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

Review Article

光场显著物体检测：综述与评测

傅可人 [1], 蒋遥 [1], 季葛鹏 [2], 周涛 [3] (✉), 赵启军 [1], 范登平 [4]

Abstract 显著物体检测（SOD）是计算机视觉领域的一个长期研究课题，其在过去十年中受到越来越多的关注。由于光场记录了在多方面有益于 SOD 的自然场景的综合信息，所以使用光场输入改善传统的基于 RGB 输入的显著性检测成为新兴的趋势。本文提供了首个光场 SOD 的综合性总结和评测，填补了显著性领域的长期空缺。首先，本文介绍了光场的理论以及数据形式，并对现有的光场 SOD 研究工作进行回顾，包括十个传统模型，七个基于深度学习的模型，一项对比研究以及一项简要总结。同时还总结了现有光场 SOD 数据集。其次，我们在四个广泛使用的光场数据集上对九个代表性光场 SOD 模型和几个前沿的 RGB-D SOD 模型进行了评测，并提供了深入讨论和分析，包括光场 SOD 和 RGB-D SOD 模型之间的比较。由于现有数据集的不一致，我们进一步生成完整的数据，为其补充焦点堆栈、深度图和多视角图像，使其一致和统一。我们补充的数据使得通用基准成为可能。最后，光场 SOD 因为其数据表示形式多样，同时高度依赖于采集数据的硬件而成为一个专门的问题，因此它与其他显著性检测任务差异巨大。我们提供了九个关于挑战和未来方向的建议，并概述了几个尚未解决的问题。本文所涉及的包括模型、数据集、评测结果和补充光场数据集在内的所有材料都已在https://github.com/kerenfu/LFSOD-Survey上公开。

Keywords 光场，显著物体检测，深度学习，评测.

1 前言

在 2021 年的 Google I/O 上，谷歌介绍了名为 Project Starline（https://blog.google/technology/research/project-starline/）的新技术，该技术结合了专门的硬件和计算机视觉技术，创造了一个可以远程连接两个人的“魔窗”（magic window），使用户体验到面对面交谈的感觉。这项沉浸式技术受益于光场显示技术，并且不需要如眼镜、耳机之类的额外设备。其中涉及的三项关键技术，包括 3D 成像、实时数据压缩和基于光场的 3D 显示，都是十分具有挑战性的工作，不过从谷歌来看已经取得了一定的突破。光场显著物体检测（SOD）可能同样有益于这三个阶段 [76]。显著物体检测（SOD） [5, 6, 15] 是计算机视觉中的一项基本任务，旨在检测和分割场景中的显著区域或物体，而光场 SOD [42, 43] 研究如何利用光场数据实现 SOD。SOD 有着广泛的应用，如物体检测和识别 [14, 48, 61, 62, 91]，语义分割 [82–84]，无监督视频物体分割 [66, 80]，多媒体压缩 [31, 46, 47, 50]，非真实性渲染 [29]，重定位 [68] 以及

1 四川大学计算机学院、四川大学视觉合成图形图像技术重点实验室。Email: fkrsuper@scu.edu.cn, yaojiangyj@foxmail.com, qjzhao@scu.edu.cn.

2 武汉大学计算机学院。Email: gepengai.ji@gmail.com

3 南京理工大学计算机科学与工程学院 PCA 实验室。Email: taozhou.ai@gmail.com.

4 南开大学计算机学院。Email: dengpfan@gmail.com.

Manuscript received: 2021-07-23; accepted: 2021-10-03.

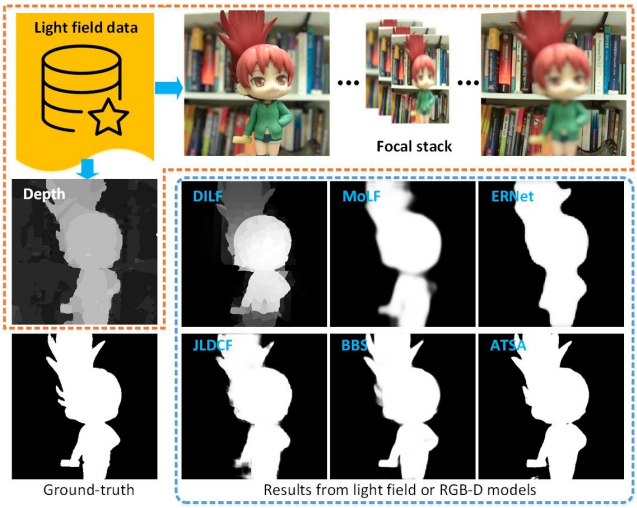

Fig. 1 在一个示例场景上用三种光场 SOD 模型：DILF [94]，MoLF [98] 和 ERNet [57] 以及三种前沿 RGB-D SOD 模型：JLDCF [25, 26]，BBS [22] 和 ATSA [96] 进行显著物体检测的结果。

人机交互 [8, 67]。一般而言，和单处理彩色图像的传统 SOD 相比，光场内丰富的线索和信息有助于算法更好地识别目标物体并提高 SOD 性能 [24, 27, 79, 100, 101]。

光场 SOD 探索如何以光场数据作为输入来检测显著物体。在 3D 空间中，光场 [28] 捕获每个空间位置和每个方向的所有光线。因此，其可以被视为一系列由网格状陈列的相机获取的图像阵列。与普通相机获取的 RGB 图像或深度传感器获取的深度图相比，全光相机获取的光场数据记录了更全面、更完整的自然场景信息，包含如深度信息 [32, 49, 54, 58, 69, 70, 77]，聚焦线索 [42, 49] 以及角度变化 [49, 56] 等。因此，光场数据可以从多方面有益于 SOD。首先，光场在获取后可重新聚焦 [49]。这可以生成一系列聚焦于不同深度的图像 [33]，该图像可提供有益于 SOD 的聚焦线索。其次，光场可以提供一系列来自于同一场景不同视角的图像。该图像包含丰富的空间视差和几何信息。最后，如 [32, 69, 70, 77] 所述，场景的深度信息嵌入在光场数据中，并可以通过不同的方式从焦点堆栈或多视角图像中估计得到。从这个意义上说，RGB-D 数据可以看作是光场数据的一种特殊退化情况。图1分别展示了在光场数据（焦点堆栈）上使用光场 SOD 方法以及在深度数据上使用 RGB-D SOD 模型获得的示例结果。虽然光场数据给 SOD 带来了巨大益处，并且该任务早在 2014 年 [42] 已被提出，但它仍然缺乏探索。具体而言，与 RGB SOD 或 RGB-D SOD 相比，光场 SOD 的研究较少。尽管文献稀少，但现有模型在技术框架和使用的光场数据集方面差异巨大。然而，据我们所知，目前还没有针对光场 SOD 的完整综述或评测。虽然 Zhang 等人 [103] 在 2015 年进行了一项对比研究，但其只将 Li 等人 [42] 提出的经典光场 SOD 模型与一组 2D 显著性模型进行比较，以证明结合光场知识的有效性。此外，该评估是在仅包含 100 张光场图像的 LFSD 数据集上进行。最近，Zhou 等人 [106] 简要总结了现有的光场 SOD 模型和相关数据集。然而，他们的工作主要集中在基于 RGB-D 的 SOD，仅一小部分用于总结光场 SOD，这导致了对模型细节和相关数据集的总结不足。此外，他们没有对光场 SOD 模型进行评测或提供任何性能评估。因此，我们认为缺乏一个现有模型和数据集的完整综述可能会阻碍该领域的进一步研究。

为此，在本文中我们提出了首个光场 *SOD* 的全面综述和评测。我们回顾了早期关于光场 SOD 的研究，包括十个传统模型 [41, 42, 55, 64, 73, 74, 76, 81, 94, 95]，七个基于深度学习的模型 [56, 57, 78, 93, 97, 98, 102]，一项对比研究 [103] 和一篇简要的综述 [106]。此外，我们还回顾了现有的光场 SOD 数据集 [42, 56, 78, 93, 95]，并对其进行统计分析，包括物体尺寸、物体与图像中心的距离、焦点切片数和物体的数量。由于数据集的不一致性（例如，部分数据集不提供焦点堆栈，而部分数据集缺少深度图或多视角图像），我们进一步生成和完善数据，包括焦点堆栈、深度图和多视角图像，从而使其一致和统一。此外，我们对可提供结果/代码的九个光场 SOD 模型 [41, 42, 56, 57, 81, 93, 94, 97, 98] 以及多个前沿的 RGB-D SOD 模型 [20, 22, 25, 38, 44, 51, 92, 96, 99] 进行评测，讨论两类模型之间的联系，并提供对挑战和未来方向的意见。本文涉及的所有材料，包括收集的模型、评测数据集和结果、补充光场数据和源代码链接，都在https://github.com/kerenfu/LFSOD-Surve

y上公开。本文的主要贡献旨在鼓励该领域的未来研究，包括如下四个方面：

- 首个关于光场 SOD 的系统综述，包括模型和数据集。该综述在该领域长期缺乏。
- 分析不同数据集的属性。由于一些数据集缺乏某些形式的数据，例如焦点堆栈或多视角图像，我们从现有数据集中生成更多数据作为补充，使它们完整和统一。
- 使用这些补充数据对九个光场 SOD 模型和几个前沿的 RGB-D SOD 模型进行了评测，并进行了综合深刻的讨论。
- 一项对光场 SOD 几个挑战的调查，以及它与其他问题关系的讨论，同时还有未来研究方向的讨论。

本文安排如下。第 Sec. 2章介绍了光场，回顾了光场 SOD 的现有模型和数据集，进行了相关的讨论和分析。第 Sec. 3章介绍了评估指标以及评测结果。第 Sec. 4章讨论了未来的研究方向，并概述了几个未解决的问题。第 Sec. 5章进行总结。

2 相关知识，模型和数据集

在本节中，我们首先简要介绍光场的理论、数据形式以及如何被用于 SOD。然后我们回顾之前关于光场 SOD 的工作，大致将它们分为传统模型和基于深度学习的模型。最后，我们总结了光场 SOD 的数据集并回顾了它们的详细信息。

2.1 光场

2.1.1 光场和光场相机

光场 [28] 由经过 3D 空间每个点和每个方向的所有光线组成。1991 年，Adelson 和 Bergen [2] 提出了全光函数 $P(\theta, \phi, \lambda, t, x, y, z)$ 来描述时间 t 时任意方向 (θ, ϕ) 和任意点 (x, y, z) 的波长 λ。在成像系统中，波长和时间可以用 RGB 通道和不同的帧表示，并且光通常沿着特定的路径传播。因此，Levoy 和 Hanrahan [37] 提出两平面参数化全光函数来表示成像系统中的光场。全光函数的两平面参数化 (L) 如图 2 (b) 所示，它可以表示为 $L(u, v, x, y)$。在该方案中，光场中的每条光线由表示

