Computational Visual Media
DOI 10.1007/s41095-xxx-xxxx-x Vol. x, No. x, ** 2020, xx–xx

Review Article

RGB-D 视觉显著性物体检测综述

周涛 [1], 范登平 [1](✉), 程明明 [2], 沈建冰 [1], 邵岭 [1]

Abstract 显著性物体检测（Salient object detection，SOD）可模拟人类视觉感知系统找到最引人注意的物体，已被广泛应用于各类计算机视觉任务。目前，随着深度传感器的广泛应用，可以更容易地获取具有丰富空间信息的深度图像，并有利于提高 SOD 的性能。尽管在过去的几年中已经提出了一些基于 RGB-D 的具有良好性能的 SOD 模型，但仍缺乏对这些模型的深入理解以及对该领域面临的挑战分析。为此，本文旨在对基于 RGB-D 的 SOD 模型进行全面系统的综述总结并详细评价了相关的基准数据集。由于光场图像也能提供深度信息，所以本文回顾总结了基于光场的 SOD 模型和主流的基准数据集。除此之外，为了充分了解现有 SOD 模型的性能，本文不仅进行了全面的性能评估还对几个具有代表性的基于 RGB-D 的 SOD 模型进行了详细的属性分析。最后，对基于 RGB-D 的 SOD 未来的研究方向及挑战进行了展望与总结。所有收集的模型、基准数据集、源代码链接、为属性评价而构建的数据集以及评估代码等详见 https://github.com/taozh2017/RGBD-SODsurvey。

Keywords RGB-D 显著性物体检测，显著性检测，综合评估，光场.

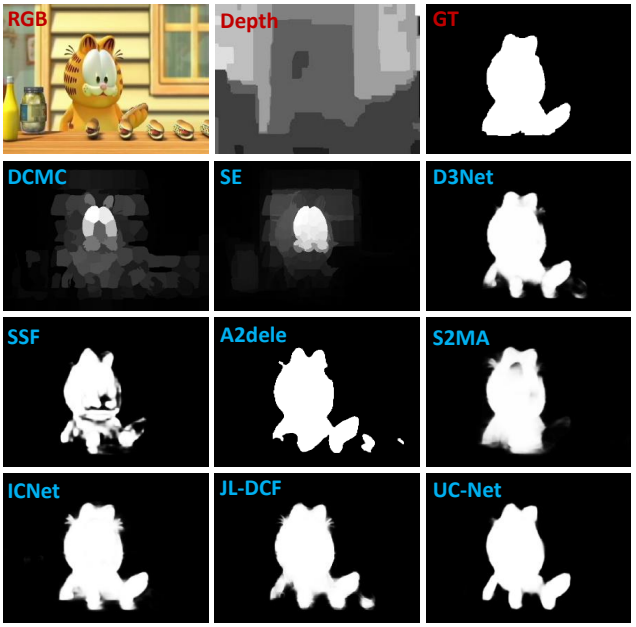


图. 1 显著性检测结果对比：两个经典的非深度模型（即，DCMC [26] 和 SE [48]）和七个最新的深度模型（即，D^3Net [38]、SSF [178]、A2dele [112]、S^2MA [88]、ICNet [76]、JL-DCF [46] 和 UC-Net [170]）。

1 前言

显著性物体检测（Salient object detection，SOD）旨在找出给定场景中视觉上最突显的对象 [32]。在一系列实际应用中 SOD 起着关键作用，如立体匹配 [101]、图像理解 [204]、协同显著性检测 [37]、动作识别 [116]、视频检测及分割 [39, 124, 147, 148]、语义分割 [121, 163]、医学图像分割 [36, 41, 154]、目标跟踪 [53, 97]、行人重识别 [98, 187]、伪装物体检测 [35]、图像检索 [87] 等。在过去几年中，虽然 SOD 领域的研究取得了重大的进展 [54, 78, 82, 125, 128, 143, 149, 155, 160, 161, 167, 180, 181, 186, 188]，但面对复杂背景、光照变化等挑战性因素，SOD 性能的提升仍有极大空间。而克服这些

1 起源人工智能研究院 (IIAI), 阿布扎比, 阿联酋。
2 南开大学计算机学院, 天津 300350, 中国。
通讯作者 (✉): 范登平 (邮箱: dengpfan@gmail.com)。
稿件接收日期: 2020-08-01; 稿件录用日期: 2020-10-08。

挑战的一种方法是采用深度图以弥补 RGB 图像缺失的空间信息，由于深度传感器 (如，Microsoft Kinect) 的广泛使用，所以也更容易获得深度图。

最近，RGB-D SOD 的研究受到越来越多的关注，涌现出了许多方法 [8, 38]。早期的 RGB-D SOD 模型倾向于提取手工特征，然后融合 RGB 图像和深度图像。例如，Lang 等人 [70]，首次提出基于 RGB-D SOD 方法，利用高斯混合模型对深度信息引导的显著性分布进行建模。Ciptadi 等人 [20] 从深度信息中提取了 3D 布局和形状特征。此外，[18, 28, 117] 采用不同区域之间的深度差来衡量深度图的对比度。在 [105] 中，作者提出一个包括局部、全局和背景对比的多上下文对比度模型，以使用深度图来检测显著性物体。更重要的是，该论文还为 SOD 研究提供了第一个大规模 RGB-D 数据集。尽管使用手工特征的传统方法可以完成 SOD 任务，但这些方法往往受限于底层特征的表达能力，且缺乏对复杂场景所需的高层推理能力。为了解决上述问题，许多基于深度学习的 RGB-D SOD 方法被提出 [38]，并提高了 SOD 性能。DF [115] 是首个基于深度学习的 RGB-D SOD 模型。最近，基于深度学习的模型 [12, 46, 76, 88, 107, 170, 185] 都致力于利用有效的多模态关联和多尺度/层级信息来提高 SOD 性能。为了更清晰地描述基于 RGB-D SOD 领域的发展进程，本文总结了一份简年表，如图 2所示。

本文旨在对 RGB-D SOD 模型进行全面系统的综述，并就未来工作的挑战和开放方向进行深入讨论。我们还将回顾另一个相关领域，即光场 SOD。在该领域中，光场可以提供更多信息（如焦点堆栈、全焦点图像和深度图）以提高显著性物体检测的性能。此外，本文还提供了一个全面的对比从而评估现有的基于 RGB-D 的 SOD 模型，并讨论这些方法的主要优势。

1.1 相关综述

目前有多个与显著性物体检测密切相关的综述，例如，Borji 等人 [3] 提供了 35 种最新的非深度学习显著性检测方法的定量评估。Cong 等人 [21] 回顾了几种不同的显著性检测模型，包括基于 RGB-D SOD 模型，协同显著性检测和视频 SOD。Zhang 等人 [165] 回顾了协同显著性检测及其发展进程，并总结了该领域中几个基准算法。Han 等人 [51] 回顾了 SOD 最近的研究进展，包括模型、基准数据集及评估指标等，并讨论了通用的物体检测、SOD 和特定类别物体检测之间的潜在联系。Nguyen 等人 [100] 回顾了与显著性应用相关的各种研究工作，并深入讨论了显著性在每个应用中的作用。Borji 等人 [2] 提供了 SOD 领域最新进展的综述，并讨论了一些相关工作，包括通用场景分割、眼动点预测和物体检测框生成等。Fan 等人 [32] 提供了几个最新的基于 CNNs 的 SOD 模型综合性能评估，并提出一个高质量 SOD 数据集 **SOC** (详情请参见 http://dpfan.net/socbenchmark/)。Zhao 等人 [190] 详细回顾了各种基于深度学习的目标检测模型和算法，以及各种特定任务，其中包括 SOD 相关研究工作。Wang 等人 [145] 主要回顾了基于深度学习的 SOD 模型。与上述关于 SOD 综述不同，本文专注于回顾已有的 RGB-D SOD 模型和基准数据集。

1.2 贡献

本文主要贡献总结如下:

- 本文首次从不同视角对 RGB-D SOD 模型进行了系统综述。将已有的 RGB-D SOD 总结为传统或深度模型、基于融合机制的模型、单数据流/多分支数据流模型以及基于注意力机制模型。
- 本文回顾了 RGB-D SOD 领域主流的 9 个 RGB-D 数据集，并提供了每个数据集的详细信息。此外，本文还对几个具有代表性的 RGB-D SOD 模型进行综合评估及基于属性的评测。
- 本文首次整理和回顾了有关光场 (light field) SOD 模型及基准数据集。
- 本文详尽地研究了 RGB-D SOD 面临的几个挑战，SOD 与其它相关研究的关系，并阐明未来研究的潜在方向。

1.3 章节安排

在第2章中，我们从不同视角对已有的 RGB-D SOD 模型进行概括及回顾。在第3章中，我们总结了当前主流的 RGB-D 视觉显著性检测基准数据集并提供它们的详细信息。在第4章中，我们全面地回顾了有关光场的 SOD 模型和基准数据集。在5章中，我们提供了几种具有代表性的 RGB-D SOD 模型的性能评估及属性分析。然后，在第6章中，我们讨论该领域面临的挑战和未来研究的潜在方向。最后，第7章总结了全文。

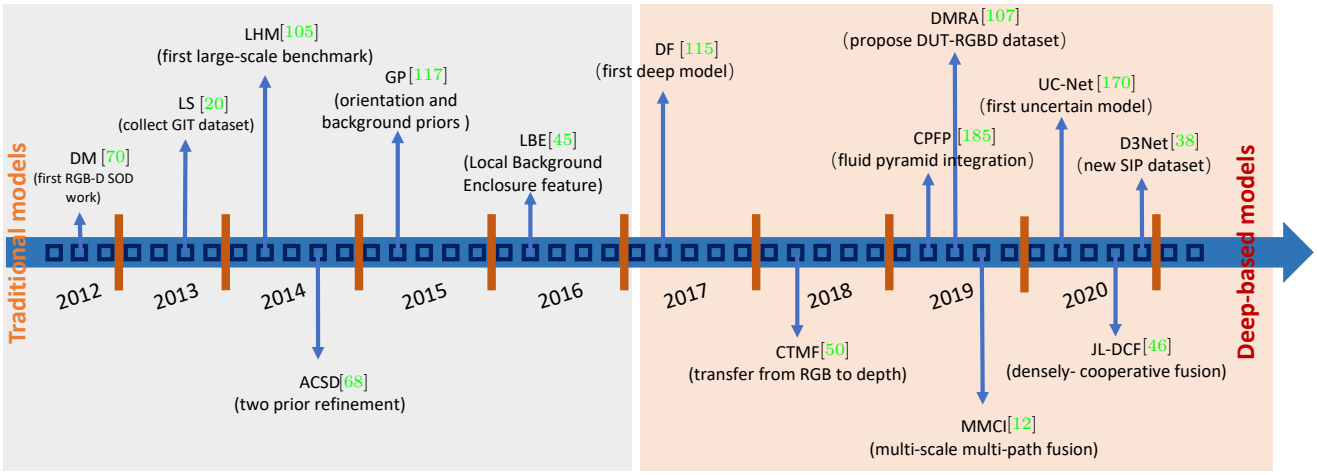


图. 2 RGB-D SOD 简年表。2012 年提出的 DM 模型 [70] 是第一个 RGB-D SOD 模型。2017 年之后，深度学习技术被广泛应用于 RGB-D SOD 研究中。更多详细内容还请参阅章节. 2.

2 RGB-D SOD 模型

在过去几年的研究中，已经涌现许多 RGB-D SOD 模型，并获得了可观的性能。这些模型总结参见表 1、2、3和 4。相关的基准数据和模型细节参见 http://dpfan.net/d3netbenchmark/。为了详细地回顾 RGB-D SOD 模型，我们从不同视角来介绍这些模型。(1) 传统/深度模型：从特征提取的视角来回顾它们，即使用手动特征或深度特征。这便于后续研究人员了解 RGB-D SOD 模型的历史发展趋势。(2) 基于融合机制的模型：在 RGB-D SOD 任务中有效地融合 RGB 和深度图是非常关键的，因此我们回顾不同的融合策略以了解其有效性。(3) 单数据流/多分支数据流模型：我们从模型参数的视角考虑这个问题。单数据流模型可以减少参数，但是最终结果可能不是最优的，而多分支数据流模型需要更多的参数。因此，这有助于理解模型的计算量和准确性之间的平衡。(4) 基于注意力机制的模型：注意力机制已广泛应用于包括 SOD 在内的各种视觉任务中。我们回顾了有关 RGB-D SOD 的相关工作，以分析这些模型如何使用注意力机制。因此，对于未来工作，设计注意力机制模块是一个可供选择的方案。

2.1 传统/深度模型

传统模型： 利用深度信息可以探索一些有效的属性，例如：边界信息、外观属性、曲面法线等，这些属性可以增强模型从复杂场景中检测显著性物体的能力。在过去的几年中，已经开发出许多基于手工特征的传统 RGB-D 模型 [10, 18, 20, 26, 28, 31, 42, 45, 48, 49, 68, 84, 105, 117, 120, 127, 202]。例如，[20] 是早期的 RGB-D SOD 模型，专注于 RGB 图像和深度图生成的布局和形状特征之间的交互进行建模。此外，一个具有代表性的研究 [105] 提出一种新颖的多级 RGB-D 模型，并构建了第一个大规模 RGB-D 基准数据集 NLPR。

深度模型： 然而，由于手工特征表达能力的限制，上述方法可能会获得不满意的 SOD 性能。为了解决该问题，一些研究则开始利用深度神经网络（deep neural networks，DNNs）融合 RGB-D 数据 [5, 9, 11, 12, 14, 22, 46, 56, 58, 72, 76, 83, 89, 91, 107, 109, 112, 115, 120, 139, 170, 178, 185, 198, 200]。这些深度模型可以学习更高层的特征表示，以挖掘 RGB 图像和深度信息之间的关联，从而提高 SOD 性能。本文将对以下一些具有代表性的研究工作进行简要回顾。

- **DF** [115] 提出一种新的卷积神经网络（convolutional neural network，CNN）将不同的低层显著性信息集成到分层特征中，从而有效地定位出 RGB-D 图像中的显著性区域。该项研究是首个把 CNN 应用于 RGB-D SOD 任务中的模型，但它仅利用了浅层架构来学习显著性图。

表. 1 基于 RGB-D 的 SOD 方法总结（发表于 2012 年至 2016 年）

序号	年份	模型	出版物	训练集	骨干网络	描述
1	2012	DM [70]	ECCV	无	无	利用高斯混合模型近似联合密度分布来建模显著性和深度图的相关性
2	2012	RCM [169]	ICCSE	无	无	提出基于信息区域对比度的 SOD 模型，并联合深度图
3	2013	LS [20]	BMVC	无	无	扩展差异性框架以对深度图信息和 RGB 图像之间的联合交互进行建模
4	2013	RC [28]	BMVC	无	无	通过基于场景区域对比度来建立 3D 显著性模型，并使用 SVM 将其融合
5	2013	SOS [71]	NEURO	无	无	通过抑制背景区域，结合深度图信息实现显著性物体分割
6	2014	SRDS [42]	ICDSP	无	无	采用颜色分布的空间紧凑性来整合深度图信息和深度色彩颜色对比权重
7	2014	LHM [105]	ECCV	无	无	使用多阶段 RGB-D 算法结合深度信息和外观信息来实现显著性物体分割
8	2014	DESM [18]	ICIMCS	无	无	结合 3 种显著性线索，即颜色对比度，空间偏差和深度对比度
9	2014	ACSD [68]	ICIP	无	无	根据它在周围的突出程度来衡量一个点的显著性，并利用两个先验信息（距观察者距离近的区域更显著，以及显著性物体往往位于中心）
10	2015	GP [117]	CVPRW	无	无	探索方向和背景的先验信息以实现显著性物体检测，并采用 PageRank 和 MRFs 方法优化显著性图像
11	2015	SFP [49]	ICIMCS	无	无	提出一种采用显著性融合和传播的 RGB-D SOD 方法
12	2015	DIC [127]	TVC	无	无	从颜色和深度信息中融合显著性图像以实现一个无噪声的显著性图像块，并利用随机游走方法来推断物体的边界
13	2015	SRD [65]	ICRA	无	无	设计基于图的分割方法，利用颜色和深度线索来判别同质区域。
14	2015	MGMR [158]	ICIP	无	无	设计相互引导的流形排序策略来实现显著性物体检测
15	2015	SF [205]	CAC	无	无	提出一种利用决策树实现自动选择判别性特征来优化显著性检测的性能
16	2016	PRC [31]	ACCESS	无	无	采用显著性融合和渐进区域分类实现优化深度自适应的显著性模型
17	2016	LBE [45]	CVPR	无	无	利用一个局部背景组件来获取角度方向的传播信息
18	2016	SE [48]	ICME	无	无	利用细胞自动机传播初始的显著性图，然后再预测并生成最终显著性结果
19	2016	DCMC [26]	SPL	无	无	提出一种新的评估深度图可靠性的机制用于减少低质量深度图对显著性目标检测结果的影响
20	2016	BF [141]	ICPR	无	无	利用贝叶斯框架融合 RGB 图像和深度图之间的对比特征
21	2016	DCI [118]	ICASSP	无	无	利用原始深度图减去拟合曲面从而生成对比增强图像
22	2016	DSF [122]	ICASSP	无	无	提出一个多阶段深度图感知的显著性物体检测模型
23	2016	GM [142]	ACCV	无	无	利用生成混合模型结合颜色和基于深度的对比特征

• **PCF** [7] 提出一种互补感知融合模块，用于集成跨模态和跨层特征表达，通过明确地使用交叉模式/层级连接和模式/层级监督来有效地利用补充信息，以降低融合的缺陷。

• **CTMF** [50] 使用一个计算模型从 RGB-D 场景中识别出显著性物体，利用 CNNs 学习 RGB 图像和深度信息的高层级表示，同时利用互补关系和联合表达。此外，该模型从源数据（RGB 图像）转换模型结构以适用于目标数据（深度图）。

• **CPFP** [185] 提出一个对比增强网络生成强化图，并提出一个流体金字塔集成模块，以分层方式有效融合跨模态信息。此外，考虑到深度图易受噪声干扰，利用一个特征增强模块来学习增强的深度信息以提高 SOD 性能。值得注意的是这是一个很有效的解决方案。

• **UC-Net** [170] 提出了一种基于概率的 RGB-D SOD 网络。利用条件变量自动解码器来模拟人类标注的不确定性，并在学习的隐空间中采样为每个输入图像生成多个显著图。它是 RGB-D SOD 任务中第一个受到数据标记过程的启发并研究不确定性的工作。UC-Net 利用多样化的显著性图来提升最终的 SOD 性能。

2.2 基于融合机制的模型

对于 RGB-D SOD 模型, 重要的是有效地融合 RGB 图像和深度图。存在的融合策略分为以下三种：1）早期融合；2）多尺度融合；3）后期融合。下面对每种融合策略进行详细的说明。

早期融合： 早期融合可遵循以下思路的其中之一：1）将 RGB 图像和深度图直接集成以形成四通道输入 [89, 105, 117, 123]，称为“输入融合”，如图 3(a) 所示；2）将 RGB 图像和深度图输入到两个独立的网络，并将它们的底层特征组合成联合表示，然后将其输入到后续网络并进一步实现显著性图的预测 [115]，称为“早期特征融合”，如图 3(a) 所示。

