80]、ECSSD[95]、PASCAL-S[96]、HKU-IS[12]、DUT-OMRON[97]和MSRA10K[27]等;RGB-D分支的数据集NJU2k[98]、NLPR[99]、LFSD[19]、SSD[100]和SIP[101]等;RGB-T分支的数据集VT821[32]、VT1000[74]和VT5000[102];光场分支的数据集LFSD[19]、HFUT[103]、Lytm Illum[92]等。

3.1.1　RGB图像显著性目标检测数据集

早期的数据集通常只包含一个显著物体,且背景相对简单,标注成本相对较低。但近年来,随着技术和设备逐渐成熟,更多复杂、杂乱背景中包含多个显著物体的数据集被引入来提高模型的检测性能。未来可以根据实际应用场景需求,结合7个关键点[55]创建有针对性的、更能反映真实场景的数据集。表5简单列出了RGB图像数据集的研究情况,图15给出了数据集图像示例及其真值图。

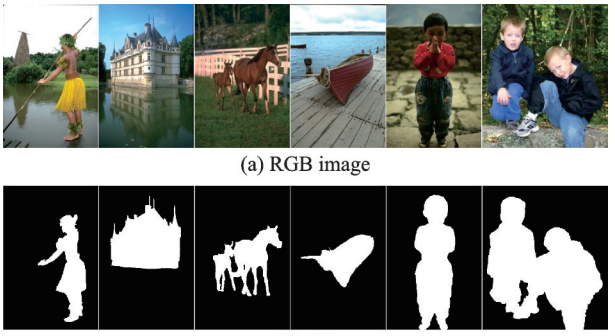
表5　RGB图像显著性目标检测主流数据集

Table 5　Mainstream datasets for RGB salient object detection

数据集	图像数量/幅	年份	标注形式	出处
MSRA[25]	20 840	2011	边界框	TPAMI
MSRA10K[27]	10 000	2012	边界框+像素标注	TPAMI
ECSSD[95]	1 000	2013	像素标注	CVPR
PASCAL-S[96]	850	2014	注视点+像素标注	CVPR
DUT-OMRON[97]	5 168	2013	像素标注	CVPR
HKU-IS[12]	4 447	2015	像素标注	CVPR
DUTS[80]	15 572	2017	像素标注	CVPR

(1)MSRA[25]是首个人工标注的显

著性目标检测数据集，旨在收集收集带有显著物体的图像，图像数量较大，20 840幅图像来自各类图像论坛和搜索引擎，通过边界框标记的方式对图像进行注释。



(a) RGB image

(b) GT

图15 RGB数据集

Fig.15 RGB dataset

(2)MSRA10K[27](也称为THUS10K)数据集由10 000幅图像组成，均选自MSRA[25]，除原有的边框标注外，额外增补了像素级标注，是目前显著性目标检测模型训练中应用最为广泛的数据集。

(3)ECSSD[95]具有复杂场景的1 000幅图像来自BSD[104]、VOC2012[105]数据集以及互联网，显著性目标的数量以一个为主，图像中前景和背景中较多的干扰信息使得该数据集具有一定的挑战性。

(4)PASCAL-S[96]从PASCAL VOC 2010[106]数据集中挑选了850幅图像，并在其原有的标注基础上增加了眼动注视点记录和显著性目标分割标记。

(5)DUT-OMRON[97]由背景复杂、内容丰富的5 168幅图像组成，每幅图像中显著目标的数量既有一个也有多个，采用二进制掩码标注显著目标，具有一定的挑战性。

(6)HKU-IS[12]带有像素级显著性目标标注，由4 447幅图像组成，具有多个显著性目标不相连、图像的边界被显著性目标所触碰、颜色对比度小于0.7的特点，对多目标显著性检测研究具有重要意义。

(7)DUTS[80]由包含10 553幅选自ImageNet[106]图像的训练集以及包含5 019幅选自ImageNet[107]和SUN[108]图像的测试集组成，并选择了50名参与者手动进行精确的像素级显著性目标标注。

3.1.2 RGB-D/T图像显著性目标检测数据集

(1) RGB-D图像显著性目标检测数据集

低质量深度图一直以来都是影响模型检测质量一个不可忽视的因素，为了获取更好的显著图，已有不少研究关注复杂和杂乱背景、拥有多个显著目标的数据集的构建，例如DUT-OMRON[97]、EC-SSD[95]、PASCAL-S[96]等数据集在标注质量和图像数量方面进行了改进；HKU-IS[12]、XPIE[109]、DUTS[80]等数据集通过手机具有多个显著目标的大量逐像素标注图像来克服标注质量不佳等缺点；Xia等[109]说明了非显著目标对显著性目标检测的重要性，进一步解决了当前数据集不包含非显著目标、无边界清晰的实例级显著目标标注的问题。但目前仍然存在数据量不够等问题。表6简单列出了RGB-D图像数据集的研究情况，图16给出了数据集图像示例及其深度图和真值图。

表6 RGB-D图像显著性目标检测主流数据集

Table 6 Mainstream datasets for RGB-D salient object detection

数据集	图像数量/幅	年份	深度图获取	出处
STERE[110]	1 000	2012	Stereo cameras	CVPR
DES[111]	135	2014	Kinect	ACM
NJU2K[98]	1 985	2014	FujiW3 camera	ICIP
NLPR[99]	1 000	2014	Microsoft Kinect	ECCV
LFSD[19]	100	2014	Lytro Illum cameras	CVPR
SSD[100]	80	2017	stereo movies	ICCVW
SIP[101]	929	2019	Huawei Mate 10	TNNLS