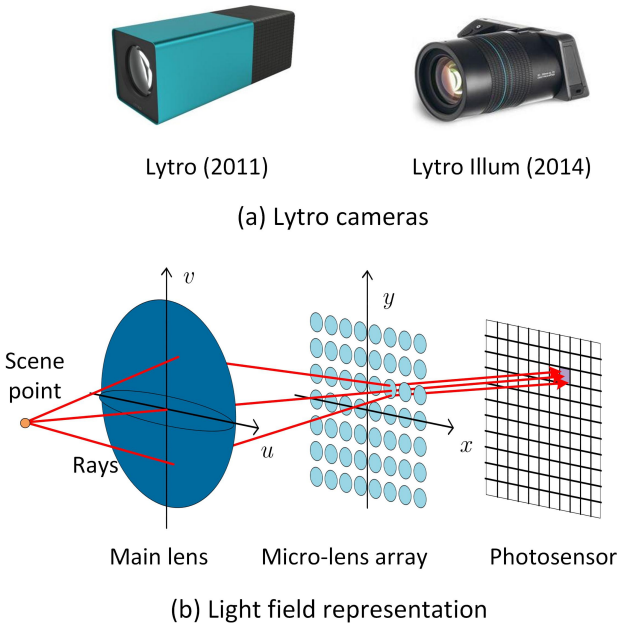

Fig. 2 Lytro 相机（a）及光场表示（b）。

空间 (x, y) 和角度 (u, v) 信息的两个平行平面决定。基于这一理论发明了可以捕获光场的设备，即图 2 (a) 所示的 Lytro 相机。这种相机包含主镜头和置于光电传感器前的微透镜阵列，其中前者作为 u-v 平面记录光线的角度信息，而后者作为 x-y 平面记录空间信息。图 2(b) 图像化地展示了光场的两平面参数化表示。根据上述四维参数，该数据通常被称为 4D 光场数据 [41–43, 55–57, 64, 73, 74, 76, 78, 81, 93–95, 97, 98, 102, 103]。

2.1.2 光场数据的格式

目前为止，所有公共光场 SOD 数据集都由 Lytro 相机拍摄，其原始数据为 LFP/LFR 文件（前者由 Lytro 获取，后者由 Lytro Illum 获取）。现有的光场数据集中的所有图像都是通过使用 Lytro Desktop 软件http://lightfield-forum.com/lytro/lytro-archive/、LFToolboxhttp://code.behnam.es/python-lfp-reader/或者 https://ww2.mathworks.cn/matlabcentral/fileexchange/75250-light-field-toolbox处理 LFP 或 LFR 文件生成。由于原始数据难以利用，现有 SOD 模型所利用的光场数据形式也是多种多样的，包括焦点堆栈和全聚焦图像 [41, 42, 55–

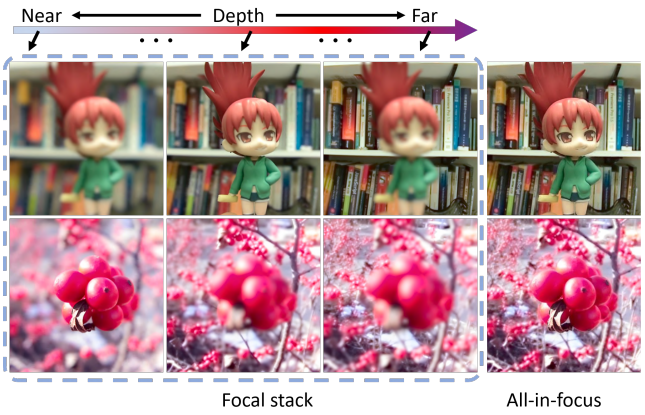

Fig. 3 焦点堆栈和全聚焦图像。

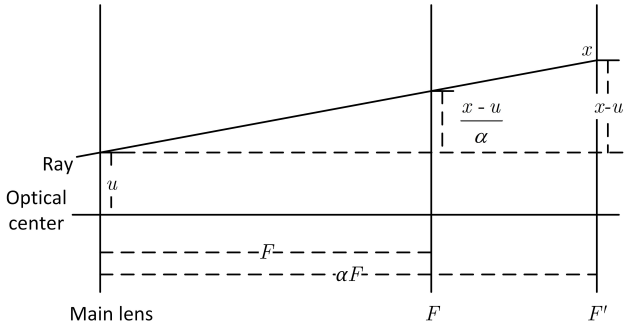

Fig. 4 重聚焦原理图。

57, 73, 76, 81, 94, 95, 97, 98]，多视角图像和中心视角图像 [56, 95, 102]，以及微透镜图像阵列 [64, 93]。如上所述，深度图像也可以从光场数据合成 [32, 69, 70, 77]，因此它们可以为 RGB-D SOD 模型形成 RGB-D 数据源（如图 1所示）。图 3展示了焦点堆栈和全聚焦图像，图 5展示了多视角图像，中心视角图像和深度图像。

具体来说，焦点堆栈（如图 Fig. 3左三列）包含一系列聚焦于不同深度的图像。该图像通过使用数字重聚焦技术处理原始光场数据生成 [49]。重聚焦原理如图 4所示，该图仅展示 u 和 x 维度。假设光线由位置 u 处进入主透镜，成像平面的位置由 F（主镜头的焦距）变为 F'，其中 $F' = \alpha F$。可按如下方式计算重聚焦图像。首先，给定 4D 光场 L_F，新成像平面 F' 处的新光场 L_α 可以表示为

$$L_\alpha(u, v, x, y) = L_F(u, v, u + \frac{x - u}{\alpha}, v + \frac{y - v}{\alpha}). \quad (1)$$

获得新的光场 $L_\alpha(u, v, x, y)$ 后，成像平面上的重聚焦图像可以合成为

$$I_\alpha(x, y) = \iint L_\alpha(u, v, x, y) du, dv. \quad (2)$$

通过改变参数 α，可以生成一系列重聚焦图像并组成一个焦点堆栈。得到焦点堆栈后，可以通过 photomontage [3] 生成全聚焦图像。例如，可以通过集合所有清晰像素来生成全聚焦图像，其中像素点的清晰度可以通过梯度来估计。全聚焦图像同样可通过计算所有焦点切片的加权平均值来获得。更多细节请见 [35]。除了焦点堆栈，多视角图像（图5）也可从光场数据中导出。如上所述，在 4D 光场的 $L_F(u, v, x, y)$ 表示中，(u, v) 对入射光线的角度信息进行编码。因此，可以通过在特定角度方向 (u^*, v^*) 进行采样以生成来自某个视点的图像，该图像可用 $L_F(u^*, v^*, x, y)$ 表示。通过改变 (u^*, v^*)，可合成多视角图像。特别的，当角度方向 (u^*, v^*) 等于中心视角的角度方向（即 (u_0, v_0)）时，此图为中心视角图像。此外，可以通过对 (x, y) 维度进行采样来生成微透镜图像。给定微透镜位置 (x^*, y^*)，可生成微透镜图像 $L_F(u, v, x^*, y^*)$，其可以捕捉场景点的多个视角图像。通过改变 (x^*, y^*)，可以获得不同的微透镜阵列，它们共同组成了一个表达完整光场信息的微透镜图像阵列。微透镜和多视角图像如图 [93] 所示。

此外，还可以从光场中估计包含场景深度信息的深度图。深度信息嵌入在聚焦和角度线索中，因此可以通过结合它们来生成深度图 [32, 54, 58, 69, 70, 77]。

2.2 光场 SOD 模型及综述

本章对现有光场 SOD 模型进行回顾与讨论，包括十个使用人工设计特征的传统方法和七个基于深度学习的方法。此外，本章还回顾了一项比较研究和一项简要综述工作。所有方法的详细信息总结于表 1中。

2.2.1 传统模型

如表 1所述，传统光场 SOD 模型通常将传统显著性检测中广泛采用的各种手工特征/假设（如全局/局部颜色对比度、背景先验以及物体位置线索） [6] 扩展到光场数据，此外，部分定制的光场特性，如聚焦度、深度和光场流，也被纳入其中。传统光场 SOD 模型倾向于