后期融合： 后期融合进一步分为两个子类：1）采用两个并行网络分别学习 RGB 和深度图的高层特征，然后将它们级联起来，用于生成最终的显著性预测结果 [28, 50, 139]，这被称为“后期特征融合”，如图 3(b) 所示。2）采用两个并行网络分别获取 RGB 图像和深度图的独立显著性图，然后将两个显著性图级联一起用于获得最终的预测图 [29]，称为“后期结果融合”，如图 3(b) 所示。

表. 2 基于 RGB-D 的 SOD 方法总结 (发表于 2017 年至 2018 年)

序号	年份	模型	出版物	训练集	骨干网络	描述
24	2017	HOSO [44]	DICTA	无	无	将表面方向分布的对比度与颜色和深度对比度相结合
25	2017	M^3Net [13]	IROS	NLPR(0.65K), NJUD(1.4K)	VGG-16	设计一个多通路多模态融合策略，并以任务激励和自适应的方法融合 RGB 图像和深度图。
26	2017	MFLN [10]	ICCVS	NLPR(0.65K), NJUD(1.4K)	AlexNet	借助于 CNN 来学习深度图的高层特征表示，并利用多模态融合网络集成 RGB 图像和深度信息以实现 RGB-D 显著性物体检测
27	2017	BED [120]	ICCVW	NLPR(0.6K), NJUD(1.2K)	GoogleNet	提出一种 RGB-D SOD 方法，利用 CNN 融合自上而下和自下而上的信息，并采用中间层特征来获取背景信息
28	2017	CDCP [202]	ICCVW	无	无	提出一种新的 RGB-D SOD 算法，利用中心暗通道先验来提高 SOD 性能
29	2017	TPF [201]	ICCVW	无	无	借助于立体视觉产生光流，它能提供额外的信息（即深度信息）来生成最后的检测结果
30	2017	MFF [134]	SPL	无	无	采用一种多阶段融合架构来集成 RGB 图像和深度图的多重视觉先验信息从而实现 SOD
31	2017	MDSF [123]	TIP	NLPR(0.5K), NJUD(1.5K)	无	提出一种 RGB-D SOD 架构，通过多尺度判别显著性融合策略，并利用自举学习实现 SOD 任务
32	2017	DF [115]	TIP	NLPR(0.75K), NJUD(1.0K)	无	将 RGB 图像和深度图特征输入到 CNN 架构中，以得到显著性置信度值，并使用拉普拉斯传播产生最终的检测结果
33	2017	MCLP [25]	TCYB	无	无	利用已有的 RGB 显著性图作为初始化，联合深度图并使用一个循环细化模型获得协同显著性检测图
34	2018	ISC [62]	SIVP	无	无	利用自底向上和自顶向下的显著性线索来融合显著性特征
35	2018	HSCS [24]	TMM	无	无	利用分层稀疏性重构和能量函数细化来完成 RGB-D 协同显著性检测
36	2018	ICS [23]	TIP	无	无	利用多个图像之间的约束相关性将深度图引入到协同显著性模型中
37	2018	CTMF [50]	TCYB	NLPR(0.65K), NJUD(1.4K)	VGG-16	将深度学习颜色网络结构迁移到深度图，并融合两种模态进而生成最终的显著性图
38	2018	PCF [7]	CVPR	NLPR(0.65K), NJUD(1.4K)	VGG-16	设计第一个多尺度融合结构，并提出一个新的具有互补性的融合模块，实现了跨模态和跨特征层的融合
39	2018	SCDL [56]	ICDSP	NLPR(0.75K), NJUD(1.0K)	VGG-16	设计一种新的损失函数来增加显著性物体间的空间连贯性
40	2018	ACCF [14]	IROS	NLPR(0.65K), NJUD(1.4K)	VGGNet	从每个特征层的不同模态中自适应选择补充的特征，然后执行更多信息性的跨模态跨特征层的结合
41	2018	CDB [84]	NEURO	无	无	利用对比度先验和深度引导背景先验构建一个 3D 立体显著性模型

表. 3 基于 RGB-D 的 SOD 方法总结 (发表于 2019 年至 2020 年)

序号	年份	模型	出版物	训练集	骨干网络	描述
42	2019	SSRC [89]	NEURO	NLPR(0.65K), NJUD(1.4K)	VGG-16	利用具有四通道输入和 DRCNN 子网的单流递归卷积神经网络
43	2019	MLF [57]	SPL	NJUD(1.588K)	VGG-16	设计一个显著性物体自适应的数据增强方法来扩展训练数据集
44	2019	TSRN [85]	ICIP	NJUD(1.387K)	VGG-16	设计一个融合细化模块来集成不同模态和分辨率下的输出特征
45	2019	DIL [30]	MTAP	NLPR(0.5K), NJUD(0.5K)	无	设计一致性融合策略来生成与深度分布一致的图像预分割结果
46	2019	CAFM [197]	TSMC	NUS [70], NCTU [95]	VGG-16	利用内容感知融合模型来集成全局和局部信息
47	2019	PDNet [200]	ICME	NLPR(0.5K), NJUD(1.5K)	VGG-16	采用一个先验模型引导的主网络处理 RGB 信息，该模型在常规 RGB 数据集上进行了预训练，以克服小训练集的限制
48	2019	MMCI [12]	PR	NLPR(0.65K), NJUD(1.4K)	VGG-16	通过多样化多模态融合路径并在多层中引入跨模态交互来优化传统的两分支结构
49	2019	TANet [9]	TIP	NLPR(0.65K), NJUD(1.4K)	VGG-16	采用一个三分支数据流的多模态融合框架，从自底向上和自顶向下的处理中探索跨模态的互补性
50	2019	DCMF [11]	TCYB	NLPR(0.65K), NJUD(1.4K)	VGG-16	形成一个基于 CNN 的跨模态迁移学习以实现深度信息引导的 SOD，并使用密集跨层级反馈策略来挖掘跨特征层间的交互
51	2019	DGT [22]	TCYB	无	无	利用深度线索并提供一个生成迁移模型将 RGB 显著性应用到 RGB-D 显著性任务
52	2019	LSF [8]	arXiv	NLPR(0.65K), NJUD(1.4K)	VGG	设计一个 RGB-D 系统包括 3 个关键组件，即模态特定的表示学习、互补信息选择及跨模态信息融合
53	2019	AFNet [139]	ACCESS	NLPR(0.65K), NJUD(1.4K)	VGG-16	学习一个开关图，用于从 RGB 图像和深度图中自适应地融合显著性预测图
54	2019	EPM [67]	ACCESS	无	无	提供一个有效的传播机制以实现 RGB-D 协同显著性检测任务
55	2019	CPFP [185]	CVPR	NLPR(0.65K), NJUD(1.4K)	VGG-16	利用一个对比增强网络来获得单通道增强图，并设计一个流式金字塔集成模型来融合跨模态和跨层级特征
56	2019	DMRA [107]	ICCV	NLPR(0.7K), NJUD(1.485K)	VGG-19	设计一个深度信息引导的多尺度循环注意力网络来实现显著性物体检测，包括一个深度细化模块和一个循环注意力模块
57	2019	DSD [29]	JVCIR	NLPR(0.5K), NJUD(1.5K)	VGG-16	利用显著性融合网络以自适应地融合颜色和深度显著性图
58	2020	DPANet [17]	arXiv	NLPR(0.65K), NJUD(1.4K), DUT(0.8K)	ResNet-50	利用显著性引导的深度感知模型来评估深度图的潜力并减低有瑕疵深度图的影响
59	2020	SSDP [153]	arXiv	NLPR(0.7K), NJUD(1.485K), DUT(0.8K)	VGG-19	充分利用现有带标签的 RGB 显著性数据集和无标签的 RGB-D 数据来提升 SOD 性能
60	2020	AttNet [198]	IVC	NLPR(0.65K), NJUD(1.4K)	VGG-16	采用注意力图以增强显著性物体的定位，对于外观信息分配更多的关注
61	2020	GFNet [91]	NEURO	NLPR(0.65K), NJUD(1.4K)	VGG-16	利用一个基于 GAN 的自适应门控融合模块，以获得更好的 RGB 图像和深度图的显著性图
62	2020	CoCNN [83]	PR	STERE, NJUD	VGG-16	在统一深度模型中融合颜色和从低到高层之间的差异特征
63	2020	cmSalGAN [64]	TMM	NLPR(0.65K), NJUD(1.4K)	ResNet-50	旨在使用对抗性学习框架为 RGB 图像和深度图像学习最优的视图不变和一致性的像素级表示
64	2020	PGHF [156]	ACCESS	NLPR(0.65K), NJUD(1.4K)	VGG-16	利用从大规模 RGB 数据集中学到的强大表示来增强建模能力

多尺度融合： 为了有效地探索 RGB 图像和深度图之间的相关性，许多方法 [12, 17, 40, 46, 76, 77, 104, 183] 提出了一种多尺度融合策略。这些模型可以分为两类，第一类是学习跨模式交互，然后将它们融合到特征学习

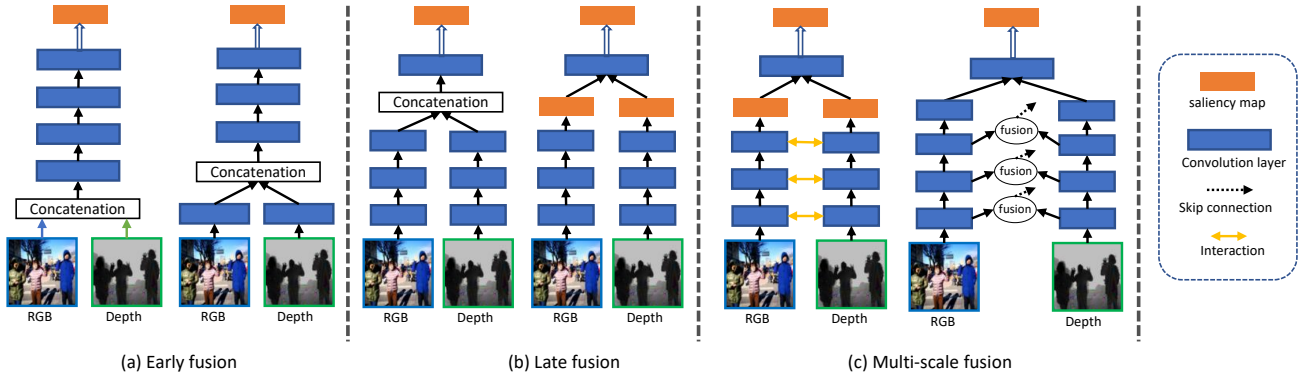


图. 3 三种融合策略对比，它们探索 RGB 图像和深度图之间的相关性。融合策略包括：a）早期融合、b）后期融合、c）多尺度融合。

网络中。例如，Chen 等人 [12] 提出一种多尺度多路径融合网络以整合 RGB 图像和深度图，以及一个交叉模式交互（称为 MMCI）模块。该方法将交叉模式交互引入到多层中，并赋予每个附加梯度以增强对深度流的学习，用于探索低层和高层表示之间的互补性。第二类是融合不同层中映射的 RGB 图像和深度图特征，然后将它们集成到解码器网络（如 skip connection）以生成最终的显著性检测图，如图 3（c）所示。下面对一些具有代表性的研究工作进行简要的讨论。

• **ICNet** [76] 提出了信息转换模块，以交互方式转换高层特征。在此模型中，引入了交叉深度加权组合（cross-modal depth-weighted combination，CDC）模块，以利用不同层级的深度图特征来增强 RGB 特征。

• **DPANet** [17] 利用门控多模态注意力（gated multi-modality attention，GMA）模块来获取长相关性。GMA 模块可以利用空间注意力机制来提取最具有判别力的特征。此外，该模型使用门函数控制交叉模式信息的融合率，以减少不可靠的深度图所带来的的影响。

• **BiANet** [183] 使用多尺度的双边注意模块（multi-scale bilateral attention module，MBAM）来获取多层更好的全局信息。

• **JL-DCF** [46] 将深度图视为彩色图像的特例，并使用共享的 CNN 进行 RGB 和深度特征的提取。该方法还提出了一种与密集协作融合策略，有效地结合从不同模态中学习到的特征。

• **BBS-Net** [40] 使二分支主干策略（bifurcated backbone strategy，BBS）将多尺度特征分为教师特征和学生特征，并开发了深度增强模块（depth-enhanced module，DEM），从空间和通道的角度提取有用的深度特征信息。

2.3 单数据流/多分支数据流模型

单数据流模型：许多 RGB-D SOD 研究 [56, 89, 115, 120, 123, 185, 200] 专注于单数据流结构来完成显著性预测。这些模型通常在输入通道或特征学习部分融合 RGB 图像和深度图。例如，MDSF [123] 利用多尺度判别显著性融合架构作为 SOD 模型，其中计算了 3 个层级的四种类型特征，然后将其融合以获得最终的显著性图。BED [120] 采用 CNN 架构为 SOD 集成了自下而上和自上而下的信息，还结合了多种特征，包括背景 enclosure 分布（background enclosure distribution，BED）和低层深度图特征（如，深度直方图距离和深度对比度）来提升 SOD 性能。PDNet [200] 使用辅助网络提取基于深度图的特征，该辅助网络充分利用了深度信息来辅助主干网络。

多分支数据流模型：双分支数据流模型 [107, 139, 198] 由两个独立分支组成，分别用于处理 RGB 图像和深度图，通常用来生成不同的高级特征或显著性图，然后在两个分支数据流的中间阶段或末尾将它们合并。值得注意的是，最新的基于深度学习的模型 [7, 8, 11, 12, 17, 64, 72, 76, 91, 112] 利用双分支数据流结构来获取 RGB 图像和深度信息之间的相关性。此外，一些模型利用多分支数据流结构 [9, 38]，并设计不同的融合模块有效地融合 RGB 和深度信息，它们能充分利用两者之间的的关联性。

2.4 基于注意力机制的模型

现有的 RGB-D SOD 方法通常使用提取的特征均等地对待所有的区域，而忽略了不同区域对最终预测结果可能有不同贡献的事实。这些方法容易受到背景杂乱的影响。此外，有些方法将 RGB 图像和深度图视为具有相同的状态，有的则过度依赖深度信息，导致这些方法不能考虑不同域（RGB 图像或深度图）的重要性。为了克服该问题，许多方法引入注意力机制来加权不同区域的重要性。

• **ASIF-Net** [72] 使用交叉融合来利用 RGB 图像和深度图的补充信息，并通过深度监督注意力机制对显著性区域进行加权。

• **AttNet** [198] 引入了注意力图以区分显著性物体和背景区域，以减少低质量深度图的负面影响。

• **TANet** [9] 提出一种多模态融合框架，从自下而上和自上而下的角度来集成 RGB 图像和深度图。然后引入一个通道注意模块来有效地融合来自不用模态和层级的补充信息。

2.5 开源项目

我们总结了本综述所回顾的基于 RGB-D SOD 模型的开源实现项目，这些模型的源代码链接见表 5。更多的源代码将在 https://github.com/taozh2017/RGBD-SODsurvey网页上更新。

3 RGB-D 数据集

随着 RGB-D SOD 研究的快速发展，在过去几年中构建了多个数据集。表 6总结了九个主流的 RGB-D 数据集，以及图 4展示了每个数据集中的图像示例（包括 RGB 图像、深度图和标注的真值图）。接下来我们对每个数据集进行详细地介绍。

• **STERE** [102] 作者利用 Flickr [1]，NVIDIA 3D Vision Live [2]，和 Stereoscopic Image Gallery [3]收集了 1250 幅立体图像。每幅图像中最显著的物体均由 3 名用户进行标注，然后根据重叠的显著性区域对所有带标注的图像进行排序，选择前 1000 幅图像来构建最终的数据集。这是该领域首个立体图像数据集。

[1]http://www.flickr.com/
[2]http://photos.3dvisionlive.com/
[3]http://www.stereophotography.com/

• **GIT** [20] 由 80 幅彩色图像和深度图组成，这些图像是在真实家庭场景中利用移动机器人采集得到的。此外，每幅图像都是基于物体的像素级分割标注。

• **DES** [18] 由 135 张室内 RGB-D 图像组成，通过 Kinect 拍摄的分辨率为 640×640。在标记该数据集时要求三个用户在每张图像中标记出显著性物体，然后标记图像的重叠区域作为显著性物体的真值。

• **NLPR** [105] 包含 1000 幅 RGB 图像和对应的深度图，这些图像都是通过标准的 Microsoft Kinect 收集得到。该数据集包括一系列室外和室内场景，如办公室、超市、校园、街道等。

• **LFSD** [80] 由 Lytro 光场相机收集的 100 张光场图像组成，包含 60 张室内场景和 40 张室外场景。为了标记该数据集，要求三个用户手动地对显著性物体进行分割，当三个结果的重叠率超过 90% 时，将分割结果作为显著性目标的真值。

• **NJUD** [68] 由 1985 个立体图像对组成，并且这些图像均来自于 Interne, 3D 电影和 Fuji W3 立体相机拍摄的照片。

• **SSD** [201] 通过 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为 960×1080。

• **DUT-RGBD** [109] 由 800 个室内场景和 400 个室外场景及相应的深度图组成。该数据集包含了几个具有挑战性的因素，即多或透明的物体、复杂的背景、相似的前景和背景以及低分辨率环境。