①STERE[110]是该研究分支首个立体图像数据集，由Flickr 1、NVIDIA 3D Vision Live 2和Stereoscopic Image Gallery三个设备采集，3名参与者人工标注图像中最为显著的目标，以3人显著区域的重合度高低对所采集到的1 250幅图像进行评估，最终选取重合度高的前1 000幅图像。

②DES[111]由135幅分辨率为640像素×640

像素的室内RGB-D图像组成,通过Kinect设备采集。在标记该数据集时要求3名参与者在每幅图像中标记出显著性目标,然后标记图像的重叠区域作为显著目标的真值。

③NLPR[99]通过Microsoft Kinect设备采集,由1 000对包含丰富室外和室内场景的RGB和深度图像组成,涵盖了办公室、商场和街道等常见场景。

④LFSD[19]由Lytro光场相机收集的100幅光场图像组成,其中60幅为室内、40幅为室外。为了标记该数据集,要求3名参与者对显著目标进行分割,当3人结果的重叠率超过90%时,将分割结果作为显著性目标的真值。

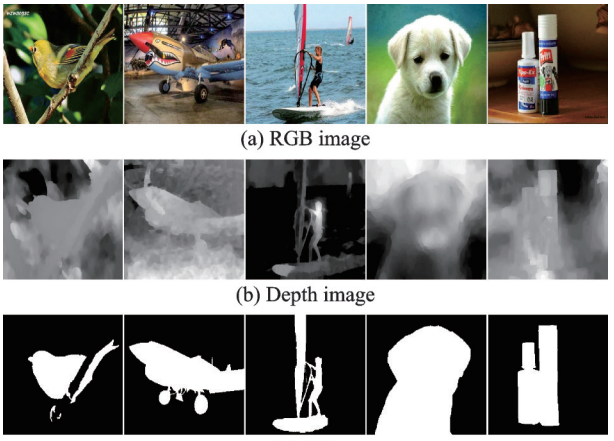


图16　RGB-D数据集

Fig.16　RGB-D dataset

⑤NJU2K[98]通过互联网、3D电影和Fuji W3立体相机获取,由1 985个立体图像对组成,数量相对较大。

⑥SSD[100]是由3部立体电影制作而成,包含室内和室外场景。该数据集共含80个样本,且每幅图像的分辨率较高,为960像素×1 080像素。

⑦SIP[101]由929幅带标注的高分辨率图像组成,每幅图像都包含多个显著性的人物。该数据集使用智能手机(华为Mate10)拍摄得到深度图。此外,该数据集还涵盖了多样性场景和各种挑战性的因素,并带有像素级标注的真值图。

(2) RGB-T图像显著性目标检测数据集

复杂场景和环境中的显著性目标检测是一个具有挑战性的研究课题,除了引入深度图,研究人员发现采用RGB图像和热红外图像作为输入使得检测模型在黑暗环境和复杂背景等现实生活中能够有较好的性能。目前对RGB-T图像显著性目标检测的研究也在逐步解决缺乏大规模数据集和综合基准限制的问题。表7简单列出了RGB-T图像数据集的研究情况,图17给出了数据集图像示例及其热红外图和真值图。

表7　RGB-T图像显著性目标检测主流数据集

Table 7　Mainstream datasets for RGB-T salient object detection

数据集	图像数量/幅	年份	深度图获取	出处
VT821[32]	821	2017	FLIR A310 and SONY TD-2073	CVPR
VT1000[74]	1 000	2019	FLIR SC620	CVPR
VT5000[102]	5 000	2020	FLIR T640 and T610	CVPR

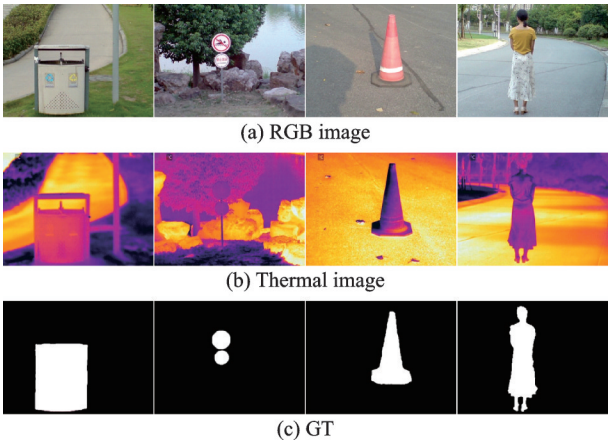


图17　RGB-T数据集

Fig.17　RGB-T dataset

①VT821[32]为2017年首个该领域的数据集,它包括由在线热像仪(FLIR A310)和CCD摄像机(SONY TD-2073)采集的821个空间对齐的RGB-T图像对以及用于显著性目标检测的真值图注释组成。图像对主要从不同环境条件下的60个场景中采集,显著目标的类别、大小、数量和空间信息也被考虑以增强检测模型的鲁棒性。但该数据集存在几个局限性:一是所使用的RGB和热红外相机具有完全不同的

成像参数，并且安装在三脚架上，它们使用单应矩阵来近似两幅图像的变换；二是对齐方法在某种形式下引入空白边界，这可能在某种程度上破坏边界；三是大多数场景非常简单，这使得数据集不那么具有挑战性和多样性。