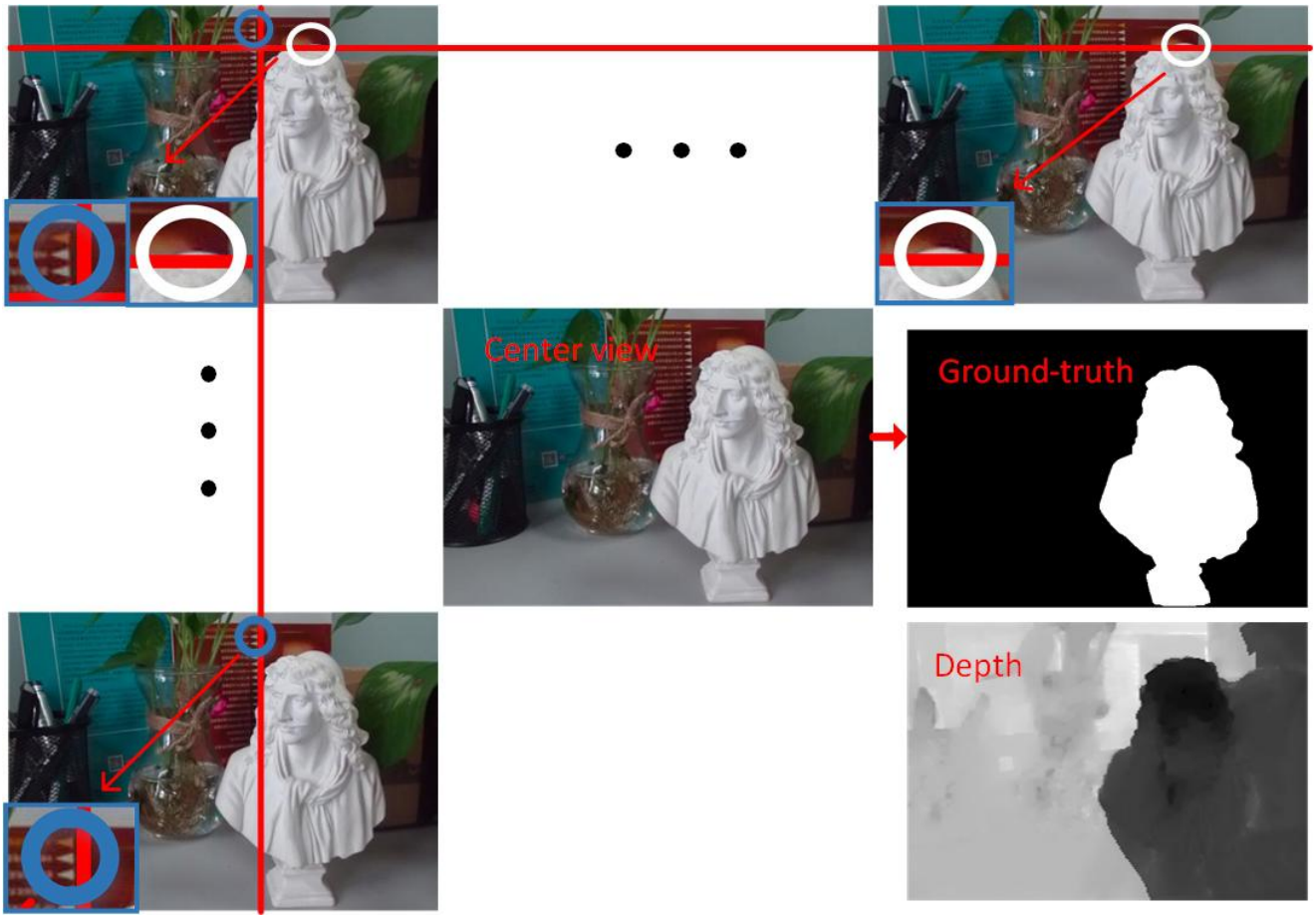

Fig. 5 多视角图像（包括中心视角图像），深度图像和真值图。注意多视角图像展现的不明显的视差（差异）（在每个多视角图像的左下角特写展示）。

采用后细化步骤，例如，优化框架 [55, 64, 76, 94, 95] 或 CRF [55]，以获得具有更好空间一致性和精确目标边界的显著图。在数据形式的利用方面，几乎所有传统模型都使用焦点堆栈，而深度图仅被其中少数模型采用。在所有传统方法中，仅有两种方法使用多视角 [95] 和微透镜数据 [64]。此外，由于早期数据集稀缺，大部分传统模型均只在 [42] 构建的小型 LFSD 数据集上进行评估。尽管传统模型取得了早期进展，但受手工设计特征的局限性影响，与现代深度学习模型相比，其很难泛化到具有挑战性和复杂的场景。后续章节将简要回顾未分类的传统模型的关键特性，虽然其采用了重叠的特征，但所采取的计算技术非常多样。

LFS [42] 是光场 SOD 的开创性和最早的工作，该工作还提出了首个光场 SOD 数据集。LFS 首先将聚焦度度量与位置先验相结合，以确定背景及前景切片。之后，其在全聚焦图像上计算背景先验及对比度线索以检测显著性候选。最后，LFS 使用对象性线索将全聚焦图像中的显著性候选与前景切片中的显著性候选进行加权合并，生成显著性图。该工作的扩展版发表于 [43]。

WSC [41] 是可处理 2D、3D 以及光场 SOD 间异构数据的统一 SOD 框架。基于加权稀疏编码框架，作者首先使用非显著性字典重建参考图像，该过程中重建误差较大的分块被选为显著性字典。该显著性字典随后通过迭代运行加权稀疏框架进行优化，以获得最终的显著性图。对于光场数据，用于字典构建的特征来自全聚焦 RGB 图像、深度图和焦点堆栈。

DILF [94] 从全聚焦图像和深度图中计算深度诱导的对比度显著性和颜色对比度显著性，并将其用于生成对比度显著性图。进一步，DILF 基于焦点堆栈中的聚焦度度量计算背景先验，并将其作为权重，以消除背景干

Tab. 1 光场 SOD 模型概述和总结工作。FS= 焦点堆栈，DE= 深度图，MV= 多视角图像，ML= 微透镜图像，OP= 开源。FS、DE、MV 和 ML 表示输入模型的数据类型。新数据集在 Main components 下用粗体显示。

	Model	Pub.	Year	Training dataset(s)	Testing dataset(s)	Main components	FS	DE	MV	ML	OP
Traditional models	LFS [42]	CVPR	2014	-	LFSD	Focusness measure, location priors, contrast cues, background prior, **new dataset (LFSD)**	✓				✓
	WSC [41]	CVPR	2015	-	LFSD	Weighted sparse coding, saliency/non-saliency dictionary construction	✓	✓			✓
	DILF [94]	IJCAI	2015	-	LFSD	Depth-induced/Color contrast, background priors by focusness	✓	✓			✓
	RL [64]	ICASSP	2016	-	LFSD	Relative locations, guided filter, micro-lens images				✓	
	BIF [73]	NPL	2017	-	LFSD	Bayesian framework, boundary prior, color/depth-induced contrast	✓	✓			
	LFS [43]	TPAMI	2017	-	LFSD	An extension of [42]	✓				✓
	MA [95]	TOMM	2017	-	LFSD + HFUT-Lytro	Superpixels intra-cue distinctiveness, light-field flow, **new dataset (HFUT-Lytro)**	✓	✓	✓		
	SDDF [74]	MTAP	2018	-	LFSD	Background priors, gradient operator, color contrast, local binary pattern histograms	✓				
	SGDC [76]	CVPR	2018	-	LFSD	Focusness cues, color and depth contrast	✓	✓			
	RDFD [81]	MTAP	2020	-	LFSD	Region-based depth feature descriptor, dark channel prior, multi-layer cellular automata	✓				
	DCA [55]	TIP	2020	-	LFSD	Depth-induced cellular automata, object-guided depth	✓	✓			
Deep learning models	DLLF [78]	ICCV	2019	DUTLF-FS	LFSD + DUTLF-FS	VGG-19, attention subnetwork, ConvLSTM, adversarial examples, **new dataset (DUTLF-FS)**	✓				
	DLSD [56]	IJCAI	2019	DUTLF-MV	DUTLF-MV	View synthesis network, multi-view detection/attention, VGG-19, **new dataset (DUTLF-MV)**			✓		✓
	MoLF [98]	NIPS	2019	DUTLF-FS	HFUT-Lytro + LFSD + DUTLF-FS	VGG-19, memory-oriented spatial fusion, memory-oriented feature integration	✓				✓
	ERNet [57]	AAAI	2020	DUTLF-FS + HFUT-Lytro	HFUT-Lytro + LFSD + DUTLF-FS	VGG-19, ResNet-18, multi-focusness recruiting/screening modules, distillation	✓				✓
	LFNet [97]	TIP	2020	DUTLF-FS	HFUT-Lytro + LFSD + DUTLF-FS	VGG-19, refine unit, attention block, ConvLSTM	✓				
	MAC [93]	TIP	2020	Lytro Illum	Lytro Illum + LFSD + HFUT-Lytro	Micro-lens images/image arrays, DeepLab-v2, model angular changes, **new dataset (Lytro Illum)**				✓	✓
	MTCNet [102]	TCSVT	2020	Lytro Illum	Lytro Illum + HFUT-Lytro	Edge detection, depth inference, feature-enhanced salient object generator			✓		
Reviews	CS [103]	NEURO	2015	-	LFSD	Comparative study between 2D *vs.* light field saliency					
	RGBDS [106]	CVM	2020	-	-	In-depth RGB-D SOD survey, brief review of light field SOD					

扰并增强显著性估计。

RL [64] 提出使用滤波处理估计场景点的相对位置。之后，该相对位置（可被视场景深度信息的另一种表示）与 [107] 中提出的鲁棒背景检测和显著性优化框架相结合，以增强显著性检测。

BIF [73] 使用贝叶斯框架融合从 RGB 图像、深度图和焦点堆栈中提取的多种特征。受 RGB SOD 方法的启发，该模型利用边界连通性先验、背景似然分数和颜色对比度生成背景概率图、前景切片、基于颜色的显著性图和深度诱导的对比度图，并通过两阶段贝叶斯方案进行融合。

MA [95] 通过计算两个超像素之间的内部线索差异来测量超像素的显著性，该过程中考虑的特征包括从不同焦平面和多个视点继承的颜色、深度和流特征。光场流 (从焦点堆栈和多视角序列中估计) 被该方法首次采用以便捕获深度不连续性/对比度。显著性度量随后使用位置先验和基于随机搜索的加权策略进行增强。此外，作者提出了一个新的光场 SOD 数据集，这是当时

最大的数据集。

SDDF [74] 利用嵌入在焦点堆栈中的深度信息进行精确的显著性检测。其首先通过对焦点堆栈图像应用梯度算子获得背景测量值，并选择测量值最高的切片作为背景层。SDDF 通过使用提取的背景区域分离全聚焦图像中的背景和前景来生成粗略预测，并通过全局计算粗略显著图的颜色和纹理（局部二值模式直方图）对比度生成最终显著图。

SGDC [76] 提出了一种用于优化多层光场显示的对比度增强显著性检测方法。它首先计算每个重聚焦图像的超像素级聚焦图，然后选择具有最高背景似然分数的重聚焦图像来获得背景线索。这种聚焦背景线索之后与颜色和深度对比显著性结合起来。最终结果通过 [107] 中提出的优化框架进行优化。

RDFD [81] 通过多线索集成框架解决光场 SOD 问题。基于暗通道先验 [30] 可用于估计离焦/模糊程度，作者提出了一种定义在焦点堆栈上的基于区域的深度特征描述符（RDFD）。RDFD 通过整合所有焦点堆栈图像的离焦度生成，从而减弱了需要精确深度图的限制。RDFD 特征用于计算基于区域的深度对比图和三维空间分布先验。该方法使用多层细胞自动机（MCA）将这些线索合并到一张图中，生成最终的显著图。

DCA [55] 提出了一种用于光场 SOD 的深度诱导细胞自动机（DCA）。首先，其利用聚焦度和深度线索计算目标诱导的深度图并选择背景种子。并基于该种子计算对比度显著图并与目标诱导的深度图相乘，以获得深度诱导显著图，该深度诱导显著图随后通过 DCA 进行优化。最后，将优化后的显著图与深度诱导显著性图相结合。贝叶斯融合策略和 CRF 被用于细化预测结果。

2.2.2 基于深度学习的模型

由于深度神经网络具有强大的学习能力，基于深度学习的模型可以实现优于传统光场 SOD 模型的精度和性能。与后者相比，深度模型的另一优点是可直接从大量数据中学习，而无需设计人工特征。如表 1所述，在深度学习时代，由于引入了三个新的数据集以更好地训练深度神经网络，数据集的稀缺性有所缓解。同样地，大多数深度模型仍然将焦点堆栈作为网络输入。由于焦点堆栈的多变量特性，注意机制 [56, 57, 78, 97, 98] 和 ConvLSTMs [57, 78, 97, 98] 等模块被多数方法采用。对于深层模型，可能存在不同的分类方法。一种简单的方法是根据所使用的光场数据类型进行分类，如表 1所示。有四种模型（DLLF [78]，MoLF [98]，ERNet [57]，LFNet [97]）使用焦点堆栈，而 DLSD [56] 和 MTCNet [102] 利用多视角图像，此外，MAC [93] 探索微透镜图像。不同的输入数据形式往往导致不同的网络设计。值得注意的是，DLSD [56] 处理的多视角图像是从输入的单视角图像渲染得到的，因此，该方法适用于所有输入场景，无论多视角图像是否可得。