• **SIP** [38] 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。

详细的数据统计分析（包括中心偏向、物体大小、背景物体、物体边界条件、以及显著物体的数量）请参见 [38]。

表. 4 基于 RGB-D 的 SOD 方法总结（发表于 2020 年）

序号	年份	模型	出版物	训练集	骨干网络	描述
65	2020	BiANet [183]	TIP	NLPR(0.7K), NJUD(1.485K)	VGG-16	采用一个双边注意模型（BAM）来探索深度图中丰富的前景和背景信息
66	2020	ASIF-Net [72]	TCYB	NLPR(0.65K), NJUD(1.4K)	VGG-16	融合 RGB-D 图像中注意力引导的互补信息，并使用对抗学习来引入全局语义约束
67	2020	Triple-Net [58]	SPL	Triple-Net	ResNe-18	使用三重互补网络实现 RGB-D SOD
68	2020	ICNet [76]	TIP	Triple-Net	VGG-16	利用一个新的信息转换模块以交互和自适应方式来融合高层级 RGB 和深度图特征
69	2020	SDF [5]	TIP	NLPR,NJUD, DEC,LFSD(1.5K)	VGG-16	提出一种示例驱动的方法来估计相对可靠的深度图，并使用一个选择深度显著性融合网络有效地结合 RGB 图像、原始深度图和新估计的深度图
70	2020	GFNet [196]	SPL	NLPR(0.8K), NJUD(1.588K)	Res2Net	设计一个门融合模块来正则化特征融合
71	2020	RGBS [90]	MTAP	NLPR(0.65K), NJUD(1.4K)	VGG-16	利用 GAN 网络来生成显著性图
72	2020	D^3Net [38]	TNNLS	NLPR(0.7K), NJUD(1.485K)	VGG-16	利用一个深度图过滤单元和一个三分支数据流的特征学习模块分别过滤低质量深度图和完成跨模态特征学习
73	2020	JL-DCF [46]	CVPR	NLPR(0.7K), NJUD(1.5K)	VGG-16, ResNet-101	使用一个联合学习策略和一个和密集协作融合模块获得了更好的 SOD 性能
74	2020	A2dele [112]	CVPR	NLPR(0.7K), NJUD(1.485K)	VGG-16	使用深度蒸馏器探索将网络预测和注意力作为深度知识转移到 RGB 图像的方法
75	2020	SSF [178]	CVPR	NLPR(0.7K), NJUD(1.485K), DUT(0.8K)	AGG-16	设计一个互补交互模块从 RGB 和深度图中选择有用的特征表示，然后融合跨模特征
76	2020	S^2MA [88]	CVPR	NLPR(0.65K), NJUD(1.4K)	VGG-16	利用自注意机制和交互注意力策略来融合多模态信息，并重新分配注意力权重以过滤掉不可靠的信息
77	2020	UC-Net [170]	CVPR	NLPR(0.7K), NJUD(1.5K)	VGG-16	使用基于变分自编码器（VAE）的概率 RGB-D 显著性检测网络来生成多个显著性图
78	2020	CMWNet [77]	ECCV	NLPR(0.65K), NJUD(1.4K)	VGG-16	使用三个跨模态、跨尺度和权重模块完成特征交互，从而改进 SOD 性能
79	2020	HDFNet [104]	ECCV	NLPR(0.7K), NJUD(1.485K), DUT(0.8K)	VGG-16	设计一个分层动态过滤网络来有效地利用跨模态的融合信息
80	2020	CAS-GNN [94]	ECCV	NLPR(0.65K), NJUD(1.4K)	VGG-16	设计级联图神经网络以利用 RGB 和深度图中的有用知识来构建强大的特征嵌入
81	2020	CMMS [73]	ECCV	NLPR(0.7K), NJUD(1.485K)	VGG-16	提供一个跨模态特征调制模型以此增强特征表示，并通过自适应的特征筛选模块来逐层筛选显著性相关的特征
82	2020	DANet [189]	ECCV	NLPR(0.65K), NJUD(1.4K)	VGG-16, VGG-19	结合深度增强双注意力机制提出一个单数据流网络来实现实时的显著性检测
83	2020	CoNet [63]	ECCV	NLPR(0.7K), NJUD(1.485K), DUT(0.8K)	ResNet	开发一个联合学习框架实现 RGB-D SOD，利用三个组件（即，边缘检测、粗略显著性物体检测和深度估计）联合提升 SOD 性能
84	2020	BBS-Net [40]	ECCV	NLPR(0.65K), NJUD(1.4K)	VGG-16, VGG-19, ResNet-50	利用二分支主干策略学习教师特征和学生特征，并采用一个深度增强模块来提取有用的深度特征信息
85	2020	ATSA [175]	ECCV	NLPR(0.7K), NJUD(1.485K), DUT(0.8K)	VGG-19	提出一个不对称的两分支数据流结构，充分考虑 RGB 图像和深度图之间的固有差异来实现显著性物体检测
86	2020	PGAR [15]	ECCV	NLPR(0.7K), NJUD(1.485K)	VGG-16	提出一个逐步引导的交替细化网络，利用多尺度残差模块来生成粗略的初始化预测
87	2020	MCINet [61]	arXiv	NLPR(0.65K), NJUD(1.4K)	ResNet-50	提出一个新的多层跨模态交互网络来实现 RGB-D SOD 任务
88	2020	DRLF [151]	TIP	NLPR(0.65K), NJUD(1.4K)	VGG-16	开发基于通道的融合网络，以进行多网络和多级选择性融合来完成 RGB-D SOD
89	2020	DQAM [152]	arXiv	NLPR(0.65K), NJUD(1.4K)	无	针对 RGB-D 图像提出一种深度图质量评估方案来构建“质量自适应的”显著性物体检测方法
90	2020	DQSD [4]	TIP	NLPR(0.65K), NJUD(1.4K)	VGG-19	将具有深度质量感知的子网络集成到双分支数据流结构中，以在进行 RGB-D 融合之前评估深度图的质量
91	2020	DASNet [184]	ACM MM	NLPR(0.7K), NJUD(1.5K)	ResNet-50	提出一种新的视角，即在学习过程中包含深度约束，而不是将深度图像直接用作输入
94	2020	DCMF [6]	TIP	NLPR(0.65K), NJUD(1.4K)	VGG-16, ResNet-50	设计一个分离式的跨模态融合网络，从 RGB 和深度中获取结构和内容表示

4 光场显著性检测

4.1 光场 SOD 模型

SOD 研究根据其输入数据的类型可分为三类，包括 RGB SOD、RGB-D SOD 和光场 SOD [176]。我们已经回顾了 RGB-D SOD 模型，其深度图提供了布局信息并能在一定程度上提升 SOD 性能。但不准确或者低质量的深度图通常会降低其性能。为了克服这个问题，许多基于光场的 SOD 方法已经提出，它们能充分地利用丰富的光场信息。具体来说，光场数据包含一个全焦点图像、一个焦点堆栈和一个简略深度图 [109]。如表 7所示，我们总结了有关光场 SOD 相关工作的进展。此外，为了深入地理解这些模型，我们对它们进行如下详细的介绍。

传统/深度模型： 经典光场 SOD 模型通常使用基于超像素的手工特征 [79–81, 109, 119, 135, 140, 150, 173, 174]。早期的研究工作 [80, 81] 证明了光场独特的重聚焦能力可以提供有用的聚焦，深度和似物性线索。随后提出了许多采用光场图像的 SOD 模型。例如，Zhang 等人 [173] 利用一组焦点切片图来计算背景先验，然后将其与位置先验结合来完成 SOD。Wang 等人 [135] 提出一个两阶段贝叶斯融合模型，为了集成多种对比度以提高 SOD 性能。最近，已经开发了许多基于深度学

图. 4 九个 RGB-D 数据集中的 RGB 图像，深度图，和显著性物体的真值实例，包括（a）STERE [102]，（b）NLPR [105]、（c）SSD [201]、（c）GIT [20]、（e）DES [18]、（f）LFSD [80]、（g）NJUD [68]、（h）DUT-RGBD [109] 和（i）SIP [38]。从左至右每个数据集依次显示为 RGB 图像，深度图和显著性真值图。

习的光场 SOD 模型 [110, 111, 144, 172, 176, 177]，取得了卓越的性能。此外，在 [144] 中，作者提出一种注意力循环 CNN 网络以融合所有焦点切片信息，同时利用对抗性示例来增加数据多样性以增强模型的鲁棒性。Zhang 等人 [177] 提出一种面向存储器的光场 SOD 解码器，该解码器采用高层信息以自上而下的方式融合更高层的信息来指导低层特征的选择。LFNet [176] 采用一个新的融合模块利用光场数据的贡献来融合特征，获得场景的空间结构以提高 SOD 性能。

细化模型：细化策略已经用来增强相邻约束或降低多模态 SOD 的同质性。例如，在 [79] 中，利用估计的显著性图对显著性字典进行了细化。MA 方法 [174] 采用了两阶段显著性细化策略来生成最终的预测图，该预测图可以使用相邻的超像素获得更相似的显著性值。此外，LFNet [176] 提供了一个有效的细化模块以降低不同模态之间的同质性以细化它们之间的差异。

4.2 光场 SOD 数据集

五个代表性的数据集已经广泛用于光场 SOD 任务。我们对每个数据集进行简单的介绍。

• **LFSD** [80] [4] 通过 Lytro 光场相机采集获取，其中包含 100 幅 360×360 分辨率不同场景的光场图像。该数据集包括 60 个室内场景和 40 个室外场景，并且大部分场景只包含一个显著性物体。此外，请求三个用户手动地对显著性物体进行分割，当三个结果的重叠率超

[4] https://sites.duke.edu/nianyi/publication/saliency-detection-on-light-field/

表. 5 RGB-D SOD 模型的开源实现总结

年份	模型	实现	源代码链接
2014	LHM [105]	Matlab	https://sites.google.com/site/rgbdsaliency/code
	DESM [18]	Matlab	https://github.com/HzFu/DES_code
2015	GP [117]	Matlab	https://github.com/JianqiangRen/Global_Priors_RGBD_Saliency_Detection
2016	DCMC [26]	Matlab	https://github.com/rmcong/Code-for-DCMC-method
	LBE [45]	Matlab & C++	http://users.cecs.anu.edu.au/ u4673113/lbe.html
2017	BED [120]	Caffe	https://github.com/sshige/rgbd-saliency
	CDCP [202]	Matlab	https://github.com/ChunbiaoZhu/ACVR2017
	MDSF [123]	Matlab	https://github.com/ivpshu
	DF [115]	Matlab	https://pan.baidu.com/s/1Y-PqAjuH9xREBjfl7H45HA
2018	CTMF [50]	Caffe	https://github.com/haochen593/CTMF
	PCF [7]	Caffe	https://github.com/haochen593/PCA-Fuse_RGBD_CVPR18
	PDNet [200]	TensorFlow	https://github.com/cai199626/PDNet
2019	AFNet [139]	TensorFlow	https://github.com/Lucia-Ningning/Adaptive_Fusion_RGBD_Saliency_Detection
	CPFP [185]	Caffe	https://github.com/JXingZhao/ContrastPrior
	DMRA [107]	PyTorch	https://github.com/jiwei0921/DMRA
	DGT [22]	Matlab	https://github.com/rmcong/Code-for-DTM-Method
2020	ICNet [76]	Caffe	https://github.com/MathLee/ICNet-for-RGBD-SOD
	JL-DCF [46]	Pytorch, Caffe	https://github.com/kerenfu/JLDCF
	A2dele [112]	PyTorch	https://github.com/OIPLab-DUT/CVPR2020-A2dele
	SSF [178]	PyTorch	https://github.com/OIPLab-DUT/CVPR_SSF-RGBD
	ASIF-Net [72]	TensorFlow	https://github.com/Li-Chongyi/ASIF-Net
	S^2MA [88]	PyTorch	https://github.com/nnizhang/S2MA
	UC-Net [170]	PyTorch	https://github.com/JingZhang617/UCNet
	D^3Net [38]	PyTorch	https://github.com/DengPingFan/D3NetBenchmark
	CMWNet [77]	Caffe	https://github.com/MathLee/CMWNet
	HDFNet [104]	PyTorch	https://github.com/lartpang/HDFNet
	CMMS [73]	TensorFlow	https://github.com/Li-Chongyi/cmMS-ECCV20
	CAS-GNN [94]	PyTorch	https://github.com/LA30/Cas-Gnn
	DANet [189]	PyTorch	https://github.com/Xiaoqi-Zhao-DLUT/DANet-RGBD-Saliency
	CoNet [63]	PyTorch	https://github.com/jiwei0921/CoNet
	DASNet [184]	PyTorch	http://cvteam.net/projects/2020/DASNet/
	BBS-Net [40]	PyTorch	https://github.com/DengPingFan/BBS-Net
	ATSA [175]	PyTorch	https://github.com/sxfduter/ATSA
	PGAR [15]	PyTorch	https://github.com/ShuhanChen/PGAR_ECCV20
	FRDT [179]	PyTorch	https://github.com/jack-admiral/ACM-MM-FRDT

过 90% 时才标注该分割结果为显著性物体的真值。

• **HFUT** [174] [5] 由 Lytro 光场相机获取，包含 255 张光场图像。该数据集大部分场景都包含多个显著性物体，在复杂和杂乱背景下，这些物体出现在不同位置和具有不同的大小比例。

• **DUTLF-FS** [144] [6] 包含 1465 个样本，其中 1000 个样本为训练样本，余下的 465 个为测试样本。该数据集图像分辨率均为 600×400，并且包含如下挑战，即显著性物体与背景之间对比度较低，多个不相邻的显著性物体，以及黑暗或强光照。

• **DUTLF-MV** [110] [7]由 Lytro Illum 相机采集，包含 1580 个样本，其中 1100 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。

• **Lytro Illum** [172] [8] 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行的因素，如不一致的光照状况，背景相似或背景杂乱中的小显著性物体。

5 模型评估及分析

5.1 评价指标

我们简要回顾几个主流的 SOD 评价指标，即 PR 曲线、F 指标（F-measure）[1, 3]、平均绝对误差（mean absolute error，MAE）[106]、S 指标（S-measure）[33] 和 E 指标（E-measure） [34]。

• **PR 曲线**。给定一个显著性图 S，我们将其转换为

[5] https://github.com/pencilzhang/HFUT-Lytro-dataset
[6] https://github.com/OIPLab-DUT/ICCV2019_Deeplightfield_Saliency
[7] https://github.com/OIPLab-DUT/IJCAI2019-Deep-Light-Field-Driven-Saliency-Detection-from-A-Single-View
[8] https://github.com/pencilzhang/MAC-light-field-saliency-net

表. 6 九个 RGB-D 基准数据集信息统计，年份 (Year)、出版物 (Pub.)、数据集大小 (Size)、图像中目标数量 (#Obj.)、场景类型 (Types)、深度传感器类型 (Sensor)、分辨率（Resolution）。更多数据集的详细信息请参阅 章节. 3，并可在http://dpfan.net/d3netbenchmark/主页下载数据集。

序号	数据集	年份	出版物.	大小	目标数量	类型	传感器	分辨率
1	**STERE** [102]	2012	CVPR	1,000	∼One	Internet	Stereo camera+sift flow	[251 ∼ 1200]×[222 ∼ 900]
2	**GIT** [20]	2013	BMVC	80	Multiple	Home environment	Microsoft Kinect	640 × 480
3	**DES** [18]	2014	ICIMCS	135	One	Indoor	Microsoft Kinect	640 × 480
4	**NLPR** [105]	2014	ECCV	1,000	Multiple	Indoor/outdoor	Microsoft Kinect	640 × 480, 480 × 640
5	**LFSD** [80]	2014	CVPR	100	One	Indoor/outdoor	Lytro Illum camera	360 × 360
6	**NJUD** [68]	2014	ICIP	1,985	∼One	Movie/internet/photo	FujiW3 camera+optical flow	[231 ∼ 1213]×[274 ∼ 828]
7	**SSD** [201]	2017	ICCVW	80	Multiple	Movies	Sun' s optical flow	960 × 1080
8	**DUT-RGBD** [109]	2019	ICCV	1,200	Multiple	Indoor/outdoor	–	400 × 600
9	**SIP** [38]	2020	TNNLS	929	Multiple	Person in the wild	Huawei Mate10	992 × 744

表. 7 主流光场 SOD 模型方法的总结

序号	年份	模型	出版物	数据集	描述
1	2014	LFS [80]	CVPR	LFSD	第一个光场显著性物体检测工作，基于光场的重聚焦能力来利用似物性和聚焦线索
2	2015	WSC [79]	CVPR	LFSD	利用一个加权的稀疏编码框架学习显著和非显著的字典
3	2015	DILF [173]	IJCAI	LFSD	结合深度对比以弥补色彩的缺点，并利用基于聚焦的背景先验来提高显著性检测性能
4	2016	RL [119]	ICASSP	LFSD	利用光场图像中的固有结构信息来提高显著性检测性能
5	2017	MA [174]	TOMM	HFUT, LFSD	从光场图像中提取的多个显著性线索，并使用基于随机搜索的加权策略整合这些显著性线索
6	2017	BIF [135]	NPL	LFSD	使用两步贝叶斯融合框架，集成基于颜色的对比度，深度引导的对比度，前景切片的聚焦图以及背景加权深度对比度
7	2017	LFS [81]	TPAMI	LFSD	该研究工作 [80] 的扩展
8	2017	RLM [75]	ICIVC	LFSD	利用光场的相对位置测量完成显著性检测任务
9	2018	SGDC [140]	CVPR	LFSD	设计用于多层光场展示的显著性引导的深度图优化框架
10	2018	DCA [108]	FiO	LFSD	提出了一个图模型和深度图引导的细胞自动机，以利用光场数据优化显著性图
11	2019	DLLF [144]	ICCV	DUTLF-FS, LFSD	利用循环注意力网络融合焦点堆栈中的每个片段，以学习最有用的特征
12	2019	DLSD [110]	IJCAI	DUTLF-MV	显著性检测被划分成两个子问题，1）从一个单视图合成光场，2）光场驱动的显著性检测
13	2019	Molf [177]	NIPS	UTLF-FS	使用一个面向存储器的解码器完成光场显著性检测
14	2020	ERNet [111]	AAAI	DUTLF-FS, HFUT, LFSD	使用非对称两分支数据流架构来克服高维光场数据中的计算复杂性和内存短缺的挑战
15	2020	DCA [109]	TIP	LFSD	提出了一种基于深度图引导的细胞自动机（DCA）模型的光场显著性检测框架。DCA 模型能增强空间一致性来优化不精确的显著性图
16	2020	RDFD [150]	MTAP	LFSD	定义从光场焦点堆栈中提取的基于区域的深度图特征描子，以促进利用低层和高层线索来完成显著性检测
17	2020	LFNet [176]	TIP	DUTLF-FS, LFSD, HFUT	利用一个光场细化模块和一个光场融合模块有效地集成光场图像中的多个线索（即焦点，深度图和似物性性特征）
18	2020	LFDCN [172]	TIP	Lytro Illum, LFSD, HFUT	使用改进的 DeepLab-v2 模型来探索光场图像的空间和多视图属性来完成显著性检测

二进制图 M，然后通过比较 M 与真实值 G 来计算精确度（Precision）和召回率（Recall）：

$$Precision = \frac{|M \cap G|}{|M|}, \ Recall = \frac{|M \cap G|}{|G|}. \quad (1)$$

一种流行的策略是使用一组阈值对显著性图 S 进行划分（即，从 0 到 255 变化）。对于每个阈值，我们首先计算一对召回率和精确度得分，然后将它们组合以获得描述该模型在不同阈值下的 PR 曲线。

• **F** 指标 (F_β)。为了全面考虑精度和召回率，F 指标是加权的调和平均值：

$$F_\beta = \left(1 + \beta^2\right) \frac{P * R}{\beta^2 P + R}, \quad (2)$$

其中，β^2 设置为 0.3 以增强精度 [1]。我们使用不同的阈值 $[0, 255]$ 来计算 F 度量值。这将产生一组 F 度量值，我们报告其最大或平均 F_β。

• **MAE** 指标。为了预测的显著性图 S 和真值 G 之间的逐像素平均绝对误差，如下式定义：

$$MAE = \frac{1}{W * H} \sum_{i=1}^{W} \sum_{i=1}^{H} |S_{i,j} - G_{i,j}|, \quad (3)$$