②VT1000[74]是Tu等针对VT821数据集存在的问题进一步改进提出的，其成像硬件是前视红外FLIR SC620，内部有1个热红外摄像机和1个CCD摄像机，捕获了大约2 000个自然RGB-T图像对，手动选择了1 500个图像对。对于每幅选中的图像，6名参与者被要求选择他们第一眼看到的最显著的目标，排除了标记一致性低的图像，并选择了前1 000个图像对。最后，4名参与者使用Adobe Photoshop裁剪与热红外图像完全重叠的RGB图像，然后从每幅图像中手动分割出显著目标，以获得像素级的真值图。但该数据集仍有一定的局限性：一是由于RGB图像和热红外图像来自于相同场景的不同传感器，二者看起来不一样，需要对齐；二是RGB图像和热红外图像是静止的，不自动对齐，不可避免地会存在由于手动对齐而引入误差的问题；三是虽然VT1000数据集比VT821数据集更大，但场景的复杂性和多样性都没有得到很大的改善。

③VT5000[102]具有大规模、高分辨率、高多样性、低偏差的特点，由Ma等在2020年提出，针对以上数据集存在的问题进行改进。首先VT5000收集了5 000对不同环境下的RGB和热图像，每对RGB-T图像都是自动对齐的，并具有它们的真值图；其次，VT5000考虑了显著目标的不同大小、类别、周围环境、成像数量和空间位置，给出了一个统计结果来显示目标的多样性；最后，VT5000除了注释具有挑战的属性外，还注释了数据集中目标的成像质量，因为目标成像质量的注释为弱监督RGB-T显著目标检测提供了标签，为下一步工作奠定了基础。但对于较深的网络以及与目标检测、分类任务相比，该数据集仍然缺乏足够数量的数据集进行训练，且对于人类真实生活场景的涵盖情况也较少。

表8 光场图像显著性目标检测常用数据集

Table 8 Mainstream datasets for light field salient object detection

数据集	图像数量/幅	年份	数据采集	出处
LFSD[19]	100	2019	LYTRO光场相机	IJCAI
HFUT[103]	255	2017	LYTRO光场相机	TPAMI
DUTLF[89]	1 465	2019	LYTRO ILLUM相机	ICCV
DUTLF-MV[93]	1 580	2019	LYTRO ILLUM相机	IJCAI
Lytro Illum[92]	640	2020	LYTRO ILLUM相机	TIP

3.1.3 光场图像显著性目标检测数据集

由于硬件采集设备的要求较高且采集处理复杂，光场图像显著性目标检测任务基准数据集目前仅有5个，规模最大的包含1 580个样本，规模最小的仅有100个样本，且图像分辨率均较低，真实场景、难点场景覆盖面低，给检测模型的训练和优化带来一定难度。此外，不同采集设备拍摄时的角度分辨率和空间分辨率均不一样，会影响不同数据集之间的统一性，而不同数据集对光场数据的处理形式也差异较大，因此构建一个大规模且统一的基准数据集是光场图像显著性目标检测任务急需解决的问题。表8简单列出了光场图像数据集的研究情况，图18给出了数据集图像示例及其焦点图和真值图。

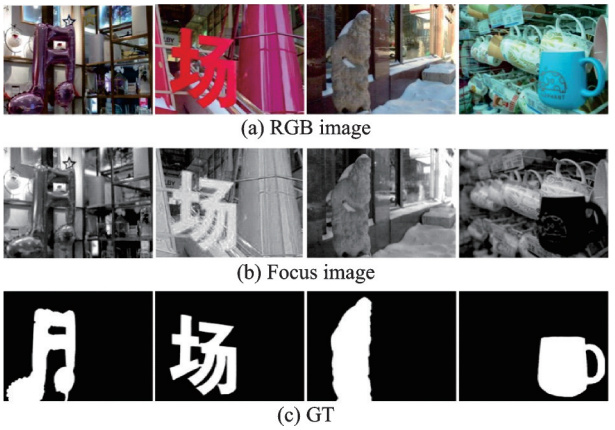


图18 光场数据集

Fig.18 Light field dataset

(1)LFSD[19]通过Lytro光场相机采集获取,其中包含360像素×360像素分辨率、60个室内和40个室外场景的光场图像,并且大部分场景只包含一个显著性目标。此外,通过3名参与者对显著目标进行分割,当3个结果的重叠率超过90%时才标注该分割结果为显著性目标的真值。

(2)HFUT[103]由Lytro光场相机获取,包含255幅光场图像。大部分场景是复杂且杂乱的,都包含位置和大小均不一致的多个显著性目标,具有一定的挑战性。

(3)DUTLF[89]含1 465个样本,其中1 000个样本为训练样本,余下的465个样本为测试样本。该数据集图像分辨率均为600像素×400像素,并且包含如下挑战:显著目标与背景之间对比度较低,包含多个不相邻的显著性目标,包括黑暗或强光照的光照条件。

(4)DUTLF-MV[93]由Lytro Illum相机采集,包含1 580个样本,其中1 100个为训练样本,余下的为测试样本。该数据集中每个光场都包含多个视角的图像并且有一个对应的真值图。

(5)Lytro Illum[92]由640个光场和相应的像素级显著性真值组成,该数据集包括几个具有挑战性的因素:如不一致的光照状况,背景相似或背景杂乱,所包含显著目标较小。