由于使用基于深度学习的技术成为光场 SOD 的主流趋势，在本文中，我们进一步将现有的深度模型按照其结构分为五类，包括**后期融合体系**、**中期融合体系**、**基于知识蒸馏的体系**、**基于重建的体系**和**单流体系**，如图 6所示。下面对各分类及其相关模型进行简要描述。

后期融合模型（图 6 (a), DLLF [78], MTCNet [102]）旨在从输入的焦点堆栈/多视角图像和全聚焦/中心视角图像中获得单独的预测，然后进行简单融合。由于其简单且易于实现，后期融合作为一种经典策略被之前的多模态检测工作（*e.g.*, RGB-D SOD [106], RGB-D 语义分割 [26, 63]）广泛采用。然而，融合过程被限制到最后一步，并且集成计算相对简单。

DLLF [78] 采用双流融合框架，分别探索焦点堆栈和全聚焦图像。在焦点堆栈流中，DLLF 首先通过全卷积网络从级联焦点切片中提取特征。之后，来自不同切片的各种特征被一个循环注意网络整合，该网络利用一个注意力子网络和 ConvLSTM [65] 自适应地加权特征并利用其空间相关性。生成的显著图最终与从全聚焦图像获得的显著图相结合。此外，作者引入了一个新的大型数据集以解决数据对深度网络训练的局限性。

MTCNet [102] 提出了一个由显著性感知的特征聚合模块（SAFA）和多视角启发的深度显著性特征提取模块（MVI-DSF）组成的双流多任务协作网络。该网络利用边缘检测、深度推断和显著物体检测的相关机

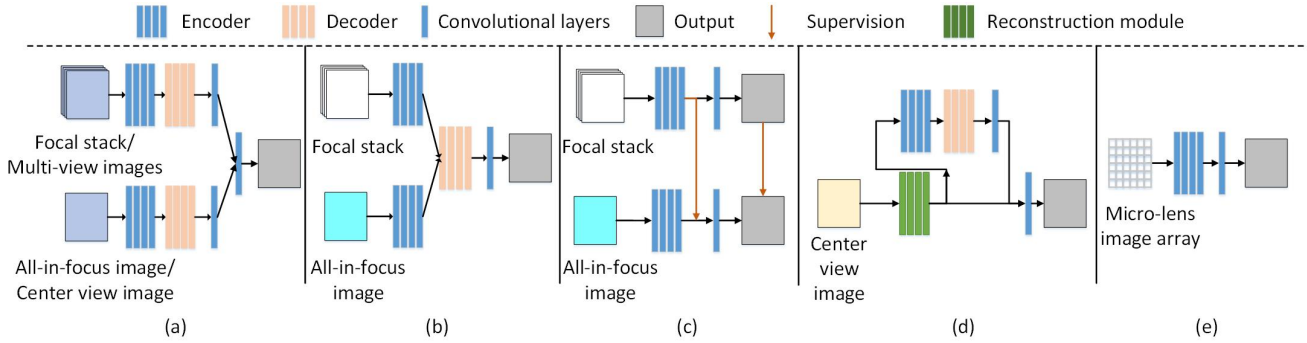

Fig. 6 现有基于深度学习的光场 SOD 模型体系。(a) 后期融合：DLLF [78]，MTCNet [102]。(b) 中期融合：MoLF [98]，LFNet [97]。(c) 基于知识蒸馏：ERNet [57]。(d) 基于重构：DLSD [56]。(e) 单流：MAC [93]。其中，(a) 使用焦点堆栈/多视角图像和全聚焦/中心视角图像，(b,c) 使用焦点堆栈和全聚焦图像，(d,e) 使用中心视角图像和微透镜图像阵列。

制提取具有代表性的显著性特征。SAFA 同时从中心视角图像中提取焦平面、边缘和启发式显著性特征，而 MVI-DSF 从一组多视角图像中推断深度显著性特征。MTCNet 使用特征增强操作组合提取的特征，以获得最终的显著图。

中期融合策略（图 6 (b)，MoLF [98]，LFNet [97]）以双流网络方式从焦点堆栈和全聚焦图像中提取特征。之后，通过一个精心设计的复杂解码器完成中间特征的融合。与图 6(a) 中的后期融合策略相比，主要区别在于融合的特征通常是分层的和中间的，并且解码器也是一个相对较深的卷积网络，可挖掘更复杂的聚合规则。

MoLF [98] 采用了一个面向内存的解码器，该解码器由空间融合模块（Mo-SFM）和特征集成模块（Mo-FIM）组成，以效仿人类融合信息的记忆机制。Mo-FSM 利用注意力机制学习不同特征图的重要性，并利用 ConvLSTM [65] 逐步细化空间信息。在 Mo-FIM 中，场景上下文集成模块（SCIM）和 ConvLSTM 被用于学习通道注意图和总结空间信息。

LFNet [97] 提出了一种双流融合网络，用于细化互补信息，并整合聚焦切片中逐渐变化的聚焦度和模糊度。从全聚焦图像和焦点堆栈中提取的特征被馈送到光场细化模块（LFRM）和整合模块（LFIM）以生成最终的显著图。在 LFRM 中，从单个切片提取的特征被馈送到细化单元以学习残差特征。LFIM 利用注意力模块自适应地加权和聚集切片特征。

基于知识蒸馏模型（图 6 (c), ERNet [57]）尝试将教师网络从焦点堆栈中学到的聚焦知识传递到处理全聚焦图像的学生网络。该方法使用来自焦点堆栈流的特征和预测监督从全聚焦流获得的特征和预测，有效地提高了后者的性能。学生网络实际上是一个在训练期间通过额外的光场知识进行增强的 RGB SOD 网络。

ERNet [57] 由基于知识蒸馏的双流师生网络组成。教师网络使用多焦点吸收模块（MFRM）和多焦点筛选模块（MFSM）从焦点切片中吸收和提炼知识，而学生网络以单个 RGB 图像作为输入以提高计算效率，并被强制逼近来自教师网络的多焦点特征以及预测。

基于重建的方法（图6 (d)，DLSD [56]）侧重于不同的方面，即从单个输入图像重建光场数据/信息。这是另一个有趣的主题，因为光场具有各种数据形式（见第 2.1.2节）。在重建光场的帮助下，可采用具有中/后期融合策略的编码器-解码器架构来完成光场 SOD 任务。该方案类似于基于知识蒸馏方案中学生网络，即其本质上是一个在训练期间由额外的光场知识（该情况下，为学习重建光场数据）增强的 RGB SOD 网络。

DLSD [56] 将光场 SOD 视为两个子问题：从单视角图像合成光场和光场驱动的 SOD。该模型首先采用光场合成网络，该网络通过两个独立的卷积网络沿水平和垂直方向估计深度图。根据深度图，单视角图像被变换为光场的水平和垂直视角图片。该方法还设计了一个由多视角显著性检测子网络和多视角注意模块组成的用

于显著性预测的光场驱动的 SOD 网络。具体地，该模型使用光场（多视角数据）作为中间桥梁从二维单视角图像推断显著图。作者提出了一个包含多视角图像和中心视图像素级真值图的新数据集。

单流模型（图 6 (e)，MAC [93]）受光场可用单个图像（微透镜图像阵列 [93]）表示启发。因此，不同于图 6 (a-b)，该方案直接使用单个自底向上的流来处理微透镜图像阵列，而无需显式特征融合。

MAC [93] 是以微透镜图像阵列作为输入的用于光场 SOD 的端到端深度卷积网络。首先，它采用了一个 MAC（Model Angular Changes）块对单个局部微透镜图像中的角度变化进行建模，然后将提取的特征馈送给修改的 DeepLab-v2 网络 [11] 以捕获多尺度信息和长距离空间依赖。作者结合该模型提出了一个新的包含高质量微透镜图像阵列的 Lytro Illum 数据集。

2.2.3 其他综述

CS [103] 对光场显著性和 2D 显著性进行了比较研究，证明在光场数据上执行 SOD 任务优于单个二维图像。它在 LFSD 数据集 [42] 上比较了经典模型 LFS [42] 和八个 2D 显著性模型。五个评价指标上的结果表明，光场显著性模型比传统的 2D 模型具有更好的鲁棒性。

RGBDS [106] 对 RGB-D SOD 进行了深入全面的调查。其从不同的角度回顾了现有 RGB-D SOD 模型及相关的基准数据集。考虑到光场同样可以提供深度图，作者简要回顾了光场 SOD 模型和数据集。然而，由于该工作主要关注 RGB-D SOD，因此仅少量内容被用于介绍光场 SDO，同时也没有进行相关的评测。

2.3 光场 SOD 数据集

2.3.1 数据集

目前针对光场 SOD 任务存在五个数据集，包括 LFSD [42]，HFUT-Lytro [95]，DUTLF-FS [78]，DUTLF-MV [56] 和 Lytro Illum [93]。我们在表 2中总结了数据集的详细信息，并在图7中展示 4 个数据集（LFSD、HFUT-Lytro、Lytro Illum 和 DUTLF-FS）中样例。关于数据集的简要介绍如下：

LFSD [42]https://sites.duke.edu/nianyi/publication/saliency-detection-on-light-field/是首个用于 SOD 的光场数据集，其包含 60 个室内场景和 40 个室外场景。该数据集通过 Lytro 相机获取，并且每个光场提供了焦点堆栈、全聚焦图像、深度图像和相应的真值图。图像的空间分辨率是 360×360。原始光场数据同样可以在 LFSD 中获取。该数据集中的大多数图像都包含单个居中物体且具有相对简单的背景。

HFUT-Lytro [95] https://github.com/pencilzhang/MAC-light-field-saliency-net 包含 255 个室内和室外的光场。每个光场包含一组焦点切片数量从 1 到 12 不等的焦点堆栈。其角度分辨率为 7×7，空间分辨率为 328×328。该数据集提供了焦点堆栈、全聚焦图像、多视角图像和粗略深度图。同时，HFUT-Lytro 存在多个关于 SOD 的挑战，例如遮挡，杂乱背景和外观变化。

DUTLF-FS [78] https://github.com/OIPLab-DUT/ICCV2019_Deeplightfield_Saliency 是目前为止最大的光场 SOD 数据集，其共包含 1462 个光场。该数据集是由 Lytro Illum 相机在室内和室外场景中拍摄的。整个数据集被分为 1000 个训练样本和 462 个测试样本。它为不同的光场提供全聚焦图像、焦点堆栈和相应的真值图。焦点堆栈的焦点切片数量从 2 到 12 不等，图像的空间分辨率为 600×400。值得注意的是，DUTLF-FS 具有各种挑战，包括不同类型的物体、显著物体与背景间的低对比度以及变化的物体位置。