其中，W 和 H 分别表示显著性图的宽和高，MAE 值被正则化为 $[0, 1]$ 区间值。

• **S** 指标 (S_α)。为了获取图像中结构信息的重要性，S_α [33] 用来评估区域感知（S_r）和目标感知（S_o）之间的结构相似性。因此，S_α 可以定义为：

$$S_\alpha = \alpha * S_o + (1 - \alpha) * S_r \quad (4)$$

其中，$\alpha \in [0, 1]$ 是平衡参数。根据 Fan 等人 [33] 建议，其默认值设置为 0.5。

• **E** 指标 (E_ϕ)。E_ϕ [34] 是基于认知视觉研究的基础上提出的，用于获取图像水平统计信息及其局部像素匹

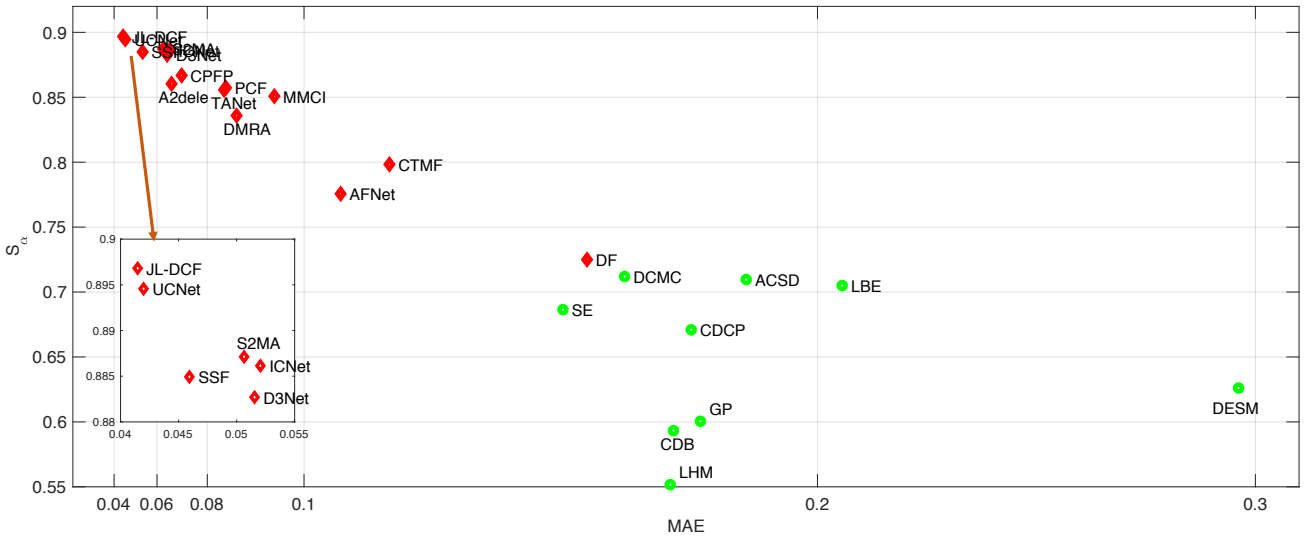


图. 5 对 24 个具有代表性的 RGB-D SOD 模型进行综合评估，包括 LHM [105]、ACSD [68], DESM [18]、GP [117]、LBE [45]、DCMC [26]、SE [48]、CDCP [202]、CDB [84]、DF [115]、PCF [7]、CTMF [50]、CPFP [185]、TANet [9], AFNet [139]、MMCI [12]、DMRA [107], D^{3}Net [38]、SSF [178]、A2dele [112]、S^{2}MA [88]、ICNet [76]、JL-DCF [46] 和 UC-Net [170]。给出每个模型在五个数据集（即，STERE [102]、NLPR [105]、LFSD [80]、DES [18] 和 SIP [38]）上的 S_α 和 MAE 平均值。性能更好的模型显示在左上角（即，具有较大的 S_α 和较小的 MAE 值），其中红色菱形表示深度模型，绿色圆圈表示传统模型。

表. 8 显著性物体尺寸属性研究，对比 24 个具有代表性的 RGB-D SOD 模型（包括 9 个传统模型和 15 个基于深度学习的模型），下表给出了基于 MAE 和 S_α 评价指标的对比结果，其中最好的 3 个结果分别用红、蓝和绿色字体标记。

		Traditional models									Deep learning-based models														
	Scale	LHM [105]	ACSD [68]	DESM [18]	GP [117]	LBE [45]	DCMC [26]	SE [48]	CDCP [202]	CDB [84]	DF [115]	PCF [7]	CTMF [50]	CPFP [185]	TANet [9]	AFNet [139]	MMCI [12]	DMRA [107]	D^{3}Net [38]	SSF [178]	A2dele [112]	S^{2}MA [88]	ICNet [76]	JL-DCF [46]	UC-Net [170]
MAE	Small	.065	.149	.319	.098	.177	.108	.056	.128	.073	.087	.042	.065	.044	.041	.046	.051	.030	.033	.031	.032	.035	.036	.032	.034
	Medium	.178	.183	.287	.180	.210	.158	.150	.173	.179	.152	.068	.107	.055	.067	.095	.079	.069	.053	.045	.054	.052	.052	.041	.042
	Large	.403	.311	.310	.377	.261	.305	.364	.308	.385	.310	.112	.183	.093	.118	.213	.130	.181	.102	.105	.114	.088	.104	.085	.072
	Overall	.166	.184	.296	.173	.206	.156	.142	.171	.167	.147	.065	.102	.055	.065	.091	.076	.067	.052	.046	.053	.051	.052	.041	.042
S_α	Small	.624	.668	.517	.650	.645	.700	.775	.661	.666	.745	.847	.789	.840	.846	.792	.832	.860	.879	.876	.859	.877	.882	.881	.883
	Medium	.543	.732	.658	.598	.723	.727	.676	.683	.585	.730	.863	.805	.877	.862	.779	.859	.838	.888	.893	.865	.893	.892	.906	.901
	Large	.386	.630	.686	.450	.731	.604	.479	.586	.424	.597	.838	.761	.855	.827	.682	.830	.734	.846	.837	.815	.863	.845	.859	.876
	Overall	.552	.710	.626	.601	.705	.712	.686	.671	.593	.725	.857	.798	.867	.856	.776	.851	.836	.883	.885	.860	.887	.886	.897	.895

配信息。因此，E_ϕ 可以定义为：

$$E_\phi = \frac{1}{W * H} \sum_{i=1}^{W} \sum_{i=1}^{H} \phi_{FM}(i,j), \tag{5}$$

其中，ϕ_{FM} 表示增强对角矩阵 [34]。

5.2 性能对比及分析

5.2.1 综合评估

为了量化不同模型的性能，我们对 24 个代表性的 RGB-D SOD 模型进行了全面评估，其中包括 1) 九个传统模型: LHM [105]、ACSD [68]、DESM [18]、GP [117]、LBE [45]、DCMC [26]、SE [48]、CDCP [202] 和 CDB [84]；2) 十五个基于深度学习的模型：DF [115]、PCF [7]、CTMF [50]、CPFP [185]、TANet [9]、AFNet [139]、MMCI [12]、DMRA [107]、D^{3}Net [38]、SSF [178]、A2dele [112]、S^{2}MA [88]、ICNet [76]、JL-DCF [46], 和 UC-Net [170]。我们提供每个模型在五个数据集（STERE [102]、NLPR [105]、LFSD [80]、DES [18] 和 SIP [38]）上的 S_α 和 MAE 的平均值，如图 5所示。值得注意的是，性能更好的模型被展示在左上角（即，S_α 越大和 MAE 越小）。由图 5中的结果可以得到如下观察：

- **传统 vs. 深度模型**：与传统的 RGB-D SOD 模型相比，深度模型的性能明显更好。这证明深度网络具有强大的特征学习能力。

- **深度模型对比**：基于深度学习的模型中，D^{3}Net [38]、JL-DCF [46]、UC-Net [170]、SSF [178]、ICNet

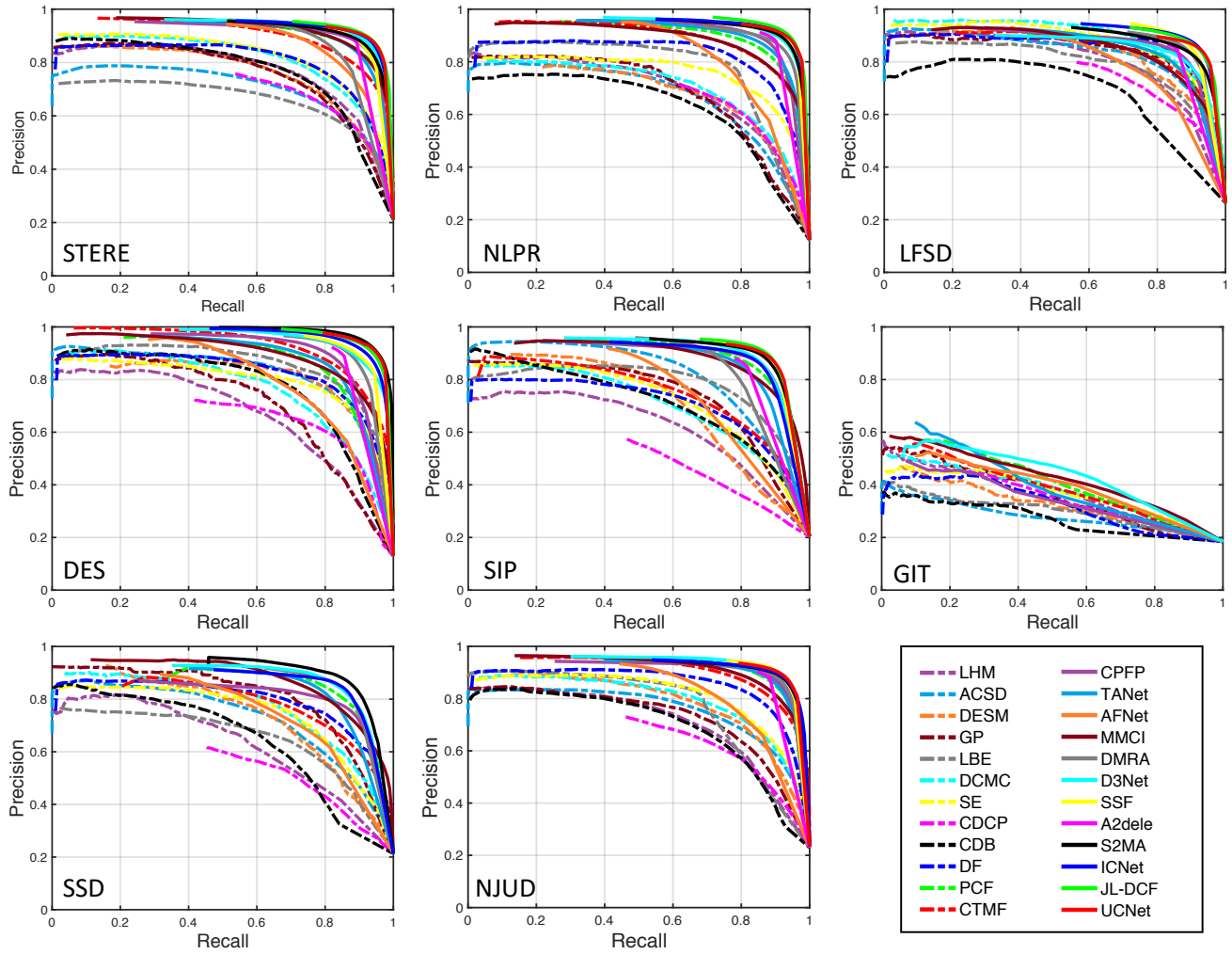


图. 6 24 个 RGB-D SOD 模型在 STERE [102]、NLPR [105]、LFSD [80]、DES [18]、SIP [38]、GIT [20]、SSD [201] 和 NJUD [68] 数据集上的 PR 曲线。

表. 9 基于背景杂斑的属性研究，对比 24 个具有代表性的 RGB-D SOD 模型（包括 9 个传统模型和 15 个基于深度学习的模型），下表给出了基于 MAE 和 S_α 评价指标的对比结果，其中最好的 3 个结果分别用红、蓝和绿色字体标记。

		Traditional models									Deep learning-based models														
	background	LHM [105]	ACSD [68]	DESM [18]	GP [117]	LBE [45]	DCMC [26]	SE [48]	CDCP [202]	CDB [84]	DF [115]	PCF [7]	CTMF [50]	CPFP [185]	TANet [9]	AFNet [139]	MMCI [12]	DMRA [107]	D^3Net [38]	SSF [178]	A2dele [112]	S^2MA [88]	ICNet [76]	JL-DCF [46]	UC-Net [170]
MAE	Simple	.100	.163	.219	.150	.202	.056	.084	.028	.136	.045	.031	.053	.018	.033	.031	.041	.028	.017	.012	.010	.016	.013	.014	.013
	Uncertain	.164	.195	.294	.175	.210	.140	.133	.139	.159	.129	.062	.081	.050	.059	.075	.070	.058	.045	.043	.043	.049	.041	.037	.037
	Complex	.159	.190	.349	.180	.205	.190	.147	.236	.143	.163	.085	.110	.079	.077	.108	.094	.087	.071	.065	.070	.072	.079	.063	.065
	Overall	.160	.193	.295	.174	.209	.140	.132	.141	.157	.127	.063	.082	.051	.059	.076	.070	.059	.046	.043	.043	.049	.043	.038	.038
S_α	Simple	.781	.787	.761	.694	.748	.930	.856	.941	.704	.944	.944	.913	.958	.937	.922	.933	.935	.960	.966	.965	.965	.969	.961	.962
	Uncertain	.572	.694	.638	.606	.695	.736	.723	.727	.610	.774	.873	.853	.882	.873	.818	.868	.854	.900	.894	.884	.895	.910	.909	.907
	Complex	.496	.627	.509	.545	.616	.577	.605	.487	.575	.627	.782	.742	.787	.790	.694	.768	.751	.822	.815	.786	.813	.808	.829	.833
	Overall	.576	.693	.633	.606	.691	.732	.720	.718	.612	.770	.869	.847	.878	.869	.813	.863	.850	.896	.891	.879	.892	.904	.904	.904

[76] 和 S^2MA [88] 获得了相对其它模型更好的性能。

此外，由图 6和图 7 所示为 24 个代表性 RGB-D SOD 模型在八个数据集（即，STERE [102]、NLPR [105]、LFSD [80]、DES [18]、SIP [38]、GIT [20]、SSD [201] 和 NJUD [68]）上的 PR 曲线和 F 指标曲线。其中数据集 NLPR、LFSD、DES、SIP、GIT 和 SSD 的测试样本分别为 1000、300、100、135、929 和 80。对于 NJUD [68] 数据集，CPFP [185]、S^2MA [88]、ICNet [76]、

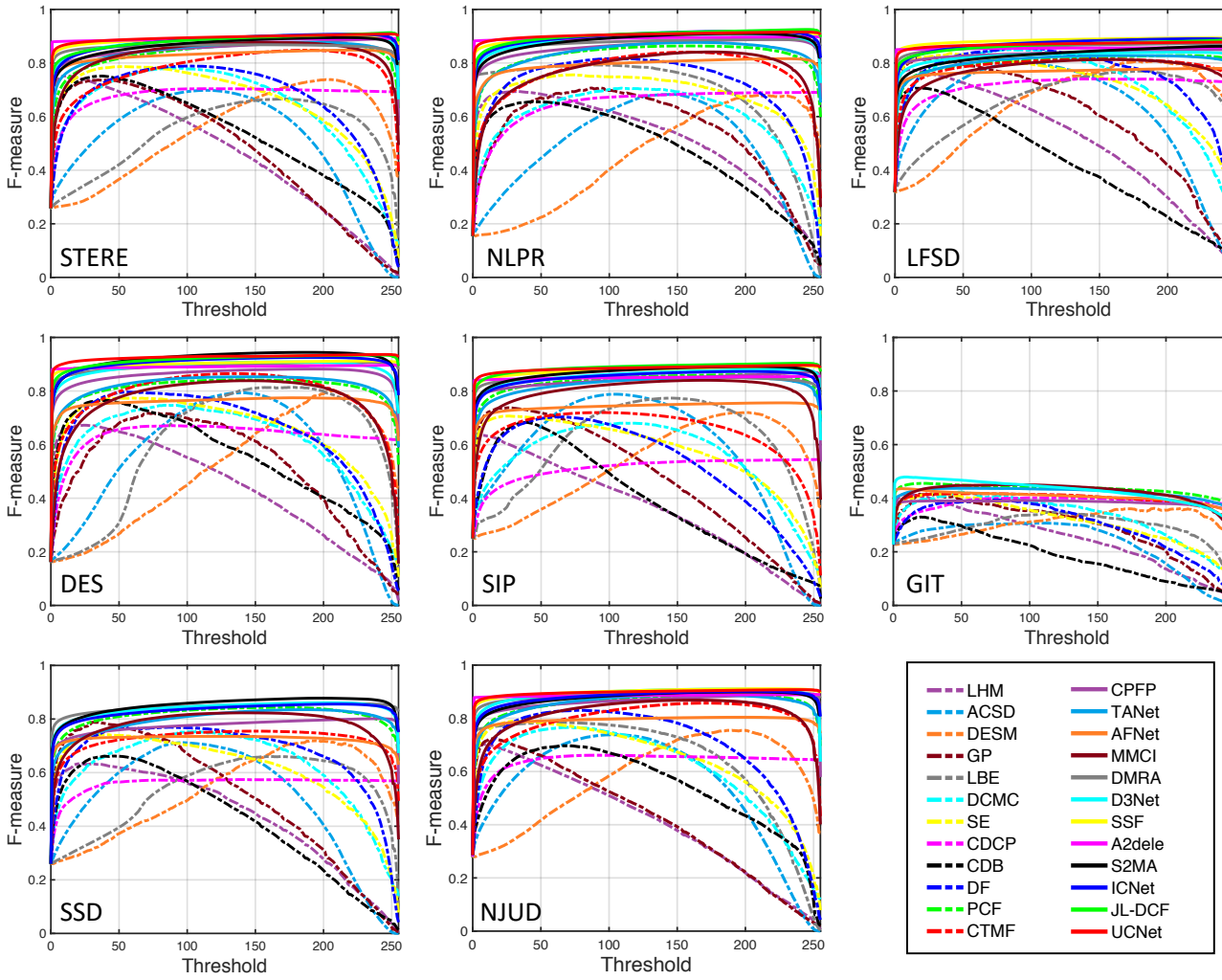


图. 7 24 个 RGB-D SOD 模型在 STERE [102]、NLPR [105]、LFS [80]、DES [18]、SIP [38]、GIT [20]、SSD [201] 和 NJUD [68] 数据集上的 F 指标曲线。

JL-DCF [46] 和 UC-Net [170] 模型的评价是基于 485 个测试样本，而余下其它模型是基于 498 个测试样本。

为了深入理解取得较好性能的前六个 RGB-D SOD 模型，我们接下来讨论它们的主要优势。

• D^3Net [38] 由两个关键组件组成，即一个三分支数据流的特征学习模块和一个深度图过滤单元（depth depurator unit，DDU）。在三分支数据流的特征学习模块中，有三个子网，即 RgbNet、RgbdNet 和 DepthNet。RgbNet 和 DepthNet 用于分别学习 RGB 和深度图的高层级特征表示，而 RgbdNet 用于学习其融合的表示。值得注意的是，此三分支数据流的特征学习模块可以获得模态的特定信息以及模态之间的相关性。因此，平衡这两个方面对于多模态学习至关重要，并且有助于提高 SOD 性能。此外，深度图过滤单元还充当了一个滤除低质量深度图的门连接，然而现有的一些方法并没有考虑这种低质量深度图的影响。由于低质量的深度图会抑制 RGB 图像和深度图之间的融合，因此 DDU 单元可以确保有效的多模态融合，以完成鲁棒的 SOD。

• JL-DCF [46] 由两个关键组成部分，即联合学习（joint learning，JL）和密集协作融合（densely-cooperative fusion，DCF）。具体而言，JL 模块用于学习鲁棒的显著性特征，而 DCF 模块用于发现互补性特征。值得注意的是，该方法使用中间融合策略从 RGB 图像和深度图中提取深度层次化特征，可以有效地利用跨模态互补性来实现准确的预测。