3.2　评估方法

3.2.1　PR曲线

精度(Precision)和召回率(Recall)也被称作查准率和查全率[26],可通过二值化下预测的显著性掩码和实际的显著性掩码计算得到,其表达式分别为

$$\text{Precision}=\frac{\text{TP}}{\text{TP}+\text{FP}} \tag{1}$$

$$\text{Recall}=\frac{\text{TP}}{\text{TP}+\text{FN}} \tag{2}$$

式中:TP(True positive)表示被模型预测为正的正样本个数;FP(False positive)表示被模型误判为正的负样本个数;FN(False negative)表示被模型误判为负的正样本个数。通过应用从0到255的阈值将显著性目标检测图二值化,每个阈值产生一组精度与召回值,从而描绘出呈现模型性能的Precision-Recall(PR)曲线。通常而言,曲线越贴近右上方,模型性能就越好。在显著性目标检测中,若将显著图转化为的二值图定义为B,真实值为G,则显著图的PR曲线计算公式为

$$\text{Precision}=\frac{|B\cap G|}{|B|} \tag{3}$$

$$\text{Recall}=\frac{|B\cap G|}{|G|} \tag{4}$$

3.2.2　F-measure

通常情况下,精度和召回率都无法完全评估显著性映射的质量,因此选择采用F-measure指标[14]就准确度和完整度进行综合判断,引入非负权重β^2来平衡精确率和召回率之间的关系,计算公式为

$$F_\beta=\frac{\left(1+\beta^2\right)\times\text{Precision}\times\text{Recall}}{\beta^2\times\text{Precision}+\text{Recall}} \tag{5}$$

式中β^2根据很多显著性目标检测工作经验常设置为0.3,提高精度在评估中对结果的影响,因为召回率并不像精确率那样重要。在显著性目标检测研究中,有Mean F-measure和Max F-measure两种计算。Mean F-measure为采取自适应阈值进行二值化时,通过计算每幅图像F_β的平均值得到;Max F-measure为采取固定阈值进行二值化时,通过计算PR曲线上的最大值得到。

3.2.3　平均绝对误差MAE

所得的显著图与真值图之间像素的MAE[112]计算方法为

$$\mathrm{MAE}=\frac{1}{W\times H}\sum_{x=1}^{W}\sum_{y=1}^{H}\left|S(x,y)-G(x,y)\right| \tag{6}$$

式中：W和H分别表示图像的宽度和高度；$S(x,y)$为模型检测得到的显著图；$G(x,y)$为人工标注的真实显著图。一般地，MAE的值越小，模型的性能也就越好。

3.2.4 S-measure指标

MAE和F-measure指标忽略了对预测图中结构信息的评估，而这一信息在人类视觉系统对场景的感知中是极其敏感的，因此结构信息的评估对显著性目标检测任务尤为重要。Fan等[113]基于结构相似性指标(Structural similarity, SSIM)的研究，提出了综合考虑区域和物体结构相似度的S-measure指标，表达式为

$$S_\alpha=\alpha S_0+(1-\alpha)S_r \tag{7}$$

式中：S_r为区域结构相似性度量；S_0为物体结构相似性度量；α是取自区间[0,1]的平衡参数，通常取$\alpha=0.5$。

4 总结与展望

本文对近5年来不同模态的3个研究分支RGB图像、RGB-D/T图像和光场图像显著性目标检测的研究难点以及研究思路进行了总结，并介绍了3个研究分支常用的数据集以及显著性目标检测常用的评估指标。与手工提取特征的传统方法相比，基于深度学习的方法能够从大量的数据中提取特征且具有较强的泛化能力，可以准确地定位显著物体，应对复杂、杂乱场景的能力也较强。但与此同时，也存在数据集与真实检测环境有差距、显著性目标检测模型易用性相对较弱、CNN网络结构有局限性等问题，仍有一定的改进空间，具体如下：

(1)更高质量、更全面的数据集。目前虽然各个分支具有对应的数据集，但仍存在图像类似、对于复杂情况和真实情况的反映较少、代表性不够强等问题，实例级标注的数据集对于多目标显著性检测以及图像编辑、视频压缩等场景有着重要作用，但目前暂未获得较多关注。RGB-T和光场图像显著性目标检测研究还不够充分，数据集的创建和数据量远远不够，其中光场数据集表示方式不一，给基准测试带来一定难度，开发大规模、统一的光场数据集对于未来的研究有重要意义。除此之外，针对路标识别等特定任务，需要对特定的显著性目标图像进行收集，构建任务型驱动的数据集，以进一步提高模型性能。

(2)适应多场景、高性能的检测模型。随着显著性目标检研究的不断发展，CNN网络结构的局限性也逐渐凸显出来，单纯的网络结构变形已不能满足当下的需求。近期，诸如Transformer、GAN等深度学习模型也被逐步扩展应用到显著性目标检测领域中，未来在进一步探索2个模型在各个分支应用的同时，其他深度学习模型在密集型检测任务上的应用也可进行拓展研究。除此之外，不断加深的网络层数也带来了模型参数量大、训练速度慢等问题，给模型部署到边缘设备带来了较大的困难，因此设计轻量化的网络模型，在降低模型计算复杂度和空间复杂度的同时保证模型性能对于深度学习技术产业化是未来非常有价值的研究方向。

(3)大幅提升跨模态融合效率的监督策略。现有的显著性目标检测模型仍以全监督学习策略居多，能够通过学习局部、全局和上下文信息来挑选显著的特征，进而更好地获取显著区域。但标注像素级显著图是一个耗时、繁琐的过程，无监督、弱监督和自监督的研究能够有效降低研究人员在数据集标注上所花费的昂贵代价，目前RGB和RGB-D图像显著性目标检测领域已有不少研究人员进行探索，而在跨模态融合领域RGB-T方向以及光场图像显著性目标检测领域暂无弱监督和无监督的相关研究。值得注意的是，相较于全监督，弱监督和半监督的特征提取能力明显弱化，因此如何在降低人工标注成

本的同时保证模型检测性能是值得展开的研究工作,尤其在多模态信息融合上,充分利用补充信息进一步获取显著特征,以达到甚至超过全监督模型的性能。

参考文献:

[1] NIE G Y, CHENG M M, LIU Y, et al. Multi-level context ultra-aggregation for stereo matching[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2019: 3283-3291.