DUTLF-MV [56] https://github.com/OIPLab-DUT/IJCAI2019-Deep-Light-Field-Driven-Saliency-Detection-from-A-Single-View 是另一个用于 SOD 的大规模光场数据集，它是从与 DUTLF-FS 相同的数据库中生成的（与 DUTLF-FS 具有 1081 个相同的场景）。与其他数据集相比，该数据的提出是为了更好地利用角度线索。因此，它只提供水平和垂直视点的多视角图像以及中心视角图像的真值图。DUTLF-MV 共包含 1580 个光场，其被分为含有 1100 个样本训练集和含有 480 个样本的测试集。空间分辨率为 400 ×

Tab. 2 光场 SOD 数据集概述。MOP= 多物体比例（整个数据集中含一个以上物体图像的比例），FS= 焦点堆栈，DE= 深度图，MV= 多视角图像，ML= 微透镜阵列图像，GT= 真值图，Raw= 原始光场数据。FS、MV、DE、ML、GT 和 Raw 表示数据集提供的数据形式。

Dataset	Number of images	Spatial resolution	Angular resolution	MOP	FS	MV	DE	ML	GT	Raw	Device
LFSD [42]	100 (No official split)	360 × 360	-	0.04	✓		✓		✓	✓	Lytro
HFUT-Lytro [95]	255 (No official split)	328 × 328	7 × 7	0.29	✓	✓	✓		✓		Lytro
DUTLF-FS [78]	1462 (1000 train, 462 test)	600 × 400	-	0.05	✓		✓		✓		Lytro Illum
DUTLF-MV [56]	1580 (1100 train, 480 test)	590 × 400	7 × 7	0.04		✓			✓		Lytro Illum
Lytro Illum [93]	640 (No official split)	540 × 375	9 × 9	0.15				✓	✓	✓	Lytro Illum

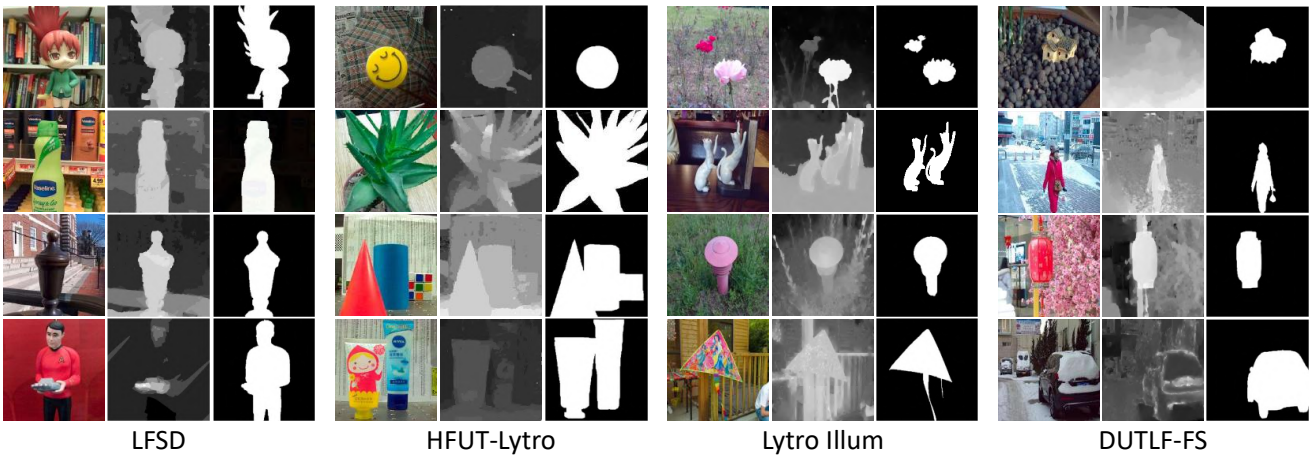

Fig. 7 来自四个数据集的 RGB 图像、深度图和真值图 (GT) 示例：LFSD [42]、HFUT-Lytro [95]、Lytro Illum [93] 和 DUTLF-FS [78]。从左到右依次为 RGB 图像、深度图和真值图。

590，角分辨率为 7 × 7。

Lytro Illum [93]https://github.com/pencilzhang/MAC-light-field-saliency-net 包含了由一台 Lytro Illum 相机拍摄的 640 个高质量光场。该数据集中的图像在物体大小、纹理、杂乱背景和光强度方面差异很大。Lytro Illum 提供了中心视角图像、微透镜图像阵列、原始光场数据以及中心视角图像对应的真值图。微透镜图像的分辨率为 4860 × 3375，而中心视图图像和真值图的空间分辨率为 540 × 375。可计算其角分辨率为 9 × 9。

2.3.2 数据集分析

如表2所总结，我们可以观察到当前数据集存在两个问题，即图片的数量限制和不统一的数据格式。相比于传统 SOD 任务构建的大型数据集，例如 DUT-OMRON（5,168 张图像）[88]、MSRA10K（10,000 张图像）[15] 和 DUTS（15,572 张图像）[75]，现有的光场 SOD 数据集仍然很小，这使得评估数据驱动模型和训练深度网络很困难。此外，其数据格式并不总是一致。例如，Lytro Illum 不提供焦点堆栈，而 DUTLF-FS 和 DUTLF-MV 仅提供焦点堆栈和多视角图像，而不提供原始数据。这使得综合性的评测变得非常困难，因为使用焦点堆栈作为输入的模型不能在 DUTLF-MV 和 Lytro Illum 上测试。我们将在3.2章节展示我们如何缓解这个问题，并在4章节讨论未来的方向。为了更好地了解上述数据集，我们进行了统计分析，包括显著物体的尺寸比例、距图像中心的归一化物体距离分布、焦点切片数量和物体数量。结果如图8和图9所示。图8（a）显示数据集中的大多数物体的大小比例都低于 0.6。HFUT-Lytro 和 Lytro Illum 具有相对较小的物体，而 LFSD 具有相对较大的物体。图8（b）和图9清晰地显示了物体的空间分布。所有五个数据集都呈现出强烈的中心偏向，图8（b）表明来自 Lytro Illum 的物体一般最

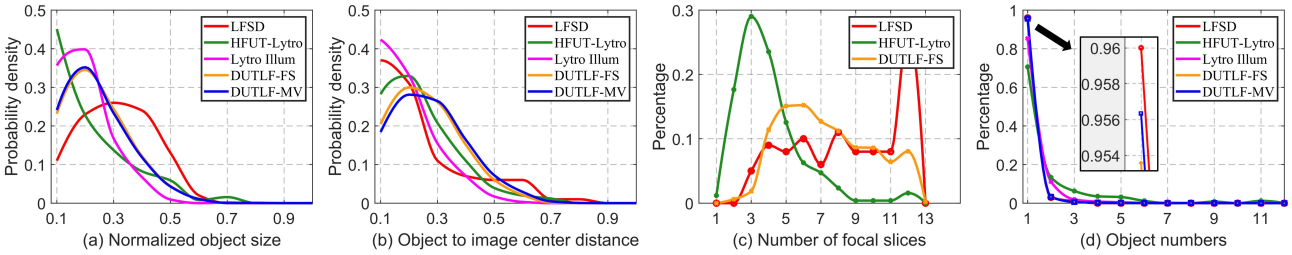

Fig. 8 光场数据集的统计汇总，包括 LFSD [42]，HFUT-Lytro [95]，Lytro Illum [93]，DUTLF-FS [78] 和 DUTLF-MV [56]。从左到右依次为：(a) 归一化的物体大小，(b) 物体与图像中心之间的归一化距离，(c) 焦点切片数量和 (d) 物体数量。

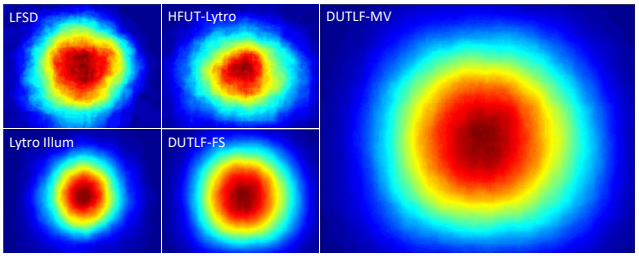

Fig. 9 五个数据集的物体位置分布图（更暖的颜色意味着更高的概率），通过平均真值图计算得到。

接近图像中心。

此外，图8 (c) 给出了焦点切片数量的统计数据。只有 LFSD、HFUT-Lytro 和 DUTLF-FS 三个数据集提供焦点切片。切片数量从 1 到 12 不等，不同数据集之间存在明显的差异。LFSD、HFUT-Lytro 和 DUTLF-FS 分布峰值对应的切片数分别为 12、3 和 6。三个数据集都有不同的切片数量，表明使用焦点堆栈的光场 SOD 模型应该能够处理不同数量的输入切片。最后，由图 8 (d) 可知，数据集中的大多数图像都有单个物体。HFUT-Lytro 和 Lytro Illum 有部分图片包含多个物体（表2 中具有更高的 “MOP” 值），这由益于验证模型在检测多个物体方面的能力。

3 模型评估和评测

在本节中，我们首先回顾五个常用的评估指标，然后提供完善数据集的过程。此外，我们进行评测并对实验结果进行分析。

3.1 评价指标

在对光场 SOD 模型的评测中，我们采用九个广泛使用的指标，其描述如下：

精确率-召回率 (PR) [1, 5, 15] 曲线可以被定义为：

$$P(T)=\frac{|M^T\cap G|}{|M^T|},\qquad R(T)=\frac{|M^T\cap G|}{|G|}, \tag{3}$$

其中 M^T 是通过以阈值 T 对显著图进行阈值化得到的二值掩码，$|\cdot|$ 是掩码的总面积。G 表示真值图。通过从 0 到 255 改变 T 可以获得完整的精确召回曲线。

F 指标 (F_β) [1, 5, 15] 被定义为精确率和召回率的调和平均值：

$$F_\beta=\frac{(1+\beta^2)PR}{\beta^2P+R}, \tag{4}$$

其中 β 是精确率和召回率之间的权重，β^2 通常设置为 0.3 以更加强调精确率。由于通过不同的精确率-召回率对可以获得不同的 F 指标分数，在本文中，我们采用从 PR 曲线计算得到的最大 F 指标 ($F_\beta^{\max}$) 和平均 F 指标 (F_β^{mean})。同时我们还采用了自适应 F 指标 (F_β^{adp}) [1]，其阈值为显著图平均值的两倍。

平均绝对误差 (M) [53] 被定义为：

$$M=\frac{1}{N}\sum_{i=1}^{N}|S_i-G_i|, \tag{5}$$

其中 S_i 和 G_i 表示显著图和真值图中第 i 个像素点处的值。N 为两张图中的像素总数。

S 指标 (S_α) [17, 104] 用来测量显著图和真值图之间的空间结构相似性。它被定义为：

$$S_\alpha=\alpha*S_o+(1-\alpha)*S_r, \tag{6}$$

其中 S_o 和 S_r 分别表示物体感知和区域感知的结构相似性，α 用于平衡 S_o 和 S_r。在本文中，我们按照 [17] 中的建议设置 $\alpha=0.5$。