• UC-Net [170] 模型不是生成单个显著性预测图，而是通过对特征输出空间的分布建模来产生多个显著性预测。因为每个人在标记显著性图时都有一些特定的偏好，当使用确定性学习流程为图像对生成单个显著性图时，它可能无法获取显著性的随机特性。因此，该模型

考虑到人们在标定显著性物体时的不确定性。此外，考虑到深度图可能会受到噪声的影响，将 RGB 图像和深度图直接融合可能会使网络拟合这种噪声。因此，提出了一个辅助的深度图校正网络，并利用语义指导的损失来精炼深度信息。因此，以上关键组件均有助于提高 SOD 性能。

• SSF [178] 开发了一个互补交互模块（complementary interaction module，CIM），用来探索可区分性跨模态互补性和融合跨模态特征，这里其引入了区域感知注意力机制为每种模态补充丰富的边界信息。此外，提出了一种补偿感知损失，以提高网络对不可靠深度图中的较难学习样本的置信度。因此，这些关键组件使所提出的模型能够有效地探索和建立跨模态特征表示的互补性，同时降低由低质量深度图带来的负面影响，从而提高 SOD 性能。

• ICNet [76] 提出了一种信息转换模块，以交互和自适应地探索高层级 RGB 与深度图特征之间的相关性。此外，引入了一种跨模态的深度图加权组合模块，以增强每个特征层级中 RGB 和深度图特征之间的差异，从而确保对这些特征进行不同的处理。还值得注意的是，ICNet 充分利用了跨模态特征的互补性，并探索了跨层级特征的连续性，这两者都有助于完成准确的预测。

• S^2MA [88] 提出了一种自交互的注意力模块（self-mutual attention module，SAM），以融合 RGB 和深度图，集成自注意力和交互注意力机制来更准确地传播上下文信息。SAM 能为多模态数据提供补充信息，以提高 SOD 性能，从而克服仅使用一种模态的自注意力机制的局限性。此外，为了减小低质量深度图（例如噪声）的影响，提出了一种选择机制来重新加权交互注意力信息。这种机制可以过滤掉不可靠的信息，从而实现更准确的显著性预测。

5.2.2 属性评估

为了研究不同属性对显著性检测性能的影响，如物体尺寸、背景杂斑、显著性物体数量、室内或室外场景、背景物体、光照条件等，本文对几种具有代表性的 RGB-D SOD 模型进行了基于属性的性能评估。

• **物体尺寸**。为了表征显著性物体的尺寸大小，本文计算显著性区域的尺寸与整幅图像的比率，并定义了三种物体尺寸：1) 当该比率小于 0.1 时，标记为“小”目标；2) 当该比率大于 0.4 时，标记为“大”目标；3) 当该比率在

图. 8 不同物体尺寸的样本图片，具体尺寸比例如黄色字体标记。

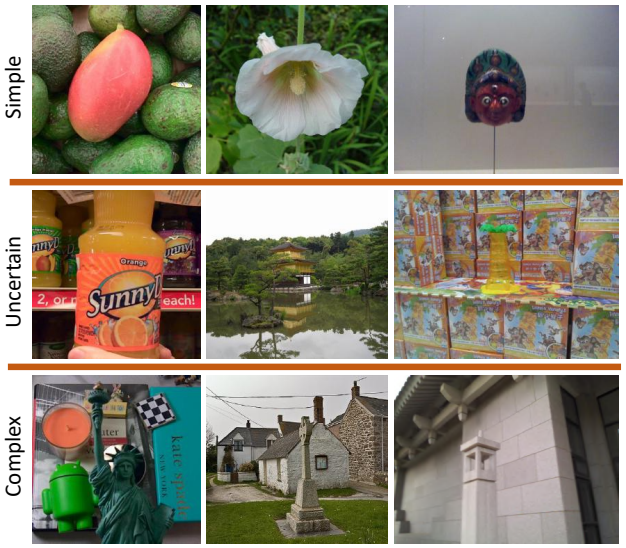


图. 9 3 种类型的背景杂斑样本图片

[0.1, 0.4] 区间内时，标记为“中等”目标。为了评估目标物体的尺寸对 SOD 性能的影响，构建了一个混合数据集，包含 STERE [102]、NLPR [105]、LFSD [80]、DES [18] 和 SIP [38]，一共收集 2464 幅图像，其中小、中等和大的显著性物体图像分别占比 24%、69.2% 和 6.8%。构建的混合数据集可在 https://github.com/taozh2017/RGBD-SODsurvey主页上获取。如图 8所示，一些具有不同目标比率的样本图像，基于属性研究的对比结果如表 8所示。由表中结果可知，所有对比方法在检测小的显著性物体时都获得了较好的性能，而检测较大的显著性物体时它们的性能都会下降。此外，最新的 JL-DCF [46]、UC-Net [170] 和 S^2MA [88] 模型取得了前三的最佳性能，而 D^3Net [38]、SSF [178]、A2dele [112] 和 ICNet [76] 也获得了相对较好的性能。

表. 10 背景物体（如汽车、障碍物、花、草、路、标牌、树木和其它类）属性研究，对比 24 个具有代表性的 RGB-D SOD 模型（包括 9 传统模型和 15 和基于深度学习的模型），显示了在 SIP [38] 数据集上基于 MAE 和 S_α 指标的评价结果，其中最好的三个结果分别用红、蓝和绿色字体标记。

		Traditional models									Deep learning-based models														
	Categories	LHM [105]	ACSD [68]	DESM [18]	GP [117]	LBE [45]	DCMC [26]	SE [48]	CDCP [202]	CDB [84]	DF [115]	PCF [7]	CTMF [50]	CPFP [185]	TANet [9]	AFNet [139]	MMCI [12]	DMRA [107]	D^{3}Net [38]	SSF [178]	A2dele [112]	S^{2}MA [88]	ICNet [76]	JL-DCF [46]	UC-Net [170]
MAE	Car	.158	.163	.301	.159	.201	.185	.154	.202	.171	.171	.085	.134	.094	.084	.101	.093	.069	.061	.063	.078	.055	.067	.058	.057
	Barrier	.197	.177	.308	.180	.201	.196	.176	.251	.203	.202	.073	.149	.060	.078	.128	.089	.093	.068	.054	.074	.057	.075	.052	.053
	Flower	.105	.122	.306	.099	.186	.158	.063	.141	.101	.132	.091	.075	.133	.100	.090	.081	.046	.095	.107	.051	.104	.025	.054	.075
	Grass	.164	.161	.279	.155	.184	.167	.138	.182	.176	.167	.041	.110	.035	.048	.088	.059	.056	.037	.030	.046	.033	.043	.023	.029
	Road	.189	.167	.281	.176	.187	.181	.164	.225	.189	.169	.070	.140	.054	.072	.125	.078	.093	.059	.049	.072	.050	.065	.045	.044
	Sign	.107	.126	.268	.110	.184	.126	.079	.134	.118	.096	.058	.101	.063	.060	.077	.083	.051	.055	.051	.054	.048	.054	.050	.057
	Tree	.192	.193	.310	.190	.241	.194	.183	.230	.219	.205	.083	.157	.083	.091	.132	.109	.106	.083	.067	.074	.092	.097	.063	.071
	Other	.246	.217	.329	.224	.229	.216	.229	.274	.233	.233	.106	.177	.111	.111	.170	.124	.140	.095	.083	.099	.100	.100	.084	.086
	Overall	.184	.172	.298	.173	.200	.186	.164	.224	.192	.185	.071	.139	.064	.075	.118	.086	.085	.063	.053	.070	.057	.069	.049	.051
S_α	Car	.516	.731	.590	.603	.714	.671	.591	.613	.546	.631	.811	.726	.786	.807	.736	.813	.817	.856	.845	.804	.870	.846	.855	.859
	Barrier	.497	.727	.609	.575	.728	.672	.612	.553	.552	.643	.837	.698	.860	.831	.708	.830	.792	.855	.874	.821	.871	.848	.876	.875
	Flower	.477	.775	.573	.673	.703	.707	.772	.667	.639	.750	.771	.738	.714	.760	.688	.785	.824	.789	.768	.845	.804	.901	.856	.811
	Grass	.537	.756	.643	.605	.760	.728	.683	.672	.559	.672	.908	.770	.908	.899	.780	.888	.876	.917	.924	.878	.928	.910	.939	.924
	Road	.521	.739	.634	.598	.751	.685	.641	.595	.576	.680	.851	.722	.871	.848	.705	.847	.807	.873	.885	.832	.885	.868	.889	.892
	Sign	.578	.786	.634	.628	.719	.745	.761	.714	.615	.757	.855	.756	.833	.857	.771	.818	.848	.849	.849	.842	.871	.861	.859	.840
	Tree	.505	.699	.606	.577	.661	.648	.600	.588	.543	.625	.802	.679	.804	.778	.691	.779	.748	.806	.837	.807	.800	.788	.848	.825
	Other	.460	.687	.594	.532	.706	.669	.563	.554	.542	.600	.786	.677	.774	.782	.647	.790	.722	.800	.828	.785	.809	.799	.821	.823
	Overall	.511	.732	.616	.588	.727	.683	.628	.595	.557	.653	.842	.716	.850	.835	.720	.833	.806	.860	.874	.828	.872	.854	.880	.875

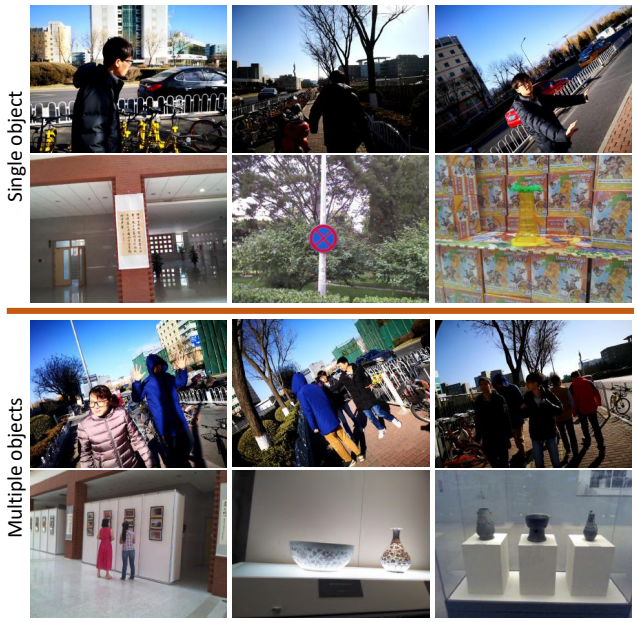


图. 10 单个或多个显著性物体的样本图片

• **背景杂斑**。如何直接地表征背景杂斑是一个困难的任务。由于经典的 SOD 方法倾向于利用先验信息或颜色对比度来定位显著性物体，但这些方法在复杂背景下会失败。因此，在该属性评估中，我们利用 5 个传统的显著性物体检测模型，即 BSCA [114]、CLC [191]、MDC [60]、MIL [55] 和 WFD [59]，检测图像中的显著性物体，然后根据检测结果将这些图像分为不同类（比如，简单或复杂的背景）。具体地，我们首先构建一个混合数据集，包含以下三个数据集 (STERE [102]、NLPR [105] 和 LFSD [80])，一共 1400 幅图像。我们将这 5 个传统 SOD 模型应用于该数据集，获得每个模型的 S_α 值，并用如下规则来表征图像：1) 若 S_α 值大于 0.9, 则该图像标记为“简单”背景；2) 若 S_α 值小于 0.6, 则该图像标记为“复杂”背景; 3) 其余的图像标记为“不确定”的背景复杂情况。图 9显示了部分不同背景杂斑的图片。创建的数据集并可通过链接 https://github.com/taozh2017/RGBD-SODsurvey获取。基于背景杂斑的属性评价的结果如表 9所示。由表可知，所有模型在处理复杂背景下的显著性检测性能均低于简单背景下的性能。在这些代表性的模型中，JL-DCF [46], UC-Net [170] 和 SSF [178] 取得了前三的最佳性能。此外，先进的四个模型 D^{3}Net [38]、S^{2}MA [88]、A2dele [112] 和 ICNet [76] 也获得相对较好的性能。

• **单个 vs. 多个物体**。在此评估中，我们构建了一个混合数据集，包含 NLPR [105] 和 SIP [38] 两个数据集中的 1229 个样本图像。对比结果如图 11所示，由结果可知检测单个显著性物体要比多个显著性物体更加容易。

• **室内 vs. 室外**。本文评估不同 RGB-D SOD 模型

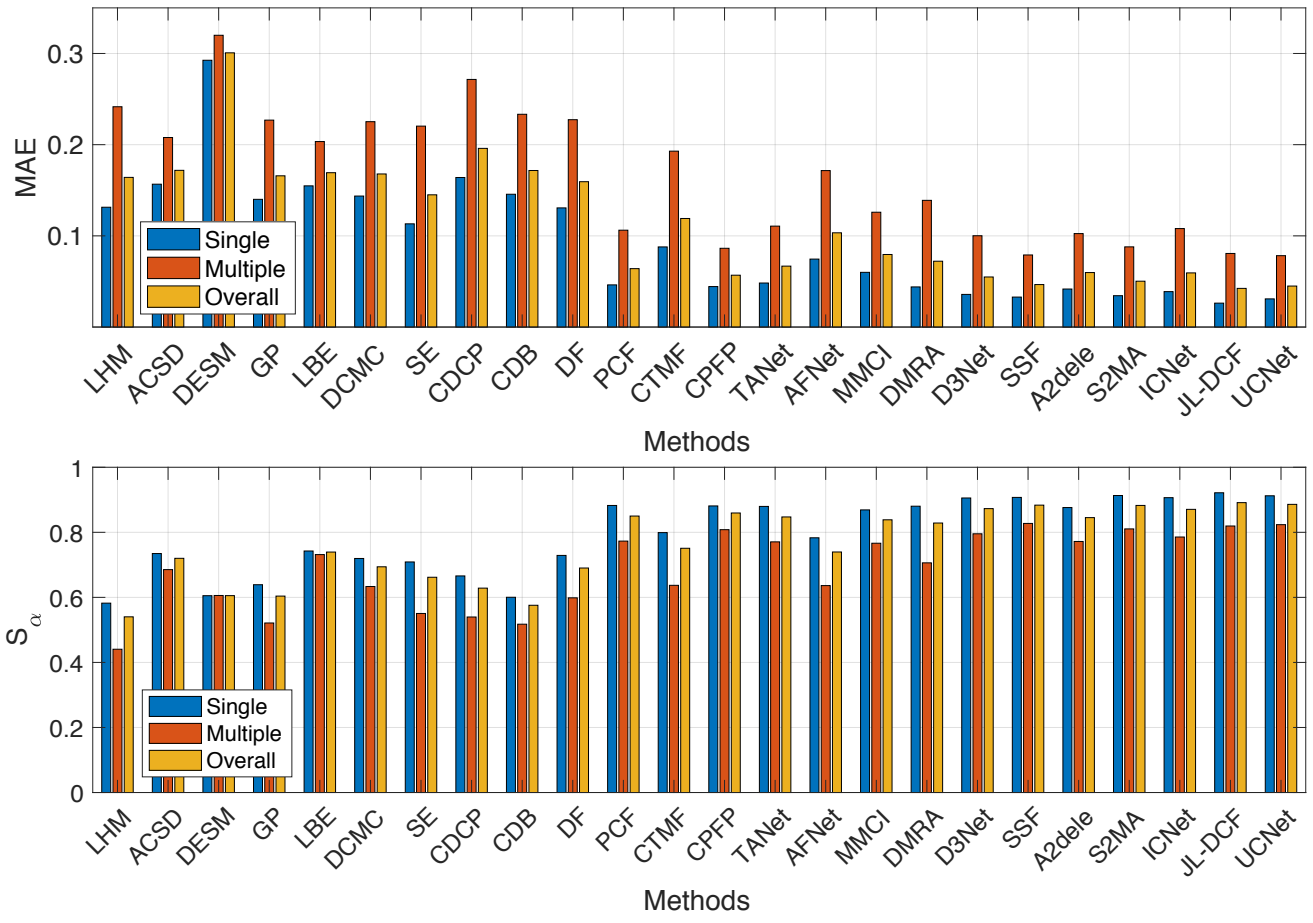


图. 11 基于显著性物体数量的属性研究（即，单个或多个）。显示了 24 个具有代表性的 RGB-D SOD 模型（即，LHM [105]、ACSD [68]、DESM [18]、GP [117]、LBE [45]、DCMC [26]、SE [48]、CDCP [202]、CDB [84]、DF [115]、PCF [7], CTMF [50]、CPFP [185]、TANet [9], AFNet [139]、MMCI [12]、DMRA [107], D^{3}Net [38]、SSF [178]、A2dele [112], S^{2}MA [88]、ICNet [76]、JL-DCF [46] 和 UC-Net [170]）的 MAE（顶部）和 S_α（底部）指标的对比结果。

表. 11 不同光照条件 (阳光 vs. 低光照) 的属性研究，对比 24 个具有代表行的 RGB-D SOD 模型（包括 9 个传统模型和 15 个基于深度学习的模型），显示了在 SIP [38] 数据集上 MAE 和 S_α 的对比结果，其中最好的三个结果分别用红、蓝和绿色字体标记。

		Traditional models									Deep learning-based models														
	Conditions	LHM [105]	ACSD [68]	DESM [18]	GP [117]	LBE [45]	DCMC [26]	SE [48]	CDCP [202]	CDB [84]	DF [115]	PCF [7]	CTMF [50]	CPFP [185]	TANet [9]	AFNet [139]	MMCI [12]	DMRA [107]	D^{3}Net [38]	SSF [178]	A2dele [112]	S^{2}MA [88]	ICNet [76]	JL-DCF [46]	UC-Net [170]
MAE	Sunny	.182	.171	.294	.171	.200	.183	.160	.218	.190	.181	.069	.137	.062	.075	.116	.085	.083	.062	.052	.068	.057	.068	.048	.051
	Low-light	.198	.178	.323	.187	.201	.207	.193	.268	.208	.211	.078	.154	.073	.076	.130	.091	.103	.067	.059	.080	.058	.081	.059	.055
	Overall	.184	.172	.298	.173	.200	.186	.164	.224	.192	.185	.071	.139	.064	.075	.118	.086	.085	.063	.053	.070	.057	.069	.049	.051
S_α	Sunny	.516	.733	.622	.593	.728	.690	.639	.607	.560	.660	.843	.718	.852	.834	.723	.833	.811	.861	.875	.831	.872	.856	.882	.876
	low-light	.481	.721	.573	.554	.722	.635	.556	.515	.543	.610	.838	.701	.838	.837	.700	.832	.775	.855	.867	.810	.871	.839	.867	.871
	Overall	.511	.732	.616	.588	.727	.683	.628	.595	.557	.653	.842	.716	.850	.835	.720	.833	.806	.860	.874	.828	.872	.854	.880	.875

在室内和室外场景下的性能，该评估实验构建了一个从 DES [18], NLPR [105] 和 LFSD [80] 数据集中收集的混合数据集。对比结果如图 12所示。由图中结果可知，大多数模型在室内检测显著性物体的性能要略差于