[2] ZHU J Y, WU J, XU Y, et al. Unsupervised object class discovery via saliency-guided multiple class learning[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 37(4): 862-875.

[3] RAPANTZIKOS K, AVRITHIS Y, KOLLIAS S. Dense saliency-based spatiotemporal feature points for action recognition [C]//Proceedings of 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami, FL, USA: IEEE, 2009: 1454-1461.

[4] WANG W, SHEN J, YANG R, et al. Saliency-aware video object segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 40(1): 20-33.

[5] HOYER L, MUNOZ M, KATIYAR P, et al. Grid saliency for context explanations of semantic segmentation[J]. Advances in Neural Information Processing Systems, 2019, 32: 6462-6473.

[6] WU Y H, GAO S H, MEI J, et al. JCS: An explainable covid-19 diagnosis system by joint classification and segmentation[J]. IEEE Transactions on Image Processing, 2021, 30: 3113-3126.

[7] HONG S, YOU T, KWAK S, et al. Online tracking by learning discriminative saliency map with convolutional neural network [C]//Proceedings of International Conference on Machine Learning. [S.l.]: PMLR, 2015: 597-606.

[8] ZHAO R, OYANG W, WANG X. Person re-identification by saliency learning[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(2): 356-370.

[9] FAN D P, JI G P, SUN G, et al. Camouflaged object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, WA, USA: IEEE, 2020: 2777-2787.

[10] LIU G, FAN D. A model of visual attention for natural image retrieval[C]//Proceedings of 2013 International Conference on Information Science and Cloud Computing Companion. Guangzhou, China: IEEE, 2013: 728-733.

[11] ITTI L, KOCH C, NIEBUR E. A model of saliency-based visual attention for rapid scene analysis[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20(11): 1254-1259.

[12] LI G, YU Y. Visual saliency based on multiscale deep features[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA,USA: IEEE, 2015: 5455-5463.

[13] FU K, JIANG Y, JI G P, et al. Light field salient object detection: A review and benchmark[J]. Computational Visual Media, 2022. DOI: 10.48550/arXiv.2010.04968.

[14] BORJI A, CHENG M M, JIANG H, et al. Salient object detection: A benchmark[J]. IEEE Transactions on Image Processing, 2015, 24(12): 5706-5722.

[15] HAN J, ZHANG D, CHENG G, et al. Advanced deep-learning techniques for salient and category-specific object detection: A survey[J]. IEEE Signal Processing Magazine, 2018, 35(1): 84-100.

[16] ZHOU T, FAN D P, CHENG M M, et al. RGB-D salient object detection: A survey[J]. Computational Visual Media, 2021, 7(1): 37-69.

[17] 罗会兰, 袁璞, 童康. 基于深度学习的显著性目标检测方法综述[J]. 电子学报, 2021, 49(7): 1417-1427.
LUO Huilan, YUAN Pu, TONG Kang. Review of the methods for salient object detection based on deep learning[J]. Acta Electronica Sinica, 2021, 49(7): 1417-1427.

[18] 史彩娟, 张卫明, 陈厚儒, 等. 基于深度学习的显著性目标检测综述[J]. 计算机科学与探索, 2021, 15(2): 219-232.
SHI Caijuan, ZHANG Weiming, CHEN Houru, et al. Survey of salient object detection based on deep learning[J]. Journal of Frontiers of Computer Science and Technology, 2021, 15(2): 219-232.

[19] LI N, YE J, JI Y, et al. Saliency detection on light field[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2014: 2806-2813.

[20] TANG L, LI B, ZHONG Y, et al. Disentangled high quality salient object detection[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal, QC, Canada: IEEE, 2021: 3580-3590.

[21] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2015: 3431-3440.

[22] ZHAO J X, LIU J J, FAN D P, et al. EGNet: Edge guidance network for salient object detection[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE, 2019: 8779-8788.

[23] WU Z, SU L, HUANG Q. Cascaded partial decoder for fast and accurate salient object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, CA, USA: IEEE, 2019: 3907-3916.

[24] QIN X, ZHANG Z, HUANG C, et al. Basnet: Boundary-aware salient object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, CA, USA: IEEE, 2019: 7479-7489.

[25] LIU T, YUAN Z, SUN J, et al. Learning to detect a salient object[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 33(2): 353-367.

[26] ACHANTA R, HEMAMI S, ESTRADA F, et al. Frequency-tuned salient region detection[C]//Proceedings of 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami, FL, USA: IEEE, 2009: 1597-1604.

[27] 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, 2014, 37(3): 569-582.

[28] 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, 2011, 34(2): 315-327.

[29] WU Y H, LIU Y, XU J, et al. MobileSal: Extremely efficient RGB-D salient object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 44(12): 10261-10269.

[30] ZHANG X, ZHOU X, LIN M, et al. ShuffleNet: An extremely efficient convolutional neural network for mobile devices[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE, 2018: 6848-6856.

[31] MA N, ZHANG X, ZHENG H T, et al. ShuffleNet v2: Practical guidelines for efficient CNN architecture design[C]// Proceedings of the European Conference on Computer Vision (ECCV). [S.l.]: Springer, 2018: 116-131.