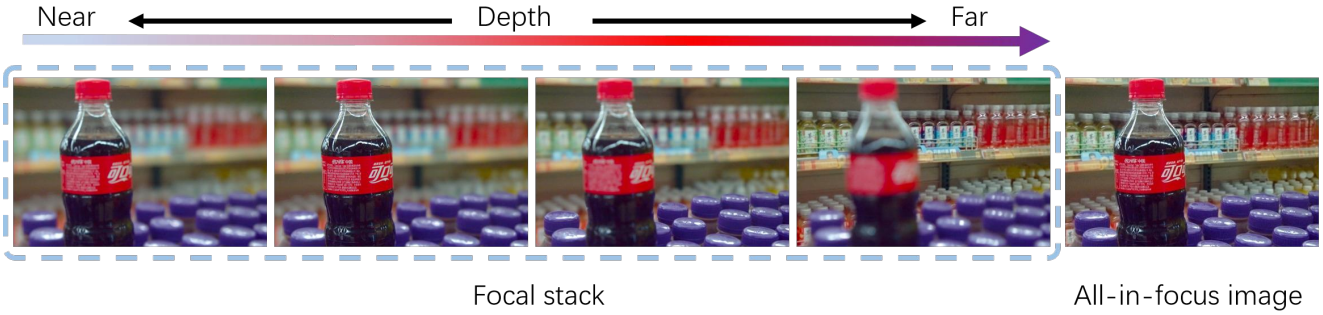

Fig. 10 从 Lytro Illum [93] 生成的焦点堆栈的样例，及其对应的全聚焦图像。

E 指标 (E_ϕ) [18] 是最近提出的评估指标，其考虑了预测值和真值之间的局部和全局相似性。它被定义为：

$$E_\phi = \frac{1}{wh}\sum_{i=1}^{w}\sum_{j=1}^{h}\phi(i,j), \tag{7}$$

其中 $\phi(\cdot)$ 表示增强对齐矩阵 [18]。w 和 h 是真值图的宽度和高度，而 (i,j) 是像素索引。由于 E_ϕ 同样进行两个二值图之间的比较，类似于 F 指标，我们使用所有可能值对显著图进行阈值化处理并求 E_ϕ 的最大值和平均值，分别表示为 $E_\phi^{\max}$ 和 E_ϕ^{mean}；自适应 E_ϕ，即 E_ϕ^{adp}，采取上述自适应 F 指标类似的计算方式，其阈值为平均显著值的两倍 [1]。注意，更高的 PR 曲线、F_β、S_α 和 E_ϕ 以及更低的 M 表示更好的性能。

3.2 数据集完善

如第 2.3章和表 2所示，现有的光场 SOD 数据集面临数据形式不统一的限制。这使得全面的评测变得困难：由于缺乏特定数据，某些模型无法在特定数据集上正确评估。为了缓解这个问题，我们为现有数据集生成补充数据，使其完整和统一，如表3所示，生成数据用 "O" 标记。生成的数据发布于https://github.com/kerenfu/LFSOD-Survey，以促进该领域未来的研究。

一般来说，我们可以使用 LFSD 和 Lytro Illum 两个数据集提供的原始光场数据来合成各种形式的数据。对于 Lytro Illum，我们使用 Lytro Desktop 软件生成了焦点堆栈（包括全聚焦图像）和深度图。在焦点堆栈生成的处理上，我们估计每个图像场景的大概的焦点范围，然后以相等的步长对焦点范围内的焦点切片进行采样。我们删除了整个模糊或重复的切片。Lytro Illum 每个场景上最终生成的焦点切片数量为 2 到 16 个，约 74% 的场景含有超过 6 个切片。图 10展示了一个生成的焦点堆栈的样例。如第 2.1.2章所述，多视角图像和微透镜图像阵列分别通过光场数据的角度和空间采样生成。因此，这两种数据形式可以相互转换。通过这种方式，我们用 Lytro Illum 的微透镜图像阵列生成了多视角图像。我们还可以通过逆向操作为 HFUT-Lytro 合成微透镜图像阵列。然而，我们无法为 DUTLF-MV 合成微透镜图像阵列，因为作者只发布了垂直/水平方向的多视角图像。通过使用原始数据，我们补充了 LFSD 的多视角图像和微透镜图像阵列（图 11）。补全的数据使更全面的模型评估成为可能。例如，基于焦点堆栈的模型，如 MoLF 和 ERNet，现在可在 Lytro Illum 数据集上进行测试。对于 DUTLF-FS/DUTLF-MV，如果作者发布原始（或其他）数据，未来可以继续补充其数据。如此，DUTLF-FS/DUTLF-MV 有可能成为未来模型的标准训练数据集，因为其有较大的规模。

Tab. 3 光场 SOD 数据集的完善；与表 2为对比。FS= 焦点堆栈，DE= 深度图，MV= 多视角图像，ML= 微透镜图像，Raw= 原始光场数据。O 表示我们完善的数据。

Datasets	FS	MV	DE	ML	Raw
LFSD [42]	✓	O	✓	O	✓
HFUT-Lytro [95]	✓	✓	✓	O	
DUTLF-FS [78]	✓		✓		
DUTLF-MV [56]		✓			
Lytro Illum [93]	O	O	O	✓	✓

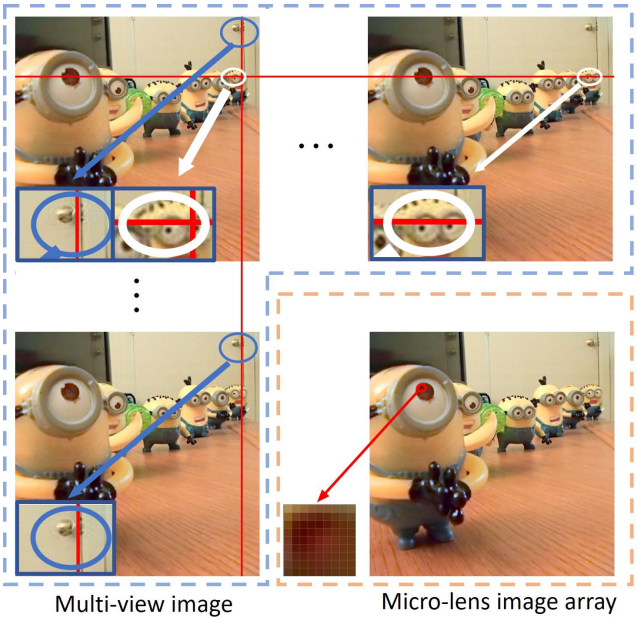

Fig. 11 从 LFSD 数据集 [42] 中生成的多视角图像（360 × 360）和微透镜图像阵列（1080 × 1080）示例。每张图片的左下角展示特写细节以更好地展示视差。微透镜图像阵列由许多微透镜图像组成 [93]。

3.3 评测与分析

3.3.1 测试

为了深入了解不同模型的性能，我们对九种光场 SOD 模型（LFS [42]，WSC [41]，DILF [94]，RDFD [81]，DLSD [56]，MoLF [98]，ERNet [57]，LFNet [97]，MAC [93]）和九个前沿的基于 RGB-D 的 SOD 方法（BBS [22]，JLDCF [25, 26]，SSF [99]，UCNet [92]，D3Net [20]，S2MA [44]，cmMS [38]，HDFNet [51]，and ATSA [96]）在 4 个现有光场数据集上进行了第一次全面的评测。4 个现有数据集包括 LFSD （100 个光场样本），HFUT-Lytro （255 个样本），Lytro Illum （640 个样本）以及 DUTLF-FS 的测试集（462 个样本）。图 7 展示了来自 4 个数据集的样例。此处测评的 RGB-D SOD 模型是在最近的一次综述 [106] 中排名靠前的模型，同样有 ECCV-2020 中最新的开源模型。每个模型所使用的深度图在整个数据集上选择性地反转，以适合此模型的最佳性能。所有参与测评模型有开源或可执行代码，或由作者提供的结果（RDFD [81] 和 LFNet [97] 的显著图由作者提供）。评测使用了上述九个评估指标：PR 曲线、S 指标、最大/平均 F 指标、最大/平均 E 指标、自适应 F 指标和 E 指标、平均绝对误差 (mean absolute error，MAE)，结果如表 4所示。PR 曲线、最大 F 指标曲线和可视化比较如图 12–15所示。

因 DUTLF-MV 数据集 [56] 只提供多视角图像，与大多数光场 SOD 模型的输入数据形式不兼容，因此未在该数据集上进行评测。此外，由于 DLSD [56] 使用了 DUTLF-FS 测试集中约 36% 的测试图像进行训练，因此未对其在 DUTLF-FS 测试集上进行测试。MAC [93] 在 Lytro Illum 数据集进行了五次交叉验证，使得它与其他模型无法直接比较，因此未在 Lytro Illum 上对 MAC [93] 进行评估。由于 DUTLF-FS 无法提供微透镜图像阵列（见表 3），并且 HFUT-Lytro 提供的微透镜图像阵列的质量过低，因此我们遵循 [93]，在这两个数据集的上采样全聚焦图像上测试 MAC。此外，对于 ERNet [57]，因其预训练的学生模型尚未公开，我们仅评估其教师模型。对测评结果的综合分析如下。

3.3.2 传统方法与深度方法

如表 1所示，与四个传统模型相比，基于深度学习的 SOD 模型在所有数据集上都具有更优结果。最佳传统模型，即 DILF，在各数据集各指标上普遍低于深度光场模型，证实了深度神经网络在该项任务上优秀性能。

3.3.3 深度学习模型

如表 1所示，MoLF、ERNet 和 LFNet 采用焦点堆栈和全聚焦图像作为输入数据，而 DLSD 和 MAC 分别使用中心视角图像和微透镜图像阵列。从表 4和图 12可知，MoLF、ERNet 和 LFNet 优于 DLSD 和 MAC。值得注意的是，MoLF 和 ERNet 是最好的两种基于深度的光场 SOD 方法，这可能是因为它们在具有约 1000 个光场样本的大规模数据集 DUTLF-FS 上训练的，或者具有优越的网络结构。测评结果同样表明，基于多视角或微透镜图像的模型性能低于基于焦点堆栈的模型。潜在原因是前者的研究较少，并且多视角和微透镜图像的有效性仍然没有得到充分的研究。此外，训练数据对其性能影响巨大，MAC 仅在 Lytro Illum 上进行训练，其规模约为 DUTLF-FS 的一半。在上述五种模型的比