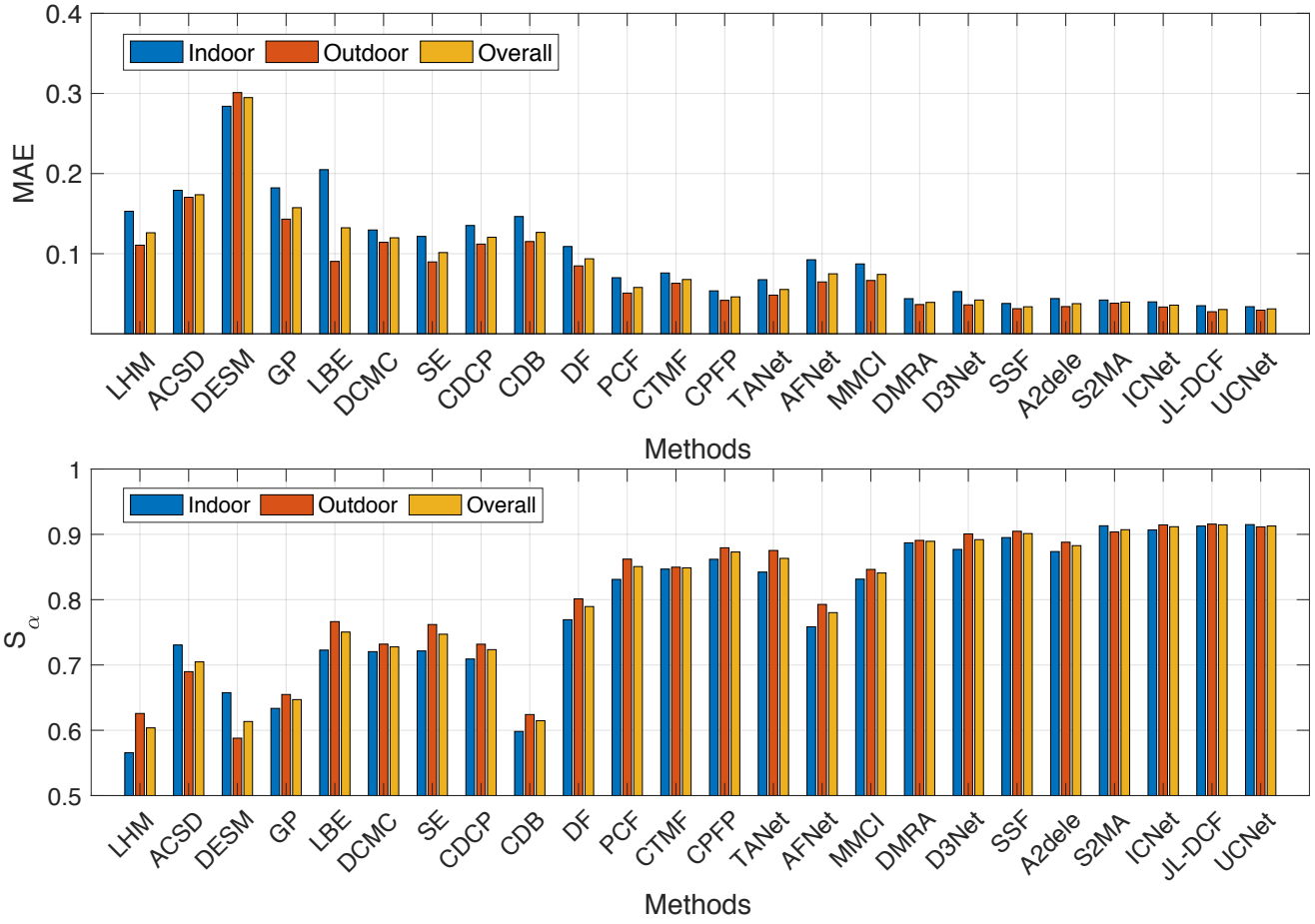


图. 12 基于室内室外场景的属性研究。显示了 24 个具有代表性的 RGB-D SOD 模型（即，LHM [105]、ACSD [68]、DESM [18]、GP [117]、LBE [45]、DCMC [26]、SE [48]、CDCP [202]、CDB [84]、DF [115]、PCF [7], CTMF [50]、CPFP [185]、TANet [9], AFNet [139]、MMCI [12]、DMRA [107], D^{3}Net [38]、SSF [178]、A2dele [112], S^{2}MA [88]、ICNet [76]、JL-DCF [46] 和 UC-Net [170]）的 MAE（顶部）和 S_α（底部）指标的对比结果。

室外场景，这可能是由于室内环境光照变化所引起的。

• **背景物体**。我们评估 RGB-D SOD 模型在不同的背景物体下的性能。我们采用 SIP [38] 数据集，将其分为八个类别，即汽车、障碍物、花、草、道路、标志、树木以及其它类。对比结果见表 10。由表可知，所有方法在不同背景物体下都获得了不同的性。在 24 个代表性的 RGB-D SOD 模型中，JL-DCF [46]、UC-Net [170] 和 SSF [178] 取得了前三的最佳性能。此外，四个新的模型 D^{3}Net [38]、S^{2}MA [88]、A2dele [112] 和 ICNet [76] 也获得了相对较好的性能。

• **光照变化**。SOD 性能可能会受到不同光照条件的影响。为了验证光照对不同 RGB-D SOD 模型性能的影响，本文在 SIP [38] 数据集上进行了该属性评估，并将该数据集分为两类，即晴朗和弱光。对比结果见表 11所示。由此可知，弱光对 SOD 性能会产生负面影响。在这些对比模型中，UC-Net [170] 在晴朗的条件下获得了最佳性能，而 JL-DCF [46] 在弱光条件下获得了最优性能。

此外，我们展示在各种挑战性的场景中所产生的显著性预测图像，以可视化不同 RGB-D SOD 模型的性能。图 13和图 14展示了两个经典的非深度学习方法（DCMC [26] 和 SE [48]）及八个基于 CNN 的最新模型（DMRA [107], D^{3}Net [38], SSF [178], A2dele [112], S^{2}MA [88], ICNet [76], JL-DCF [46], 和 UC-Net [170]）的显著性预测结果图。其中第一行显示小的显著性物体，第二行展示了大的显著性物体。第三行和第四行的

图. 13 两个经典非深度学习方法（DCMC [26] 和 SE [48]）和三个基于 CNN 的最新模型（DMRA [107]、D^{3}Net [38]、SSF [178]）的可视化对比。

场景中分别包含复杂的背景和边界，第五和第六行的场景中包含多个显著性物体。第七行为弱光环境，第八行包含物体边界不准确的深度图，它们可能会抑制 SOD 性能。由图 13和图 14中可视化结果可看出，在这些挑战性的场景中，基于深度学习的模型其性能要优于非深度学习模型，证明了深度特征相对于手工特征具有更强大的表达能力。此外，D^{3}Net [38], S^{2}MA [88], JL-DCF [46], 和 UC-Net [170] 性能要优于其它深度模型。

6 挑战与展望

6.1 不完美深度图的影响

低质量深度图的影响。事实证明，具有丰富空间信息的深度图有助于从杂乱的背景中检测出显著性物体，而深度图的质量也直接影响 SOD 的性能。由于深度传感器的局限性，深度图的质量在不同的情况下会发生很大的变化，如何减小低质量深度图的影响仍是一个挑战。然而，大多数已有的方法直接将 RGB 图像和深度图的原始数据融合在一起，而不考虑低质量深度图的影响。有一些值得注意的例外，在 [185] 方法中，提出了一个对比度增强网络来学习增强的深度图，该图与原始的深度图相比具有更高的对比度。在 [178] 中，提出了

图. 14 五个基于 CNN 的最新模型（A2dele [112]、S^2MA [88]、ICNet [76]、JL-DCF [46] 和 UC-Net [170]）的可视化对比结果。

一种补偿损失函数，以便更多地关注包含不可靠深度信息的样本。此外，D^3Net [38] 利用一个深度图过滤单元（DDU）将深度图分为两类（即合格和低质量）。该 DDU 还可以作为过滤单元过滤掉低质量的深度图。但上述方法通常使用两步策略来实现深度图增强和多模态融合 [178, 185]，或者采用独立的门连接操作来过滤掉较差的深度图，这可能会导致次最优问题。因此，需要开发一种端到端框架，在多模态融合阶段，实现深度图增强或者自适应地对深度图进行加权（例如，为较质量差的深度图分配低权重），这将有助于减少低质量深度图的影响来提高 SOD 的性能。

不完整的深度图。在 RGB-D 数据集中，由于采集设备的限制，不可避免地会获得一些低质量的深度图。如前所述，许多深度增强算法已经用来改善深度图的质量。但是，经常会丢弃遭受严重噪声或边缘模糊的深度图。在这种情况下，我们有完整的 RGB 图像，但有些样本没有深度图，这类似于不完整的多视图/模态学习问题 [157, 192–195]。因此，我们称其为“不完整 RGB-D SOD”。由于当前 SOD 模型专注于使用完整的 RGB 图像和深度图，因此我们认为这可能是 RGB-D SOD 的任务中的新方向。

深度估计。深度估计提供了一种有效的解决方案，可恢复高质量的深度图并克服低质量深度图的影响。目前

已有各种深度估计方法 [47, 66, 86, 138]，可将其引入基于 RGB-D 的 SOD 任务中以提高性能。

6.2 有效的融合策略

基于对抗学习的融合。对于 RGB-D SOD 任务，有效融合 RGB 图像和深度图是至关重要的。现有模型常采用不同的融合策略（例如，早期融合、中期融合或者后期融合）来挖掘 RGB 图像和深度图之间的相关性。最近，生成对抗网络（generative adversarial networks，GANs） [99] 应用于 SOD 检测任务 [103, 203] 引起了广泛的关注。在基于 GAN 的 SOD 模型中，生成器将 RGB 图像作为输入并生成相应的显著性图，同时采用鉴别器来判断给定图像是合成的还是真实的图像。GAN 模型能容易地扩展到 RGB-D SOD，由于其卓越的特征学习能力，这将有助于提升 SOD 性能。此外，GAN 还可以用于学习 RGB 图像和深度图的公共特征表示 [64]，这有助于特征或显著性图融合并进一步提高 SOD 性能。

注意力引导的融合。注意力机制已广泛应用于许多基于深度学习的任务中 [43, 133, 136, 146]，该机制允许网络选择性地关注一些区域的子集，以提取更具区分性和强大能力的特征。此外，已经提出了协同注意机制来探索多种模态之间的潜在相关性，并且在视觉问答 [92, 162] 和视频目标分割 [93] 中展开了广泛的研究。因此，对于 RGB-D SOD 任务，可以开发基于注意力机制的融合算法，以挖掘 RGB 图像和深度图之间的相关性来提高性能。

6.3 不同的监督策略

现有的 RGB-D 模型通常使用全监督策略来学习显著性预测模型。然而，标注像素级显著性图是一个繁琐且耗时的过程。为了缓解这个问题，人们越来越关注弱监督和半监督学习，这些策略已经应用于显著性物体检测 [113, 159, 164, 168, 199]。通过利用图像级标签 [164] 和伪逐像素注释 [159, 166]，可以将半/弱监督引入 RGB-D SOD 模型中，以提高显著性检测性能。此外，一些研究 [16, 27] 表明，使用自监督预训练模型可以获得更好的性能。因此，我们也可以采用自监督的方式在大量带标注的 RGB 图像上训练显著性检测模型，然后将预训练的模型迁移到 RGB-D SOD 任务。

6.4 数据收集

数据集大小。尽管 SOD 领域已有 9 个公开的 RGB-D 数据集，但它们的规模是非常限制的。例如，最大的 NJUD 数据集仅有大约 2000 个样本 [68]。与用于通用物体检测或动作识别的其他 RGB-D 数据集相比 [69, 171]，SOD 领域的 RGB-D 数据集都非常小。因此，开发新的大规模 RGB-D 数据集作为未来研究的基准至关重要。

复杂背景 & 任务驱动的数据集。大多数 RGB-D 数据集都会收集包含一个或多个显著性对象，但背景都是相对干净的场景。然而，实际应用中通常会遇到更为复杂的情况（例如，遮挡、外观变化、低光照等），这可能会降低 SOD 性能。因此，收集具有复杂背景的图像对于提高 RGB-D SOD 模型的泛化能力至关重要。此外，对于某些任务，必须收集具有特定显著性物体的图像。例如，路标识别作为驾驶员辅助系统中的一项重要的技术，这需要收集带有路标的图像。因此，构建像 SIP [38] 这样的任务驱动的 RGB-D 数据集至关重要。

6.5 真实场景中的模型设计

许多智能手机可以获取深度图 (例如，使用华为 Mate10 获取的 SIP 数据集)。因此，在现实世界应用中，如在智能设备上进行 SOD 任务将是可行的。然而，大多数现有的方法都包括复杂和较深的 DNN 网络，通过增加模型容量并实现更好的性能，这限制了将其直接应用于实际工作平台。为了克服这个问题，可以使用模型压缩 [19, 52] 技术来学习轻量紧凑的 RGB-D SOD 模型去获得较好的检测精度。此外，JL-DCF [46] 采用 RGB 图像和深度图学习一个共享网络来定位显著物体，这极大地减少了模型参数，并使实际应用变得可行。

6.6 扩展到 RGB-T SOD

除了 RGB-D SOD 外，还有其它几种融合了不同模态以进行更好检测的模型，例如 RGB-T SOD，它融合了 RGB 和红外数据。热红外摄像机可以获取温度高于绝对零值的任何物体发出的辐射，从而使热红外图像对光照条件不敏感 [96]。因此，当显著的物体遭受反射光或阴影光照变化时，热图像可以提供补充信息以提高 SOD 性能。在过去几年中，涌现了许多 RGB-T SOD 模型 [74, 96, 126, 129–132, 137, 182] 和相关数据集 (VT821 [137]、VT1000 [132] 以及 VT5000 [130])。与 RGB-D SOD 相似，RGB-T SOD 的主要目标是融合 RGB 和红外热图像，并利用两种模态之间的相关性。因此，可以将 RGB-D SOD 中先进的多模态融合技术扩展到 RGB-T SOD 任务中。

7 结论

本文首次对基于 RGB-D 的 SOD 模型进行了系统全面的综述。我们首先从不同角度审视与总结 RGB-D SOD 模型，以及整理流行的 RGB-D SOD 数据集并简述每个数据集的细节。考虑到光场也能提供深度信息这一事实，我们还回顾了主流的光场 SOD 模型和相关的基准数据集。接下来，我们构建新的数据集并对 24 个代表性的 RGB SOD 模型进行了综合评估以及基于属性的评估。此外，我们讨论了一些挑战并强调了未来研究的开放方向。我们也简要地讨论了 RGB-T SOD 的扩展工作，以提高当物体遭受反射光或阴影等光照变化的情况下显著性物体检测的性能。尽管基于 RGB-D 的 SOD 在过去的几十年中取得了显著的进步，但仍有很大的改进空间。我们希望本综述会激发学者们对该领域的更多研究兴趣。

References

[1] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk. Frequency-tuned salient region detection. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 1597–1604. IEEE, 2009.

[2] A. Borji, M.-M. Cheng, Q. Hou, H. Jiang, and J. Li. Salient object detection: A survey. *Computational Visual Media*, pages 1–34, 2014.

[3] A. Borji, M.-M. Cheng, H. Jiang, and J. Li. Salient object detection: A benchmark. *IEEE Transactions on Image Processing*, 24(12):5706–5722, 2015.

[4] C. Chen, J. Wei, C. Peng, and H. Qin. Depth quality aware salient object detection. *IEEE Transactions on Image Processing*, 2020.

[5] C. Chen, J. Wei, C. Peng, W. Zhang, and H. Qin. Improved saliency detection in RGB-D images using two-phase depth estimation and selective deep fusion. *IEEE Transactions on Image Processing*, 29:4296–4307, 2020.

[6] H. Chen, Y. Deng, Y. Li, T.-Y. Hung, and G. Lin. Rgbd salient object detection via disentangled cross-modal fusion. *IEEE Transactions on Image Processing*, 29:8407–8416, 2020.

[7] H. Chen and Y. Li. Progressively complementarity-aware fusion network for RGB-D salient object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3051–3060, 2018.

[8] H. Chen and Y. Li. Cnn-based rgb-d salient object detection: Learn, select and fuse. *arXiv preprint arXiv:1909.09309*, 2019.

[9] H. Chen and Y. Li. Three-stream attention-aware network for RGB-D salient object detection. *IEEE Transactions on Image Processing*, 28(6):2825–2835, 2019.

[10] H. Chen, Y. Li, and D. Su. RGB-D saliency detection by multi-stream late fusion network. In *Proceedings of the International Conference on Computer Vision Systems*, pages 459–468. Springer, 2017.

[11] H. Chen, Y. Li, and D. Su. Discriminative cross-modal transfer learning and densely cross-level feedback fusion for RGB-D salient object detection. *IEEE Transactions on Cybernetics*, 2019.

[12] H. Chen, Y. Li, and D. Su. Multi-modal fusion network with multi-scale multi-path and cross-modal interactions for RGB-D salient object detection. *Pattern Recognition*, 86:376–385, 2019.

[13] H. Chen, Y.-F. Li, and D. Su. M3net: Multi-scale multi-path multi-modal fusion network and example application to rgb-d salient object detection. In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 4911–4916. IEEE, 2017.

[14] H. Chen, Y.-F. Li, and D. Su. Attention-aware cross-modal cross-level fusion network for RGB-D salient object detection. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 6821–6826. IEEE, 2018.

[15] S. Chen and Y. Fu. Progressively guided alternate refinement network for rgb-d salient object detection. In *Proceedings of the European Conference on Computer Vision*. Springer, 2020.

[16] T. Chen, S. Liu, S. Chang, Y. Cheng, L. Amini, and Z. Wang. Adversarial robustness: From self-supervised pre-training to fine-tuning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 699–708, 2020.

[17] Z. Chen and Q. Huang. Depth potentiality-aware gated attention network for RGB-D salient object detection. *arXiv preprint arXiv:2003.08608*, 2020.

[18] Y. Cheng, H. Fu, X. Wei, J. Xiao, and X. Cao. Depth enhanced saliency detection method. In *Proceedings of the International Conference on Internet Multimedia Computing and Service*, pages 23–27, 2014.

[19] Y. Cheng, D. Wang, P. Zhou, and T. Zhang. A survey of model compression and acceleration for deep neural networks. *arXiv preprint arXiv:1710.09282*, 2017.

[20] A. Ciptadi, T. Hermans, and J. M. Rehg. An in depth view of saliency. Georgia Institute of Technology, 2013.

[21] R. Cong, J. Lei, H. Fu, M.-M. Cheng, W. Lin, and Q. Huang. Review of visual saliency detection with comprehensive information. *IEEE Transactions on Circuits and Systems for Video Technology*, 29(10):2941–2959, 2018.

[22] R. Cong, J. Lei, H. Fu, J. Hou, Q. Huang, and S. Kwong. Going from RGB to RGBD saliency: A depth-guided transformation model. *IEEE Transactions on Cybernetics*, 2019.

[23] R. Cong, J. Lei, H. Fu, Q. Huang, X. Cao, and C. Hou. Co-saliency detection for RGBD images based on multi-constraint feature matching and cross label propagation. *IEEE Transactions on Image Processing*, 27(2):568–579, 2017.

[24] R. Cong, J. Lei, H. Fu, Q. Huang, X. Cao, and N. Ling. HSCS: Hierarchical sparsity based co-saliency detection for RGBD images. *IEEE Transactions on Multimedia*, 21(7):1660–1671, 2018.

[25] R. Cong, J. Lei, H. Fu, W. Lin, Q. Huang, X. Cao, and C. Hou. An iterative co-saliency framework for RGBD images. *IEEE Transactions on Cybernetics*, 49(1):233–246, 2017.