[32] WANG G, LI C, MA Y, et al. RGB-T saliency detection benchmark: Dataset, baselines, analysis and a novel approach[C]// Proceedings of Chinese Conference on Image and Graphics Technologies. Singapore: Springer, 2018: 359-369.

[33] PANG Y, ZHAO X, ZHANG L, et al. Multi-scale interactive network for salient object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, WA, USA: IEEE, 2020: 9413-9422.

[34] LIU Y, ZHANG X Y, BIAN J W, et al. SAMNet: Stereoscopically attentive multi-scale network for lightweight salient object detection[J]. IEEE Transactions on Image Processing, 2021, 30: 3804-3814.

[35] FU K, FAN D P, JI G P, et al. JL-DCF: Joint learning and densely-cooperative fusion framework for RGB-D salient object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, WA, USA: IEEE, 2020: 3052-3062.

[36] ZHOU T, FU H, CHEN G, et al. Specificity-preserving rgb-d saliency detection[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal, QC, Canada: IEEE, 2021: 4681-4691.

[37] GAO W, LIAO G, MA S, et al. Unified information fusion network for multi-modal RGB-D and RGB-T salient object detection[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 32(4): 2091-2106.

[38] JIN W D, XU J, HAN Q, et al. CDNet: Complementary depth network for RGB-D salient object detection[J]. IEEE Transactions on Image Processing, 2021, 30: 3376-3390.

[39] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation[C]// Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer , 2015: 234-241.

[40] CHEN L C, PAPANDREOU G, SCHROFF F, et al. Rethinking atrous convolution for semantic image segmentation[EB/OL].(2017-01-01)[2022-03-15]. https://arxiv.org/abs/1706.05587.

[41] ZHAO H, SHI J, QI X, et al. Pyramid scene parsing network[C]//Proceedings of the IEEE Conference on Computer Vision

and Pattern Recognition. Honolulu, HI, USA: IEEE, 2017: 2881-2890.

[42] SUN H, BIAN Y, LIU N, et al. Multi-scale edge-based U-shape network for salient object detection[C]//Proceedings of Pacific Rim International Conference on Artificial Intelligence. Cham: Springer, 2021: 501-514.

[43] SHEN K, ZHOU X, WAN B, et al. Fully squeezed multiscale inference network for fast and accurate saliency detection in optical remote-sensing images[J]. IEEE Geoscience and Remote Sensing Letters, 2022, 19: 1-5.

[44] WEI J, WANG S, HUANG Q. F^3Net: Fusion, feedback and focus for salient object detection[C]//Proceedings of the AAAI Conference on Artificial Intelligence. [S.l.]: AAAI, 2020: 12321-12328.

[45] CHEN Z, XU Q, CONG R, et al. Global context-aware progressive aggregation network for salient object detection[C]//Proceedings of the AAAI Conference on Artificial Intelligence. [S.l.]: AAAI, 2020: 10599-10606.

[46] LIU J J, HOU Q, CHENG M M, et al. A simple pooling-based design for real-time salient object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, CA, USA: IEEE, 2019: 3917-3926.

[47] ZHAO T, WU X. Pyramid feature attention network for saliency detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, CA, USA: IEEE, 2019: 3085-3094.

[48] LI J, SU J, XIA C, et al. Salient object detection with purificatory mechanism and structural similarity loss[J]. IEEE Transactions on Image Processing, 2021, 30: 6855-6868.

[49] LIN G, LIU Z, LIN W, et al. Multi-content complementation network for salient object detection in optical remote sensing images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021. DOI: 10.1109/TGRS.2021.3131221.

[50] CONG R, ZHANG Y, FANG L, et al. RRNet: Relational reasoning network with parallel multiscale attention for salient object detection in optical remote sensing images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 60: 1-11.

[51] ZHANG Q, CONG R, LI C, et al. Dense attention fluid network for salient object detection in optical remote sensing images [J]. IEEE Transactions on Image Processing, 2020, 30: 1305-1317.

[52] CHEN S, TAN X, WANG B, et al. Reverse attention for salient object detection[C]//Proceedings of the European Conference on Computer Vision (ECCV). [S.l.]: Springer, 2018: 234-250.

[53] TANG L, LI B. Class: Cross-level attention and supervision for salient objects detection[C]//Proceedings of the Asian Conference on Computer Vision. [S.l.]: [s.n.], 2020.

[54] KE Y Y, TSUBONO T. Recursive contour-saliency blending network for accurate salient object detection[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. Waikoloa, HI, USA: IEEE, 2022: 2940-2950.

[55] YANG Z, SOLTANIAN-ZADEH S, FARSIU S. BiconNet: An edge-preserved connectivity-based approach for salient object detection[J]. Pattern Recognition, 2022, 121: 108231.

[56] WU Z, SU L, HUANG Q. Decomposition and completion network for salient object detection[J]. IEEE Transactions on Image Processing, 2021, 30: 6226-6239.

[57] LIU J J, HOU Q, CHENG M M. Dynamic feature integration for simultaneous detection of salient object, edge, and skeleton [J]. IEEE Transactions on Image Processing, 2020, 29: 8652-8667.

[58] GAO S, ZHANG W, WANG Y, et al. Weakly-supervised salient object detection using point supervison[EB/OL].(2022-07-22)[2022--3-15]. https://arxiv.org/abs/2203.11652v1.

[59] PIAO Y, WANG J, ZHANG M, et al. MFNet: Multi-filter directive network for weakly supervised salient object detection [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal, QC, Canada: IEEE, 2021: 4136-4145.

[60] ZHANG J, ZHANG T, DAI Y, et al. Deep unsupervised saliency detection: A multiple noisy labeling perspective[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE, 2018: 9029-9038.