Tab. 4 定量评估：九个前沿光场 SOD 模型：(LFS [42], WSC [41], DILF [94], RDFD [81], DLSD [56], MoLF [98], ERNet [57], LFNet [97] 和 MAC [93]) 和九个前沿 RGB-D SOD 模型 (BBS [22], JLDCF [25] , SSF [99], UCNet [92], D3Net [20], S2MA [44], cmMS [38], HDFNet [51] 和 ATSA [96]) 上的 S 指标 (S_α) [17]，最大 F 指标 ($F_\beta^{\max}$)，平均 F 指标 (F_β^{mean}) [1]，自适应 F 指标 (F_β^{adp}) [1]，最大 E 指标 ($E_\phi^{\max}$)，平均 E 指标 (E_ϕ^{mean}) [18]，自适应 E 指标 (E_ϕ^{adp}) [1] 和 MAE(M) [53] 结果。光场 SOD 模型用 † 标记。N/T 表示模型没有被测试。排名前三的光场和 RGB-D 模型分别用红色、蓝色和绿色突出表示。↑/↓ 表示更大/更小的值更好。

		Traditional				Deep learning-based													
	Metric	LFS† [42]	WSC† [41]	DILF† [94]	RDFD† [81]	DLSD† [56]	MoLF† [98]	ERNet† [57]	LFNet† [97]	MAC† [93]	BBS [22]	JLDCF [25]	SSF [99]	UCNet [92]	D3Net [20]	S2MA [44]	cmMS [38]	HDFNet [51]	ATSA [96]
LFSD[42]	S_α ↑	0.681	0.702	0.811	0.786	0.786	0.825	0.831	0.820	0.789	0.864	0.862	0.859	0.858	0.825	0.837	0.850	0.846	0.858
	$F_\beta^{\max}$ ↑	0.744	0.743	0.811	0.802	0.784	0.824	0.842	0.824	0.788	0.858	0.867	0.868	0.859	0.812	0.835	0.858	0.837	0.866
	F_β^{mean} ↑	0.513	0.722	0.719	0.735	0.758	0.800	0.829	0.794	0.753	0.842	0.848	0.862	0.848	0.797	0.806	0.850	0.818	0.856
	F_β^{adp} ↑	0.735	0.743	0.795	0.802	0.779	0.810	0.831	0.806	0.793	0.840	0.827	0.862	0.838	0.788	0.803	0.857	0.818	0.852
	$E_\phi^{\max}$ ↑	0.809	0.789	0.861	0.851	0.859	0.880	0.884	0.885	0.836	0.900	0.902	0.901	0.898	0.863	0.873	0.896	0.880	0.902
	E_ϕ^{mean} ↑	0.567	0.753	0.764	0.758	0.819	0.864	0.879	0.867	0.790	0.883	0.894	0.890	0.893	0.850	0.855	0.881	0.869	0.899
	E_ϕ^{adp} ↑	0.773	0.788	0.846	0.834	0.852	0.879	0.882	0.882	0.839	0.889	0.882	0.896	0.890	0.853	0.863	0.890	0.872	0.897
	M ↓	0.205	0.150	0.136	0.136	0.117	0.092	0.083	0.092	0.118	0.072	0.070	0.067	0.072	0.095	0.094	0.073	0.086	0.068
HFUT-Lytro[95]	S_α ↑	0.565	0.613	0.672	0.619	0.711	0.742	0.778	0.736	0.731	0.751	0.789	0.725	0.748	0.749	0.729	0.723	0.763	0.772
	$F_\beta^{\max}$ ↑	0.427	0.508	0.601	0.533	0.624	0.662	0.722	0.657	0.667	0.676	0.727	0.647	0.677	0.671	0.650	0.626	0.690	0.729
	F_β^{mean} ↑	0.323	0.493	0.513	0.469	0.594	0.639	0.709	0.628	0.620	0.654	0.707	0.639	0.672	0.651	0.623	0.617	0.669	0.706
	F_β^{adp} ↑	0.427	0.485	0.530	0.518	0.592	0.627	0.706	0.615	0.638	0.654	0.677	0.636	0.675	0.647	0.588	0.636	0.653	0.689
	$E_\phi^{\max}$ ↑	0.637	0.695	0.748	0.712	0.784	0.812	0.841	0.799	0.797	0.801	0.844	0.778	0.804	0.797	0.777	0.784	0.801	0.833
	E_ϕ^{mean} ↑	0.524	0.684	0.657	0.623	0.749	0.790	0.832	0.777	0.733	0.765	0.825	0.763	0.793	0.773	0.756	0.746	0.788	0.819
	E_ϕ^{adp} ↑	0.666	0.680	0.693	0.691	0.755	0.785	0.831	0.770	0.772	0.804	0.811	0.781	0.810	0.789	0.744	0.779	0.789	0.810
	M ↓	0.221	0.154	0.150	0.214	0.111	0.094	0.082	0.092	0.107	0.089	0.075	0.100	0.090	0.091	0.112	0.097	0.095	0.084
Lytro Illum[93]	S_α ↑	0.619	0.709	0.756	0.738	0.788	0.834	0.843	N/T	N/T	0.879	0.890	0.872	0.865	0.869	0.853	0.881	0.873	0.883
	$F_\beta^{\max}$ ↑	0.545	0.662	0.697	0.696	0.746	0.820	0.827	N/T	N/T	0.850	0.878	0.850	0.843	0.843	0.823	0.857	0.855	0.875
	F_β^{mean} ↑	0.385	0.646	0.604	0.624	0.713	0.766	0.800	N/T	N/T	0.829	0.848	0.836	0.827	0.818	0.788	0.839	0.823	0.848
	F_β^{adapt} ↑	0.547	0.639	0.659	0.679	0.720	0.747	0.796	N/T	N/T	0.828	0.830	0.835	0.824	0.813	0.778	0.835	0.823	0.842
	$E_\phi^{\max}$ ↑	0.721	0.804	0.830	0.816	0.871	0.908	0.911	N/T	N/T	0.913	0.931	0.913	0.910	0.909	0.895	0.914	0.913	0.929
	E_ϕ^{mean} ↑	0.546	0.791	0.726	0.738	0.830	0.882	0.900	N/T	N/T	0.900	0.919	0.907	0.904	0.894	0.873	0.907	0.898	0.919
	E_ϕ^{adapt} ↑	0.771	0.797	0.812	0.815	0.855	0.876	0.900	N/T	N/T	0.912	0.914	0.917	0.907	0.907	0.878	0.915	0.904	0.917
	M ↓	0.197	0.115	0.132	0.142	0.086	0.065	0.056	N/T	N/T	0.047	0.042	0.044	0.048	0.050	0.063	0.045	0.051	0.041
DUTLF-FS[78]	S_α ↑	0.585	0.656	0.725	0.658	N/T	0.887	0.899	0.878	0.804	0.894	0.905	0.908	0.870	0.852	0.845	0.906	0.868	0.905
	$F_\beta^{\max}$ ↑	0.533	0.617	0.671	0.599	N/T	0.903	0.908	0.891	0.792	0.884	0.908	0.915	0.864	0.840	0.829	0.906	0.857	0.915
	F_β^{mean} ↑	0.358	0.607	0.582	0.538	N/T	0.855	0.891	0.843	0.746	0.867	0.885	0.907	0.854	0.820	0.806	0.893	0.841	0.899
	F_β^{adapt} ↑	0.525	0.617	0.663	0.599	N/T	0.843	0.885	0.831	0.790	0.872	0.874	0.903	0.850	0.826	0.791	0.887	0.835	0.893
	$E_\phi^{\max}$ ↑	0.711	0.788	0.802	0.774	N/T	0.939	0.949	0.930	0.863	0.923	0.943	0.946	0.909	0.891	0.883	0.936	0.898	0.943
	E_ϕ^{mean} ↑	0.511	0.759	0.695	0.686	N/T	0.921	0.943	0.912	0.806	0.908	0.932	0.939	0.904	0.874	0.866	0.928	0.889	0.938
	E_ϕ^{adapt} ↑	0.742	0.787	0.813	0.782	N/T	0.923	0.943	0.913	0.872	0.924	0.930	0.942	0.905	0.895	0.870	0.931	0.895	0.936
	M ↓	0.227	0.151	0.156	0.191	N/T	0.051	0.039	0.054	0.102	0.054	0.043	0.036	0.059	0.069	0.079	0.041	0.065	0.039

较中，ERNet 具有最高准确度。

3.3.4 光场与 RGB-D SOD 模型

从表 4和图 12中的定量结果可知，最新前沿 RGB-D 模型取得了与光场 SOD 模型相当甚至更好的结果。在大多数数据集上，JLDCF、SSF 和 ATSA 通常优于 ERNet。根本原因可能有两方面。首先，近年来 RGB-D SOD 引起了广泛的研究兴趣，大量功能强大且复杂的模型被提出。受早期对 RGB SOD 问题研究 [24, 60, 86] 的启发，这些模型通常从深度神经网络中寻求边缘保持结果，并采用功能模块和架构，例如边界补充单元 [99]，多尺度特征聚合模块 [99]，或 UNet 型的自下而上/自上而下架构 [25, 39, 44]。相反，光场 SOD 较少被探索，并且其模型和体系结构发展缓慢。大多数现有模型尚未考虑边缘感知特性。例如，尽管在 ERNet 中采用了注意机制和 ConvLSTM，但没有使用类似于 UNet 的自上而下细化来生成边缘感知显著图。如图 1和 14所示，RGB-D SOD 模型比现有的深度光场 SOD 模型更能检测出精确的边界。其次，另一个潜在原因是 RGB-D SOD 模型是在更多数据上训练的。例如，RGB-D SOD 任务普遍使用的训练集包含 2200 个 RGB-D 场景 [25]，而 ERNet [57] 只在约 1000 个光场上进行训练。因此，前者更可能具有更好的泛化能力。

Tab. 5 定量评估：一个重新训练的光场 SOD 模型 (ERNet [57]) 和七个重新训练的 RGB-D SOD 模型（BBS [22]，SSF [99]，ATSA [96]，S2MA [44]，D3Net [20]，HDFNet [51] 和 JLDCF [25]）上 S 指标 (S_α) [17]，最大 F 指标 ($F_\beta^{\max}$)，最大 E 指标 ($E_\phi^{\max}$) 和 MAE (M) [53] 结果。原始模型数据来自表 4，重新训练的模型用 * 标记。重新训练的模型的最好的结果用粗体突出显示。↑/↓ 表示更大/更小的值更好。