[26] R. Cong, J. Lei, C. Zhang, Q. Huang, X. Cao, and C. Hou. Saliency detection for stereoscopic images based on depth confidence analysis and multiple cues fusion. *IEEE Signal Processing Letters*, 23(6):819–823, 2016.

[27] A. Dai, C. Diller, and M. Nießner. Sg-nn: Sparse generative neural networks for self-supervised scene completion of rgb-d scans. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 849–858, 2020.

[28] K. Desingh, K. M. Krishna, D. Rajan, and C. Jawahar. Depth really matters: Improving visual salient region detection with depth. In *Proceedings of the British Machine Vision Conference*, 2013.

[29] Y. Ding, Z. Liu, M. Huang, R. Shi, and X. Wang. Depth-aware saliency detection using convolutional neural networks. *Journal of Visual Communication and Image Representation*, 61:1–9, 2019.

[30] H. Du, Z. Liu, and R. Shi. Salient object segmentation based on depth-aware image layering. *Multimedia Tools and Applications*, 78(9):12125–12138, 2019.

[31] H. Du, Z. Liu, H. Song, L. Mei, and Z. Xu. Improving RGBD saliency detection using progressive region classification and saliency fusion. *IEEE Access*, 4:8987–8994, 2016.

[32] D.-P. Fan, M.-M. Cheng, J.-J. Liu, S.-H. Gao, Q. Hou, and A. Borji. Salient objects in clutter: Bringing salient object detection to the foreground. In *Proceedings of the European Conference on Computer Vision*, pages 186–202. Springer, 2018.

[33] D.-P. Fan, M.-M. Cheng, Y. Liu, T. Li, and A. Borji. Structure-measure: A new way to evaluate foreground maps. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4548–4557, 2017.

[34] D.-P. Fan, C. Gong, Y. Cao, B. Ren, M.-M. Cheng, and A. Borji. Enhanced-alignment measure for binary foreground map evaluation. In *Proceedings of the International Joint Conferences on Artificial Intelligence*, pages 698–704, 2018.

[35] D.-P. Fan, G.-P. Ji, G. Sun, M.-M. Cheng, J. Shen, and L. Shao. Camouflaged object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2777–2787, 2020.

[36] D.-P. Fan, G.-P. Ji, T. Zhou, G. Chen, H. Fu, J. Shen, and L. Shao. Pranet: Parallel reverse attention network for polyp segmentation. In *Medical Image Computing and Computer-Assisted Intervention*, 2020.

[37] D.-P. Fan, T. Li, Z. Lin, G.-P. Ji, D. Zhang, M.-M. Cheng, H. Fu, and J. Shen. Re-thinking co-salient object detection. *arXiv preprint arXiv:2007.03380*, 2020.

[38] D.-P. Fan, Z. Lin, Z. Zhang, M. Zhu, and M.-M. Cheng. Rethinking RGB-D salient object detection: Models, data sets, and large-scale benchmarks. *IEEE Transactions on Neural Networks and Learning Systems*, 2020.

[39] D.-P. Fan, W. Wang, M.-M. Cheng, and J. Shen. Shifting more attention to video salient object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8554–8564, 2019.

[40] D.-P. Fan, Y. Zhai, A. Borji, J. Yang, and L. Shao. Bbs-net: Rgb-d salient object detection with a bifurcated backbone strategy network. In *Proceedings of the European Conference on Computer Vision*. Springer, 2020.

[41] D.-P. Fan, T. Zhou, G.-P. Ji, Y. Zhou, G. Chen, H. Fu, J. Shen, and L. Shao. Inf-net: Automatic covid-19

lung infection segmentation from ct images. *IEEE Transactions on Medical Imaging*, 2020.

[42] X. Fan, Z. Liu, and G. Sun. Salient region detection for stereoscopic images. In *Proceedings of the International Conference on Digital Signal Processing*, pages 454–458. IEEE, 2014.

[43] H.-S. Fang, J. Cao, Y.-W. Tai, and C. Lu. Pairwise body-part attention for recognizing human-object interactions. In *Proceedings of the European Conference on Computer Vision*, pages 51–67. Springer, 2018.

[44] D. Feng, N. Barnes, and S. You. Hoso: Histogram of surface orientation for rgb-d salient object detection. In *Proceedings of the International Conference on Digital Image Computing: Techniques and Applications*, pages 1–8. IEEE, 2017.

[45] D. Feng, N. Barnes, S. You, and C. McCarthy. Local background enclosure for RGB-D salient object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2343–2350, 2016.

[46] K. Fu, D.-P. Fan, G.-P. Ji, and Q. Zhao. Jl-dcf: Joint learning and densely-cooperative fusion framework for RGB-D salient object detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2020.

[47] C. Godard, O. Mac Aodha, and G. J. Brostow. Unsupervised monocular depth estimation with left-right consistency. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 270–279, 2017.

[48] J. Guo, T. Ren, and J. Bei. Salient object detection for RGB-D image via saliency evolution. In *Proceedings of the IEEE International Conference on Multimedia and Expo*, pages 1–6. IEEE, 2016.

[49] J. Guo, T. Ren, J. Bei, and Y. Zhu. Salient object detection in rgb-d image based on saliency fusion and propagation. In *Proceedings of the International Conference on Internet Multimedia Computing and Service*, pages 1–5, 2015.

[50] J. Han, H. Chen, N. Liu, C. Yan, and X. Li. Cnns-based RGB-D saliency detection via cross-view transfer and multiview fusion. *IEEE Transactions on Cybernetics*, 48(11):3171–3183, 2017.

[51] J. Han, D. Zhang, G. Cheng, N. Liu, and D. Xu. Advanced deep-learning techniques for salient and category-specific object detection: a survey. *IEEE Signal Processing Magazine*, 35(1):84–100, 2018.

[52] Y. He, J. Lin, Z. Liu, H. Wang, L.-J. Li, and S. Han. Amc: Automl for model compression and acceleration on mobile devices. In *Proceedings of the European Conference on Computer Vision*, pages 784–800. Springer, 2018.

[53] S. Hong, T. You, S. Kwak, and B. Han. Online tracking by learning discriminative saliency map with convolutional neural network. In *Proceedings of the International Conference on Machine Learning*, pages 597–606, 2015.

[54] X. Hou and L. Zhang. Saliency detection: A spectral residual approach. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2007.

[55] F. Huang, J. Qi, H. Lu, L. Zhang, and X. Ruan. Salient object detection via multiple instance learning. *IEEE Transactions on Image Processing*, 26(4):1911–1922, 2017.

[56] P. Huang, C.-H. Shen, and H.-F. Hsiao. Rgbd salient object detection using spatially coherent deep learning framework. In *Proceedings of the IEEE International Conference on Digital Signal Processing*, pages 1–5. IEEE, 2018.

[57] R. Huang, Y. Xing, and Z. Wang. RGB-D salient object detection by a CNN with multiple layers fusion. *IEEE Signal Processing Letters*, 26(4):552–556, 2019.

[58] R. Huang, Y. Xing, and Y. Zou. Triple-complementary network for RGB-D salient object detection. *IEEE Signal Processing Letters*, 2020.

[59] X. Huang and Y. Zhang. Water flow driven salient object detection at 180 fps. *Pattern Recognition*, 76:95–107, 2018.

[60] X. Huang and Y.-J. Zhang. 300-fps salient object detection via minimum directional contrast. *IEEE Transactions on Image Processing*, 26(9):4243–4254, 2017.

[61] Z. Huang, H.-X. Chen, T. Zhou, Y.-Z. Yang, and C.-Y. Wang. Multi-level cross-modal interaction network for rgb-d salient object detection. *arXiv preprint arXiv:2007.14352*, 2020.

[62] N. Imamoglu, W. Shimoda, C. Zhang, Y. Fang, A. Kanezaki, K. Yanai, and Y. Nishida. An integration of bottom-up and top-down salient cues on rgb-d data: saliency from objectness versus non-objectness. *Signal, Image and Video Processing*, 12(2):307–314, 2018.

[63] W. Ji, J. Li, M. Zhang, Y. Piao, and H. Lu. Accurate rgb-d salient object detection via collaborative

learning. In *ECCV*, 2020.

[64] B. Jiang, Z. Zhou, X. Wang, J. Tang, and B. Luo. cmsalgan: RGB-D salient object detection with cross-view generative adversarial networks. *IEEE Transactions on Multimedia*, 2020.

[65] L. Jiang, A. Koch, and A. Zell. Salient regions detection for indoor robots using rgb-d data. In *Proceedings of the IEEE International Conference on Robotics and Automation*, pages 1323–1328. IEEE, 2015.

[66] L. Jin, Y. Xu, J. Zheng, J. Zhang, R. Tang, S. Xu, J. Yu, and S. Gao. Geometric structure based and regularized depth estimation from 360 indoor imagery. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 889–898, 2020.

[67] Z. Jin, J. Li, and D. Li. Co-saliency detection for rgbd images based on effective propagation mechanism. *IEEE Access*, 7:141311–141318, 2019.

[68] R. Ju, L. Ge, W. Geng, T. Ren, and G. Wu. Depth saliency based on anisotropic center-surround difference. In *Proceedings of the IEEE International Conference on Image Processing*, pages 1115–1119. IEEE, 2014.

[69] K. Lai, L. Bo, X. Ren, and D. Fox. A large-scale hierarchical multi-view rgb-d object dataset. In *Proceedings of the IEEE International Conference on Robotics and Automation*, pages 1817–1824. IEEE, 2011.

[70] C. Lang, T. V. Nguyen, H. Katti, K. Yadati, M. Kankanhalli, and S. Yan. Depth matters: Influence of depth cues on visual saliency. In *Proceedings of the European Conference on Computer Vision*, pages 101–115. Springer, 2012.

[71] J. Lei, H. Zhang, L. You, C. Hou, and L. Wang. Evaluation and modeling of depth feature incorporated visual attention for salient object segmentation. *Neurocomputing*, 120:24–33, 2013.

[72] C. Li, R. Cong, S. Kwong, J. Hou, H. Fu, G. Zhu, D. Zhang, and Q. Huang. ASIF-Net: Attention steered interweave fusion network for RGB-D salient object detection. *IEEE Transactions on Cybernetics*, 2020.

[73] C. Li, R. Cong, Y. Piao, Q. Xu, and C. C. Loy. Rgb-d salient object detection with cross-modality modulation and selection. In *Proceedings of the European Conference on Computer Vision*. Springer, 2020.

[74] C. Li, G. Wang, Y. Ma, A. Zheng, B. Luo, and J. Tang. A unified rgb-t saliency detection benchmark: dataset, baselines, analysis and a novel approach. *arXiv preprint arXiv:1701.02829*, 2017.

[75] C. Li, B. Zhan, S. Zhang, and H. Sheng. Saliency detection with relative location measure in light field image. In *Proceedings of the International Conference on Image, Vision and Computing*, pages 8–12. IEEE, 2017.

[76] G. Li, Z. Liu, and H. Ling. Icnet: Information conversion network for RGB-D based salient object detection. *IEEE Transactions on Image Processing*, 29:4873–4884, 2020.

[77] G. Li, Z. Liu, L. Ye, Y. Wang, and H. Ling. Cross-modal weighting network for rgb-d salient object detection. In *Proceedings of the European Conference on Computer Vision*. Springer, 2020.

[78] G. Li and Y. Yu. Deep contrast learning for salient object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 478–487, 2016.

[79] N. Li, B. Sun, and J. Yu. A weighted sparse coding framework for saliency detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5216–5223, 2015.

[80] N. Li, J. Ye, Y. Ji, H. Ling, and J. Yu. Saliency detection on light field. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2806–2813, 2014.

[81] N. Li, J. Ye, Y. Ji, H. Ling, and J. Yu. Saliency detection on light field. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(8):1605–1616, 2017.

[82] X. Li, F. Yang, H. Cheng, W. Liu, and D. Shen. Contour knowledge transfer for salient object detection. In *Proceedings of the Proceedings of the European Conference on Computer Vision*. Springer, September 2018.

[83] F. Liang, L. Duan, W. Ma, Y. Qiao, Z. Cai, J. Miao, and Q. Ye. Cocnn: RGB-D deep fusion for stereoscopic salient object detection. *Pattern Recognition*, page 107329, 2020.

[84] F. Liang, L. Duan, W. Ma, Y. Qiao, Z. Cai, and L. Qing. Stereoscopic saliency model using contrast and depth-guided-background prior. *Neurocomputing*, 275:2227–2238, 2018.

[85] D. Liu, Y. Hu, K. Zhang, and Z. Chen. Two-stream refinement network for rgb-d saliency detection. In *Proceedings of IEEE International Conference on*

Image Processing, pages 3925–3929. IEEE, 2019.

[86] F. Liu, C. Shen, and G. Lin. Deep convolutional neural fields for depth estimation from a single image. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5162–5170, 2015.

[87] G. Liu and D. Fan. A model of visual attention for natural image retrieval. In *Proceedings of the IEEE Conference on Information Science and Cloud Computing Companion*, pages 728–733. IEEE, 2013.

[88] N. Liu, N. Zhang, and J. Han. Learning selective self-mutual attention for RGB-D saliency detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2020.

[89] Z. Liu, S. Shi, Q. Duan, W. Zhang, and P. Zhao. Salient object detection for RGB-D image by single stream recurrent convolution neural network. *Neurocomputing*, 363:46–57, 2019.

[90] Z. Liu, J. Tang, Q. Xiang, and P. Zhao. Salient object detection for rgb-d images by generative adversarial network. *Multimedia Tools and Applications*, pages 1–23, 2020.

[91] Z. Liu, W. Zhang, and P. Zhao. A cross-modal adaptive gated fusion generative adversarial network for RGB-D salient object detection. *Neurocomputing*, 2020.

[92] J. Lu, J. Yang, D. Batra, and D. Parikh. Hierarchical question-image co-attention for visual question answering. In *Proceedings of the International Conference on Neural Information Processing Systems*, pages 289–297, 2016.

[93] X. Lu, W. Wang, C. Ma, J. Shen, L. Shao, and F. Porikli. See more, know more: Unsupervised video object segmentation with co-attention siamese networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3623–3632, 2019.

[94] A. Luo, X. Li, F. Yang, Z. Jiao, H. Cheng, and S. Lyu. Cascade graph neural networks for rgb-d salient object detection. In *Proceedings of the Proceedings of the European Conference on Computer Vision*. Springer, 2020.

[95] C.-Y. Ma and H.-M. Hang. Learning-based saliency model with depth information. *Journal of vision*, 15(6):19–19, 2015.

[96] Y. Ma, D. Sun, Q. Meng, Z. Ding, and C. Li. Learning multiscale deep features and svm regressors for adaptive rgb-t saliency detection. In *Proceedings of the International Symposium on Computational Intelligence and Design*, volume 1, pages 389–392. IEEE, 2017.

[97] V. Mahadevan and N. Vasconcelos. Saliency-based discriminant tracking. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1007–1013. IEEE, 2009.

[98] N. Martinel, C. Micheloni, and G. L. Foresti. Kernelized saliency-based person re-identification through multiple metric learning. *IEEE Transactions on Image Processing*, 24(12):5645–5658, 2015.

[99] M. Mirza and S. Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014.

[100] T. V. Nguyen, Q. Zhao, and S. Yan. Attentive systems: A survey. *International Journal of Computer Vision*, 126(1):86–110, 2018.

[101] G.-Y. Nie, M.-M. Cheng, Y. Liu, Z. Liang, D.-P. Fan, Y. Liu, and Y. Wang. Multi-level context ultra-aggregation for stereo matching. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3283–3291, 2019.

[102] Y. Niu, Y. Geng, X. Li, and F. Liu. Leveraging stereopsis for saliency analysis. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 454–461. IEEE, 2012.

[103] J. Pan, C. C. Ferrer, K. McGuinness, et al. Salgan: Visual saliency prediction with generative adversarial networks. *arXiv preprint arXiv:1701.01081*, 2017.

[104] Y. Pang, L. Zhang, X. Zhao, and H. Lu. Hierarchical dynamic filtering network for RGB-D salient object detection. In *Proceedings of the European Conference on Computer Vision*. Springer, 2020.

[105] H. Peng, B. Li, W. Xiong, W. Hu, and R. Ji. Rgbd salient object detection: a benchmark and algorithms. In *Proceedings of the European Conference on Computer Vision*, pages 92–109. Springer, 2014.

[106] F. Perazzi, P. Krähenbühl, Y. Pritch, and A. Hornung. Saliency filters: Contrast based filtering for salient region detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 733–740. IEEE, 2012.

[107] Y. Piao, W. Ji, J. Li, M. Zhang, and H. Lu. Depth-induced multi-scale recurrent attention network for saliency detection. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 7254–7263, 2019.

[108] Y. Piao, X. Li, and M. Zhang. Depth-induced cellular

automata for light field saliency. In *Frontiers in Optics*, pages FTh3E–3. Optical Society of America, 2018.

[109] Y. Piao, X. Li, M. Zhang, J. Yu, and H. Lu. Saliency detection via depth-induced cellular automata on light field. *IEEE Transactions on Image Processing*, 29:1879–1889, 2020.

[110] Y. Piao, Z. Rong, M. Zhang, X. Li, and H. Lu. Deep light-field-driven saliency detection from a single view. In *Proceedings of the International Joint Conference on Artificial Intelligence*, 2019.

[111] Y. Piao, Z. Rong, M. Zhang, and H. Lu. Exploit and replace: An asymmetrical two-stream architecture for versatile light field saliency detection. In *Proceedings of the Association for the Advancement of Artificial Intelligence*, 2020.

[112] Y. Piao, Z. Rong, M. Zhang, W. Ren, and H. Lu. A2dele: Adaptive and attentive depth distiller for efficient RGB-D salient object detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2020.

[113] M. Qian, J. Qi, L. Zhang, M. Feng, and H. Lu. Language-aware weak supervision for salient object detection. *Pattern Recognition*, 96:106955, 2019.

[114] Y. Qin, H. Lu, Y. Xu, and H. Wang. Saliency detection via cellular automata. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 110–119, 2015.

[115] L. Qu, S. He, J. Zhang, J. Tian, Y. Tang, and Q. Yang. RGBD salient object detection via deep fusion. *IEEE Transactions on Image Processing*, 26(5):2274–2285, 2017.

[116] K. Rapantzikos, Y. Avrithis, and S. Kollias. Dense saliency-based spatiotemporal feature points for action recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1454–1461. IEEE, 2009.

[117] J. Ren, X. Gong, L. Yu, W. Zhou, and M. Ying Yang. Exploiting global priors for RGB-D saliency detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 25–32, 2015.

[118] H. Sheng, X. Liu, and S. Zhang. Saliency analysis based on depth contrast increased. In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 1347–1351. IEEE, 2016.

[119] H. Sheng, S. Zhang, X. Liu, and Z. Xiong. Relative location for light field saliency detection. In *Proceedings of the IEEE Conference on Acoustics, Speech and Signal Processing*, pages 1631–1635. IEEE, 2016.

[120] R. Shigematsu, D. Feng, S. You, and N. Barnes. Learning RGB-D salient object detection using background enclosure, depth contrast, and top-down features. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 2749–2757, 2017.

[121] W. Shimoda and K. Yanai. Distinct class-specific saliency maps for weakly supervised semantic segmentation. In *Proceedings of the European Conference on Computer Vision*, pages 218–234. Springer, 2016.