[61] YAN P, WU Z, LIU M, et al. Unsupervised domain adaptive salient object detection through uncertainty-aware pseudo-label learning[EB/OL]. (2022-02-26)[2022-03-15]. https://arxiv.org/abs/2202.13170.

[62] LIU N, ZHANG N, WAN K, et al. Visual saliency transformer[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal, QC, Canada: IEEE, 2021: 4722-4732.

[63] ZHANG M, REN W, PIAO Y, et al. Select, supplement and focus for RGB-D saliency detection[C]//Proceedings of the

IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, WA, USA: IEEE, 2020: 3472-3481.

[64] PANG Y, ZHANG L, ZHAO X, et al. Hierarchical dynamic filtering network for RGB-D salient object detection[C]// Proceedings of European Conference on Computer Vision. Cham: Springer, 2020: 235-252.

[65] LIU Z, SHI S, DUAN Q, et al. Salient object detection for RGB-D image by single stream recurrent convolution neural network[J]. Neurocomputing, 2019, 363: 46-57.

[66] QU L, HE S, ZHANG J, et al. RGBD salient object detection via deep fusion[J]. IEEE Transactions on Image Processing, 2017, 26(5): 2274-2285.

[67] FAN D P, LIN Z, ZHANG Z, et al. Rethinking RGB-D salient object detection: Models, data sets, and large-scale benchmarks[J]. IEEE Transactions on Neural Networks and Learning Systems, 2020, 32(5): 2075-2089.

[68] LI G, LIU Z, LING H. ICNet: Information conversion network for RGB-D based salient object detection[J]. IEEE Transactions on Image Processing, 2020, 29: 4873-4884.

[69] FAN D P, ZHAI Y, BORJI A, et al. BBS-Net: RGB-D salient object detection with a bifurcated backbone strategy network [C]//Proceedings of European Conference on Computer Vision. Cham: Springer, 2020: 275-292.

[70] HUANG N, LUO Y, ZHANG Q, et al. Discriminative unimodal feature selection and fusion for RGB-D salient object detection[J]. Pattern Recognition, 2022, 122: 108359.

[71] HAN J, CHEN H, LIU N, et al. CNNs-based RGB-D saliency detection via cross-view transfer and multiview fusion[J]. IEEE Transactions on Cybernetics, 2017, 48(11): 3171-3183.

[72] WANG N, GONG X. Adaptive fusion for RGB-D salient object detection[J]. IEEE Access, 2019, 7: 55277-55284.

[73] DING Y, LIU Z, HUANG M, et al. Depth-aware saliency detection using convolutional neural networks[J]. Journal of Visual Communication and Image Representation, 2019, 61: 1-9.

[74] TU Z, XIA T, LI C, et al. RGB-T image saliency detection via collaborative graph learning[J]. IEEE Transactions on Multimedia, 2019, 22(1): 160-173.

[75] ZHANG Q, HUANG N, YAO L, et al. RGB-T salient object detection via fusing multi-level CNN features[J]. IEEE Transactions on Image Processing, 2019, 29: 3321-3335.

[76] ZHANG Q, XIAO T, HUANG N, et al. Revisiting feature fusion for RGB-T salient object detection[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2020, 31(5): 1804-1818.

[77] ZHOU W, ZHU Y, LEI J, et al. APNet: Adversarial learning assistance and perceived importance fusion network for all-day RGB-T salient object detection[J]. IEEE Transactions on Emerging Topics in Computational Intelligence, 2021, 6(4): 957-968.

[78] LIU Z, TAN Y, HE Q, et al. SwinNet: Swin transformer drives edge-aware RGB-D and RGB-T salient object detection[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 32(7): 4486-4497.

[79] LIU Y, WANG P, CAO Y, et al. Weakly-supervised salient object detection with saliency bounding boxes[J]. IEEE Transactions on Image Processing, 2021, 30: 4423-4435.

[80] WANG L, LU H, WANG Y, et al. Learning to detect salient objects with image-level supervision[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA: IEEE, 2017: 136-145.

[81] LI G, XIE Y, LIN L. Weakly supervised salient object detection using image labels[C]//Proceedings of the AAAI Conference on Artificial Intelligence. [S.l.]: AAAI, 2018. DOI:10.48550/arXiv.1803.06503.

[82] ZHANG H, ZENG Y, LU H, et al. Learning to detect salient object with multi-source weak supervision[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(7): 3577-3589.

[83] ZHANG J, YU X, LI A, et al. Weakly-supervised salient object detection via scribble annotations[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, WA, USA: IEEE, 2020: 12546-12555.

[84] YU S, ZHANG B, XIAO J, et al. Structure-consistent weakly supervised salient object detection with local saliency coherence [C]//Proceedings of the AAAI Conference on Artificial Intelligence. [S.l.]: AAAI, 2021: 3234-3242.

[85] XU Y, YU X, ZHANG J, et al. Weakly supervised RGB-D salient object detection with prediction consistency training and active scribble boosting[J]. IEEE Transactions on Image Processing, 2022, 31: 2148-2161.

[86] LI C, WANG X, ZHANG L, et al. Weighted low-rank decomposition for robust grayscale-thermal foreground detection[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2016, 27(4): 725-738.

[87] ZHANG Y, CHEN G, CHEN Q, et al. Learning synergistic attention for light field salient object detection[EB/OL]. (2021-10-22)[2022-03-15]. https://arxiv.org/abs/2104.13916v2.

[88] ZHANG M, LI J, WEI J, et al. Memory-oriented decoder for light field salient object detection[J]. Advances in Neural Information Processing Systems, 2019, 32: 1-11.

[89] WANG T, PIAO Y, LI X, et al. Deep learning for light field saliency detection[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE, 2019: 8838-8848.

[90] ZHANG M, JI W, PIAO Y, et al. LFNet: Light field fusion network for salient object detection[J]. IEEE Transactions on Image Processing, 2020, 29: 6276-6287.

[91] PIAO Y, RONG Z, ZHANG M, et al. Exploit and replace: An asymmetrical two-stream architecture for versatile light field saliency detection[C]//Proceedings of the AAAI Conference on Artificial Intelligence. [S.l.]: AAAI, 2020: 11865-11873.

[92] ZHANG J, LIU Y, ZHANG S, et al. Light field saliency detection with deep convolutional networks[J]. IEEE Transactions on Image Processing, 2020, 29: 4421-4434.

[93] PIAO Y, RONG Z, ZHANG M, et al. Deep light-field-driven saliency detection from a single view[C]//Proceedings of the 28th International Joint Conference on Artificial Intelligence. [S.l.]: ACM, 2019: 904-911.

[94] ZHANG Q, WANG S, WANG X, et al. A multi-task collaborative network for light field salient object detection[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2020, 31(5): 1849-1861.

[95] YAN Q, XU L, SHI J, et al. Hierarchical saliency detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Portland, OR, USA: IEEE, 2013: 1155-1162.

[96] LI Y, HOU X, KOCH C, et al. The secrets of salient object segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH, USA: IEEE, 2014: 280-287.

[97] YANG C, ZHANG L, LU H, et al. Saliency detection via graph-based manifold ranking[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Portland, OR, USA: IEEE, 2013: 3166-3173.

[98] JU R, GE L, GENG W, et al. Depth saliency based on anisotropic center-surround difference[C]//Proceedings of 2014 IEEE International Conference on Image Processing (ICIP). Paris, France: IEEE, 2014: 1115-1119.

[99] PENG H, LI B, XIONG W, et al. RGBD salient object detection: A benchmark and algorithms[C]//Proceedings of European Conference on Computer Vision. Cham: Springer, 2014: 92-109.

[100] ZHU C, LI G. A three-pathway psychobiological framework of salient object detection using stereoscopic technology[C]//Proceedings of the IEEE International Conference on Computer Vision Workshops. Venice, Italy: IEEE, 2017: 3008-3014.

[101] FAN D P, LIN Z, ZHANG Z, et al. Rethinking RGB-D salient object detection: Models, data sets, and large-scale benchmarks[J]. IEEE Transactions on Neural Networks and Learning Systems, 2020, 32(5): 2075-2089.

[102] TU Z, MA Y, LI Z, et al. RGBT salient object detection: A large-scale dataset and benchmark[J]. IEEE Transactions on Multimedia, 2022. DOI: 10.1109/TMM.2022.3171688.

[103] ZHANG J, WANG M, LIN L, et al. Saliency detection on light field: A multi-cue approach[J]. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 2017, 13(3): 1-22.

[104] 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]//Proceedings of Eighth IEEE International Conference on Computer Vision. Vancouver, BC, Canada: IEEE, 2001, 2: 416-423.

[105] EVERINGHAM M, WINN J. The PASCAL visual object classes challenge 2012 (VOC2012) development kit[M]. [S.l.]: [s.n.], 2012: 1-45.

[106] EVERINGHAM M, VAN GOOL L, WILLIAMS C K I, et al. The pascal visual object classes (VOC) challenge[J]. International Journal of Computer Vision, 2010, 88(2): 303-338.

[107] DENG J, DONG W, SOCHER R, et al. Imagenet: A large-scale hierarchical image database[C]//Proceedings of 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami, FL, USA: IEEE, 2009: 248-255.

[108] XIAO J, HAYS J, EHINGER K A, et al. Sun database: Large-scale scene recognition from abbey to zoo[C]//Proceedings of 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2010: 3485-3492.

[109] XIA C, LI J, CHEN X, et al. What is and what is not a salient object? learning salient object detector by ensembling linear

exemplar regressors[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2017: 4142-4150.

[110] NIU Y, GENG Y, LI X, et al. Leveraging stereopsis for saliency analysis[C]//2012 IEEE Conference on Computer Vision and Pattern Recognition. San Francisco, CA, USA: IEEE, 2012: 454-461.

[111] CHENG Y, FU H, WEI X, et al. Depth enhanced saliency detection method[C]//Proceedings of International Conference on Internet Multimedia Computing and Service. [S.l.]: ACM, 2014: 23-27.

[112] PERAZZI F, KRÄHENBÜHL P, PRITCH Y, et al. Saliency filters: Contrast based filtering for salient region detection[C]//Proceedings of 2012 IEEE Conference on Computer Vision and Pattern Recognition. Providence, RI, USA: IEEE, 2012: 733-740.

[113] FAN D P, CHENG M M, LIU Y, et al. Structure-measure: A new way to evaluate foreground maps[C]//Proceedings of the IEEE International Conference on Computer Vision. Venice, Italy: IEEE, 2017: 4548-4557.

作者简介:

孙涵(1978-),男,副教授,硕士生导师,研究方向:计算机视觉和模式识别等,E-mail:sunhan@nuaa.edu.cn。

刘译善(1996-),**通信作者**,女,硕士研究生,研究方向:显著性目标检测,E-mail:liuyishan@nuaa.edu.cn。

林昱涵(1999-),男,硕士研究生,研究方向:显著性目标检测。

(编辑:张黄群)