[122] H. Song, Z. Liu, H. Du, and G. Sun. Depth-aware saliency detection using discriminative saliency fusion. In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 1626–1630. IEEE, 2016.

[123] H. Song, Z. Liu, H. Du, G. Sun, O. Le Meur, and T. Ren. Depth-aware salient object detection and segmentation via multiscale discriminative saliency fusion and bootstrap learning. *IEEE Transactions on Image Processing*, 26(9):4204–4216, 2017.

[124] H. Song, W. Wang, S. Zhao, J. Shen, and K.-M. Lam. Pyramid dilated deeper convlstm for video salient object detection. In *Proceedings of the European Conference on Computer Vision*, pages 715–731. Springer, 2018.

[125] J. Su, J. Li, Y. Zhang, C. Xia, and Y. Tian. Selectivity or invariance: Boundary-aware salient object detection. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3799–3808, 2019.

[126] D. Sun, S. Li, Z. Ding, and B. Luo. Rgb-t saliency detection via robust graph learning and collaborative manifold ranking. In *Proceedings of the International Conference on Bio-Inspired Computing: Theories and Applications*, pages 670–684. Springer, 2019.

[127] Y. Tang, R. Tong, M. Tang, and Y. Zhang. Depth incorporating with color improves salient object detection. *The Visual Computer*, 32(1):111–121, 2016.

[128] W.-C. Tu, S. He, Q. Yang, and S.-Y. Chien. Real-time salient object detection with a minimum spanning tree. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 2334–2342, 2016.

[129] Z. Tu, Z. Li, C. Li, Y. Lang, and J. Tang. Multi-

interactive encoder-decoder network for rgbt salient object detection. *arXiv preprint arXiv:2005.02315*, 2020.

[130] Z. Tu, Y. Ma, Z. Li, C. Li, J. Xu, and Y. Liu. Rgbt salient object detection: A large-scale dataset and benchmark. *arXiv preprint arXiv:2007.03262*, 2020.

[131] Z. Tu, T. Xia, C. Li, Y. Lu, and J. Tang. M3s-nir: Multi-modal multi-scale noise-insensitive ranking for rgb-t saliency detection. In *Proceedings of the IEEE Conference on Multimedia Information Processing and Retrieval*, pages 141–146. IEEE, 2019.

[132] Z. Tu, T. Xia, C. Li, X. Wang, Y. Ma, and J. Tang. Rgb-t image saliency detection via collaborative graph learning. *IEEE Transactions on Multimedia*, 22(1):160–173, 2019.

[133] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin. Attention is all you need. In *Proceedings of the conference on Neural Information Processing Systems*, pages 5998–6008, 2017.

[134] A. Wang and M. Wang. RGB-D salient object detection via minimum barrier distance transform and saliency fusion. *IEEE Signal Processing Letters*, 24(5):663–667, 2017.

[135] A. Wang, M. Wang, X. Li, Z. Mi, and H. Zhou. A two-stage bayesian integration framework for salient object detection on light field. *Neural Processing Letters*, 46(3):1083–1094, 2017.

[136] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, and X. Tang. Residual attention network for image classification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3156–3164. Springer, 2017.

[137] G. Wang, C. Li, Y. Ma, A. Zheng, J. Tang, and B. Luo. Rgb-T saliency detection benchmark: Dataset, baselines, analysis and a novel approach. In *Proceedings of the Chinese Conference on Image and Graphics Technologies*, pages 359–369. Springer, 2018.

[138] L. Wang, J. Zhang, O. Wang, Z. Lin, and H. Lu. Sdc-depth: Semantic divide-and-conquer network for monocular depth estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 541–550, 2020.

[139] N. Wang and X. Gong. Adaptive fusion for RGB-D salient object detection. *IEEE Access*, 7:55277–55284, 2019.

[140] S. Wang, W. Liao, P. Surman, Z. Tu, Y. Zheng, and J. Yuan. Salience guided depth calibration for perceptually optimized compressive light field 3d display. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2031–2040, 2018.

[141] S.-T. Wang, Z. Zhou, H.-B. Qu, and B. Li. Rgb-d saliency detection under bayesian framework. In *Proceedings of International Conference on Pattern Recognition*, pages 1881–1886. IEEE, 2016.

[142] S.-T. Wang, Z. Zhou, H.-B. Qu, and B. Li. Visual saliency detection for RGB-D images with generative model. In *Proceedings of the Asian Conference on Computer Vision*, pages 20–35. Springer, 2016.

[143] T. Wang, A. Borji, L. Zhang, P. Zhang, and H. Lu. A stagewise refinement model for detecting salient objects in images. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4019–4028, 2017.

[144] T. Wang, Y. Piao, X. Li, L. Zhang, and H. Lu. Deep learning for light field saliency detection. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 8838–8848, 2019.

[145] W. Wang, Q. Lai, H. Fu, J. Shen, H. Ling, and R. Yang. Salient object detection in the deep learning era: An in-depth survey. *arXiv preprint arXiv:1904.09146*, 2019.

[146] W. Wang and J. Shen. Deep visual attention prediction. *IEEE Transactions on Image Processing*, 27(5):2368–2378, 2017.

[147] W. Wang, J. Shen, and L. Shao. Video salient object detection via fully convolutional networks. *IEEE Transactions on Image Processing*, 27(1):38–49, 2017.

[148] W. Wang, J. Shen, R. Yang, and F. Porikli. Saliency-aware video object segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(1):20–33, 2017.

[149] W. Wang, S. Zhao, J. Shen, S. C. Hoi, and A. Borji. Salient object detection with pyramid attention and salient edges. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1448–1457, 2019.

[150] X. Wang, Y. Dong, Q. Zhang, and Q. Wang. Region-based depth feature descriptor for saliency detection on light field. *Multimedia Tools and Applications*, 2020.

[151] X. Wang, S. Li, C. Chen, Y. Fang, A. Hao, and H. Qin. Data-level recombination and lightweight fusion scheme for rgb-d salient object detection. *IEEE Transactions on Image Processing*, 2020.

[152] X. Wang, S. Li, C. Chen, A. Hao, and H. Qin. Knowing depth quality in advance: A depth quality assessment method for rgb-d salient object detection. *arXiv preprint arXiv:2008.04157*, 2020.

[153] Y. Wang, Y. Li, J. H. Elder, H. Lu, and R. Wu. Synergistic saliency and depth prediction for RGB-D saliency detection. *arXiv preprint arXiv:2007.01711*, 2020.

[154] Y.-H. Wu, S.-H. Gao, J. Mei, J. Xu, D.-P. Fan, C.-W. Zhao, and M.-M. Cheng. Jcs: An explainable covid-19 diagnosis system by joint classification and segmentation. *arXiv preprint arXiv:2004.07054*, 2020.

[155] C. Xia, J. Li, X. Chen, A. Zheng, and Y. Zhang. What is and what is not a salient object? learning salient object detector by ensembling linear exemplar regressors. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4142–4150, 2017.

[156] F. Xiao, B. Li, Y. Peng, C. Cao, K. Hu, and X. Gao. Multi-modal weights sharing and hierarchical feature fusion for rgbd salient object detection. *IEEE Access*, 8:26602–26611, 2020.

[157] C. Xu, D. Tao, and C. Xu. Multi-view learning with incomplete views. *IEEE Transactions on Image Processing*, 24(12):5812–5825, 2015.

[158] H. Xue, Y. Gu, Y. Li, and J. Yang. Rgb-d saliency detection via mutual guided manifold ranking. In *Proceedings of IEEE International Conference on Image Processing*, pages 666–670. IEEE, 2015.

[159] P. Yan, G. Li, Y. Xie, Z. Li, C. Wang, T. Chen, and L. Lin. Semi-supervised video salient object detection using pseudo-labels. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 7284–7293, 2019.

[160] Q. Yan, L. Xu, J. Shi, and J. Jia. Hierarchical saliency detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1155–1162, 2013.

[161] C. Yang, L. Zhang, H. Lu, X. Ruan, and M.-H. Yang. Saliency detection via graph-based manifold ranking. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3166–3173, 2013.

[162] Z. Yu, J. Yu, Y. Cui, D. Tao, and Q. Tian. Deep modular co-attention networks for visual question answering. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6281–6290, 2019.

[163] Y. Zeng, Y. Zhuge, H. Lu, and L. Zhang. Joint learning of saliency detection and weakly supervised semantic segmentation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 7223–7233. Springer, 2019.

[164] Y. Zeng, Y. Zhuge, H. Lu, L. Zhang, M. Qian, and Y. Yu. Multi-source weak supervision for saliency detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6074–6083, 2019.

[165] D. Zhang, H. Fu, J. Han, A. Borji, and X. Li. A review of co-saliency detection algorithms: Fundamentals, applications, and challenges. *ACM Transactions on Intelligent Systems and Technology*, 9(4):1–31, 2018.

[166] D. Zhang, J. Han, and Y. Zhang. Supervision by fusion: Towards unsupervised learning of deep salient object detector. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4048–4056, 2017.

[167] D. Zhang, D. Meng, and J. Han. Co-saliency detection via a self-paced multiple-instance learning framework. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(5):865–878, 2016.

[168] D. Zhang, D. Meng, L. Zhao, and J. Han. Bridging saliency detection to weakly supervised object detection based on self-paced curriculum learning. *arXiv preprint arXiv:1703.01290*, 2017.

[169] H. Zhang, J. Lei, X. Fan, M. Wu, P. Zhang, and S. Bu. Depth combined saliency detection based on region contrast model. In *Proceedings of International Conference on Computer Science & Education*, pages 763–766. IEEE, 2012.

[170] J. Zhang, D.-P. Fan, Y. Dai, S. Anwar, F. S. Saleh, T. Zhang, and N. Barnes. Uc-net: uncertainty inspired rgb-d saliency detection via conditional variational autoencoders. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2020.

[171] J. Zhang, W. Li, P. Wang, P. Ogunbona, S. Liu, and C. Tang. A large scale rgb-d dataset for action recognition. In *Proceedings of the International Workshop on Understanding Human Activities through 3D Sensors*, pages 101–114. Springer, 2016.

[172] J. Zhang, Y. Liu, S. Zhang, R. Poppe, and M. Wang. Light field saliency detection with deep convolutional networks. *IEEE Transactions on Image Processing*, 29:4421–4434, 2020.

[173] J. Zhang, M. Wang, J. Gao, Y. Wang, X. Zhang, and X. Wu. Saliency detection with a deeper investigation of light field. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 2212–2218, 2015.

[174] J. Zhang, M. Wang, L. Lin, X. Yang, J. Gao, and Y. Rui. Saliency detection on light field: A multi-cue approach. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 13(3):1–22, 2017.

[175] M. Zhang, S. X. Fei, J. Liu, S. Xu, Y. Piao, and H. Lu. Asymmetric two-stream architecture for accurate rgb-d saliency detection. In *Proceedings of the European Conference on Computer Vision*. Springer, 2020.

[176] M. Zhang, W. Ji, Y. Piao, J. Li, Y. Zhang, S. Xu, and H. Lu. Lfnet: Light field fusion network for salient object detection. *IEEE Transactions on Image Processing*, 29:6276–6287, 2020.

[177] M. Zhang, J. Li, J. WEI, Y. Piao, and H. Lu. Memory-oriented decoder for light field salient object detection. In *Proceedings of the International Conference on Neural Information Processing Systems*, pages 896–906, 2019.

[178] M. Zhang, W. Ren, Y. Piao, Z. Rong, and H. Lu. Select, supplement and focus for RGB-D saliency detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2020.

[179] M. Zhang, Y. Zhang, Y. Piao, B. Hu, and H. Lu. Feature reintegration over differential treatment: A top-down and adaptive fusion network for rgb-d salient object detection. In *ACM Multimedia*, 2020.

[180] P. Zhang, D. Wang, H. Lu, H. Wang, and X. Ruan. Amulet: Aggregating multi-level convolutional features for salient object detection. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 202–211, 2017.

[181] P. Zhang, D. Wang, H. Lu, H. Wang, and B. Yin. Learning uncertain convolutional features for accurate saliency detection. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 212–221, 2017.

[182] Q. Zhang, N. Huang, L. Yao, D. Zhang, C. Shan, and J. Han. Rgb-t salient object detection via fusing multi-level cnn features. *IEEE Transactions on Image Processing*, 29:3321–3335, 2019.

[183] Z. Zhang, Z. Lin, J. Xu, W. Jin, S.-P. Lu, and D.-P. Fan. Bilateral attention network for rgb-d salient object detection. *arXiv preprint arXiv:2004.14582*, 2020.

[184] J. Zhao, Y. Zhao, J. Li, and X. Chen. Is depth really necessary for salient object detection. In *ACM Multimedia*, 2020.

[185] J.-X. Zhao, Y. Cao, D.-P. Fan, M.-M. Cheng, X.-Y. Li, and L. Zhang. Contrast prior and fluid pyramid integration for RGBD salient object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3927–3936, 2019.

[186] J.-X. Zhao, J.-J. Liu, D.-P. Fan, Y. Cao, J. Yang, and M.-M. Cheng. Egnet: Edge guidance network for salient object detection. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 8779–8788, 2019.

[187] R. Zhao, W. Oyang, and X. Wang. Person re-identification by saliency learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(2):356–370, 2016.

[188] T. Zhao and X. Wu. Pyramid feature attention network for saliency detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3085–3094, 2019.

[189] X. Zhao, L. Zhang, Y. Pang, H. Lu, and L. Zhang. A single stream network for robust and real-time rgb-d salient object detection. In *Proceedings of the European Conference on Computer Vision*. Springer, 2020.

[190] Z.-Q. Zhao, P. Zheng, S.-t. Xu, and X. Wu. Object detection with deep learning: A review. *IEEE Transactions on Neural Networks and Learning Systems*, 30(11):3212–3232, 2019.

[191] L. Zhou, Z. Yang, Q. Yuan, Z. Zhou, and D. Hu. Salient region detection via integrating diffusion-based compactness and local contrast. *IEEE Transactions on Image Processing*, 24(11):3308–3320, 2015.

[192] T. Zhou, H. Fu, G. Chen, J. Shen, and L. Shao. Hi-net: hybrid-fusion network for multi-modal MR image synthesis. *IEEE Transactions on Medical Imaging*, 39(9):2772–2781, 2020.

[193] T. Zhou, M. Liu, K.-H. Thung, and D. Shen. Latent representation learning for alzheimer' s disease diagnosis with incomplete multi-modality neuroimaging and genetic data. *IEEE Transactions on Medical Imaging*, 38(10):2411–2422, 2019.

[194] T. Zhou, K.-H. Thung, M. Liu, F. Shi, C. Zhang, and D. Shen. Multi-modal latent space inducing

ensemble svm classifier for early dementia diagnosis with neuroimaging data. *Medical Image Analysis*, 60:101630, 2020.

[195] T. Zhou, K.-H. Thung, X. Zhu, and D. Shen. Effective feature learning and fusion of multimodality data using stage-wise deep neural network for dementia diagnosis. *Human Brain Mapping*, 40(3):1001–1016, 2019.

[196] W. Zhou, Y. Chen, C. Liu, and L. Yu. GFNet: Gate fusion network with res2net for detecting salient objects in rgb-d images. *IEEE Signal Processing Letters*, 2020.

[197] W. Zhou, Y. Lv, J. Lei, and L. Yu. Global and local-contrast guides content-aware fusion for rgb-d saliency prediction. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2019.

[198] X. Zhou, G. Li, C. Gong, Z. Liu, and J. Zhang. Attention-guided RGBD saliency detection using appearance information. *Image and Vision Computing*, 95:103888, 2020.

[199] Y. Zhou, S. Huo, W. Xiang, C. Hou, and S.-Y. Kung. Semi-supervised salient object detection using a linear feedback control system model. *IEEE Transactions on Cybernetics*, 49(4):1173–1185, 2018.

[200] C. Zhu, X. Cai, K. Huang, T. H. Li, and G. Li. PDNet: Prior-model guided depth-enhanced network for salient object detection. In *Proceedings of the IEEE International Conference on Multimedia and Expo*, pages 199–204. IEEE, 2019.

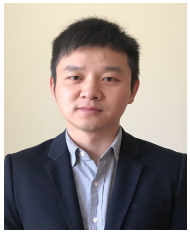
[201] C. Zhu and G. Li. A three-pathway psychobiological framework of salient object detection using stereoscopic technology. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 3008–3014, 2017.

[202] C. Zhu, G. Li, W. Wang, and R. Wang. An innovative salient object detection using center-dark channel prior. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 1509–1515, 2017.

[203] D. Zhu, L. Dai, Y. Luo, G. Zhang, X. Shao, L. Itti, and J. Lu. Multi-scale adversarial feature learning for saliency detection. *Symmetry*, 10(10):457, 2018.

[204] J.-Y. Zhu, J. Wu, Y. Xu, E. Chang, and Z. Tu. Unsupervised object class discovery via saliency-guided multiple class learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(4):862–875, 2014.

[205] L. Zhu, Z. Cao, Z. Fang, Y. Xiao, J. Wu, H. Deng, and J. Liu. Selective features for rgb-d saliency. In *Proceedings of Chinese Automation Congress*, pages 512–517. IEEE, 2015.



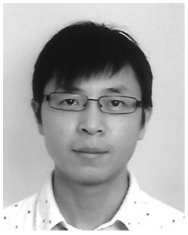
Tao Zhou received Ph.D. degree in Pattern Recognition and Intelligent System from the Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, in 2016. He is currently a Research Scientist at the Inception Institute of Artificial Intelligence (IIAI), Abu Dhabi, United Arab Emirates. His research interests include machine learning, computer vision, and medical image analysis.

Deng-Ping Fan received his Ph.D. degree from the Nankai University in 2019. He joined the Inception Institute of Artificial Intelligence (IIAI) in 2019. He has published about 20 top journal and conference papers such as CVPR, ICCV, *etc.* His research interests include computer vision, deep learning, and saliency detection, especially on co-salient object detection, RGB salient object detection, RGB-D salient object detection, and video salient object detection.

Ming-Ming Cheng (Senior Member, IEEE) received his Ph.D. degree from Tsinghua University in 2012. He then he did 2 years research fellow, with Prof. Philip Torr in Oxford. He is now a professor at Nankai University, leading the Media Computing Lab. His research interests include computer graphics, machine learning, computer vision, and image processing. He is an Associate Editor of IEEE TIP. He received several research awards, including the ACM China Rising Star Award, the IBM Global SUR Award, *etc.*

Jianbing Shen is currently acting as the Lead Scientist with the Inception Institute of Artificial Intelligence (IIAI), Abu Dhabi, United Arab Emirates. He is also a Full Professor with the School of Computer Science, Beijing Institute of Technology. He has published about

100 journal and conference papers such as IEEE TPAMI, CVPR, and ICCV. His research interests include computer vision and deep learning. He is an Associate Editor of IEEE TNNLS, IEEE TIP, *etc.*



Ling Shao is the CEO and Chief Scientist of the Inception Institute of Artificial Intelligence (IIAI), Abu Dhabi, United Arab Emirates. His research interests include computer vision, machine learning and medical imaging. He is an associate editor of IEEE TRANSACTIONS ON IMAGE PROCESSING, IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, and several other journals. He is a fellow of the International Association of Pattern Recognition, the Institution of Engineering and Technology and the British Computer Society.