Models	LFSD [42]				HFUT-Lytro [95]				Lytro Illum [93]				DUTLF-FS [78]			
	$S_\alpha\uparrow$	$F_\beta^{\max}\uparrow$	$E_\phi^{\max}\uparrow$	$M\downarrow$	$S_\alpha\uparrow$	$F_\beta^{\max}\uparrow$	$E_\phi^{\max}\uparrow$	$M\downarrow$	$S_\alpha\uparrow$	$F_\beta^{\max}\uparrow$	$E_\phi^{\max}\uparrow$	$M\downarrow$	$S_\alpha\uparrow$	$F_\beta^{\max}\uparrow$	$E_\phi^{\max}\uparrow$	$M\downarrow$
BBS [22]	0.864	0.858	0.900	0.072	0.751	0.676	0.801	0.089	0.879	0.850	0.913	0.047	0.894	0.884	0.923	0.054
SSF [99]	0.859	0.868	0.901	0.067	0.725	0.647	0.778	0.100	0.872	0.850	0.913	0.044	0.908	0.915	0.946	0.036
ATSA [96]	0.858	0.866	0.902	0.068	0.772	0.729	0.833	0.084	0.883	0.875	0.929	0.041	0.905	0.915	0.943	0.039
ERNet [57]	0.831	0.842	0.884	0.083	0.778	0.722	0.841	0.082	0.843	0.827	0.911	0.056	0.899	0.908	0.949	0.039
S2MA [44]	0.837	0.835	0.873	0.094	0.729	0.650	0.777	0.112	0.853	0.823	0.895	0.063	0.845	0.829	0.883	0.079
D3Net [20]	0.825	0.812	0.863	0.095	0.749	0.671	0.797	0.091	0.869	0.843	0.909	0.050	0.852	0.840	0.891	0.069
HDFNet [51]	0.846	0.837	0.879	0.086	0.763	0.690	0.801	0.095	0.873	0.855	0.913	0.051	0.868	0.857	0.898	0.065
JLDCF [25]	0.862	0.867	0.902	0.070	0.789	0.727	0.844	0.075	0.890	0.878	0.931	0.042	0.905	0.908	0.943	0.043
BBS* [22]	0.739	0.738	0.812	0.123	0.708	0.622	0.773	0.102	0.825	0.788	0.878	0.065	0.873	0.870	0.919	0.051
SSF* [99]	0.790	0.793	0.861	0.097	0.687	0.612	0.781	0.099	0.833	0.799	0.886	0.059	0.881	0.889	0.930	0.050
ATSA* [96]	0.816	0.823	0.873	0.087	0.727	0.673	0.805	0.094	0.844	0.822	0.905	0.054	0.880	0.892	0.936	0.045
ERNet* [57]	0.822	0.825	0.885	0.085	0.707	0.632	0.766	0.117	0.840	0.810	0.900	0.059	0.898	0.903	0.946	0.040
S2MA* [44]	0.827	0.829	0.873	0.086	0.672	0.572	0.735	0.120	0.839	0.802	0.885	0.060	0.894	0.893	0.934	0.046
D3Net* [20]	0.827	0.821	0.877	0.086	0.720	0.645	0.801	0.092	0.859	0.835	0.906	0.051	0.906	0.911	0.947	0.039
HDFNet* [51]	0.849	0.850	0.891	0.073	0.747	0.673	0.801	**0.085**	0.874	0.854	0.915	0.045	0.922	**0.931**	0.955	**0.030**
JLDCF* [25]	**0.850**	**0.860**	**0.900**	**0.071**	**0.755**	**0.694**	**0.823**	0.086	**0.877**	**0.855**	**0.919**	**0.042**	**0.924**	**0.931**	**0.958**	**0.030**

尽管如此，我们仍无法否认光场对提高 SOD 性能的潜力，因为最近 RGB-D SOD 比光场 SOD 更加活跃，并且大量新的具有竞争力模型被提出。此外，在评测数据集上，ERNet 和 MoLF 的性能仅略低于 RGB-D 模型，这进一步表明光场数据对 SOD 的有效性 [103]。由于光场可以提供比成对的 RGB 和深度图更多的信息，光场 SOD 仍有很大的改进空间。

此外，为消除训练差异，我们在统一的训练集（即包含 1000 个场景的 DUTLF-FS 训练集）上重新训练 RGB-D 模型。我们还重新训练 ERNet，以移除其使用的额外的 HFUT-Lytro 训练数据，如表 1所示。对比结果如表 5所示，其中所有模型普遍发生性能退化。有趣的是，经过重新训练，SSF* 不再优于 ERNet*，而 ATSA* 在 LFSD 和 DUTLF-FS 上则不如 ERNet*。JLDCF* 和 HDFNet* 始终明显优于 ERNet*。

3.3.5 数据集间的准确度

表 4和图 12清楚地表明，模型在不同的数据集上表现不同。一般来说，模型在 LFSD 上比在其他三个数据集上能获得更好的结果，这表明 LFSD 是光场 SOD 最简单的数据集，传统模型 DILF 甚至可在该数据集上优于 DLSD 和 MAC 等深度模型。相比之下，HFUT-Lytro、Lytro Illum 和 DUTLF-FS 更具挑战性。MoLF、ERNet 和 ATSA 在 DUTLF-FS 上表现好，可能是由于它们是在 DUTLF-FS 训练集或训练数据上进行训练（见表 1）。此外，如章节 2.3所述，HFUT-Lytro 有大量小尺寸显著物体，同时单图具有多个物体。模型在该数据集上的性能降低表明，无论对基于 RGB-D 的模型还是光场模型，检测小尺寸/多个显著物体对现有方案仍非常具有挑战。这使得 HFUT-Lytro 成为现有最具挑战性的光场数据集。

3.3.6 可视化结果

五个光场模型（包括两种传统方法 LFS 和 DILF，三种基于深度学习的模型 DLSD、MoLF 和 ERNet）以及三种最新的基于 RGB-D 的模型（JLDCF、BBS 和 ATSA）的可视化结果如图 14所示。图 14中前两行展示简单场景，第三到第五行展示具有复杂背景或复杂边

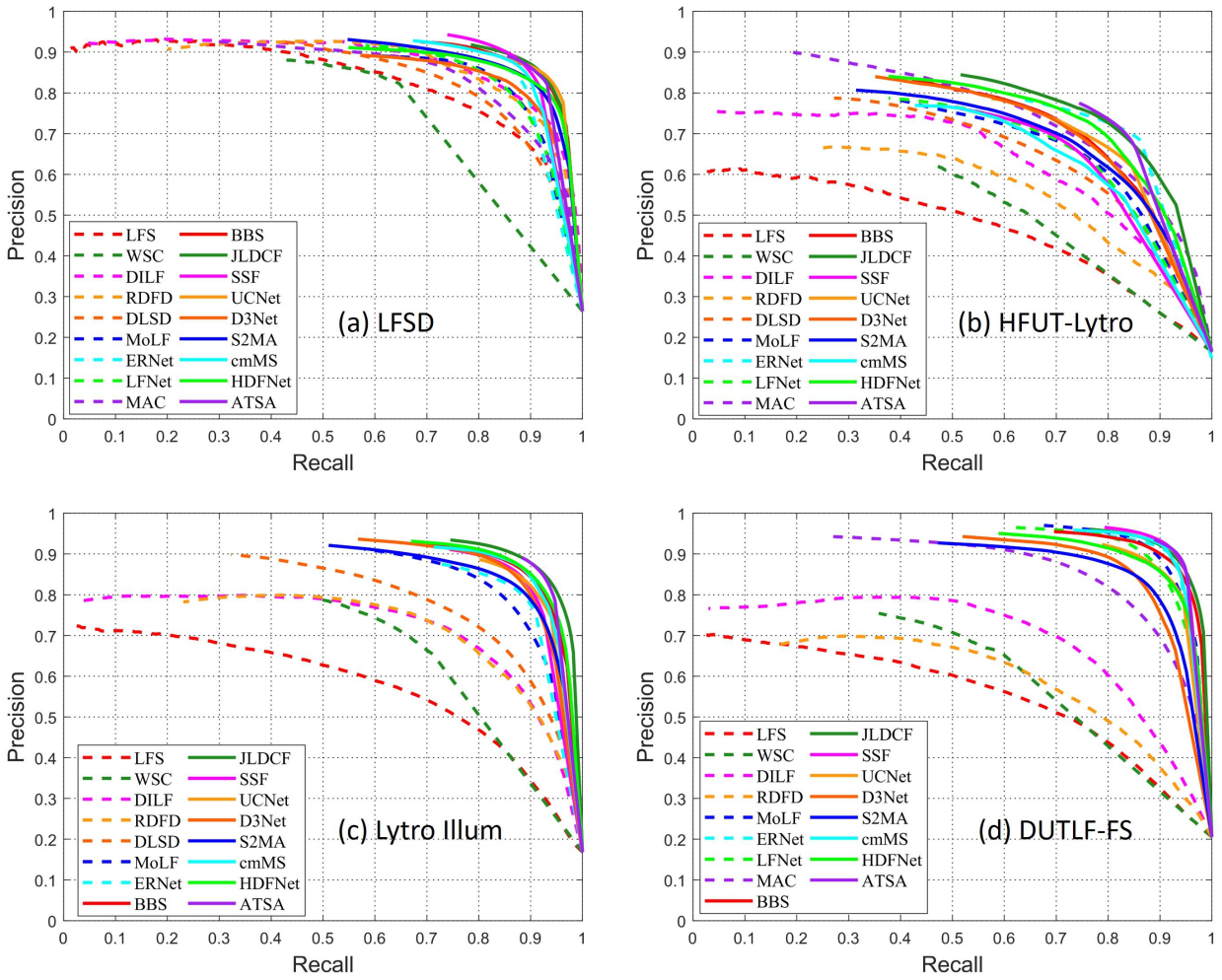

Fig. 12 九个光场 SOD 模型：LFS [42]，WSC [41]，DILF [94]，RDFD [81]，DLSD [56]，MoLF [98]，ERNet [57]，LFNet [97] 和 MAC [93]，以及九个 RGB-D SOD 模型：BBS [22]，JLDCF [25, 26] ，SSF [99]，UCNet [92]，D3Net [20]，S2MA [44]，cmMS [38]，HDFNet [51] 和 ATSA [96] 在四个数据集 ((a) LFSD [42]，(b) HFUT-Lytro [95]，(c) Lytro Illum [93] 和 (d)DUTLF-FS [78]) 上的 PR 曲线。实线和虚线分别代表了 *RGB-D SOD* 模型和光场 *SOD* 模型的 PR 曲线。

界场景。最后一行展示前景和背景颜色对比度较低的场景。如图所示，RGB-D 模型的性能与光场模型相当甚至更好，这表明该领域的研究仍不够充分。图 15进一步展示了具有小尺寸物体和多个显著物体的场景，其中前三行展示具有多个显著物体的情况，其他行展示具有小尺寸物体的情况。在这种情况下，基于 RGB-D 的模型和光场模型都更有可能产生错误检测，这证实了现有技术处理小尺寸或多个物体的能力较差。

4 挑战和开放性的方向

本节重点介绍光场 SOD 的几个未来研究方向，并概述了几个尚未解决的问题。

4.1 数据集收集和统一

如第 2.3节所述，现有光场数据集规模有限，并且数据表示不统一，使评估不同的模型和概括深度网络变得困难。不同于其他 SOD 任务，如 RGB-D SOD [25, 96, 99] 和视频 SOD [21, 72]，数据形式不统一问题对于光场 SOD 尤其严重，因其具有不同的数据表示并且高度依赖特殊采集硬件。因此，创建大规模并且统一的数据集

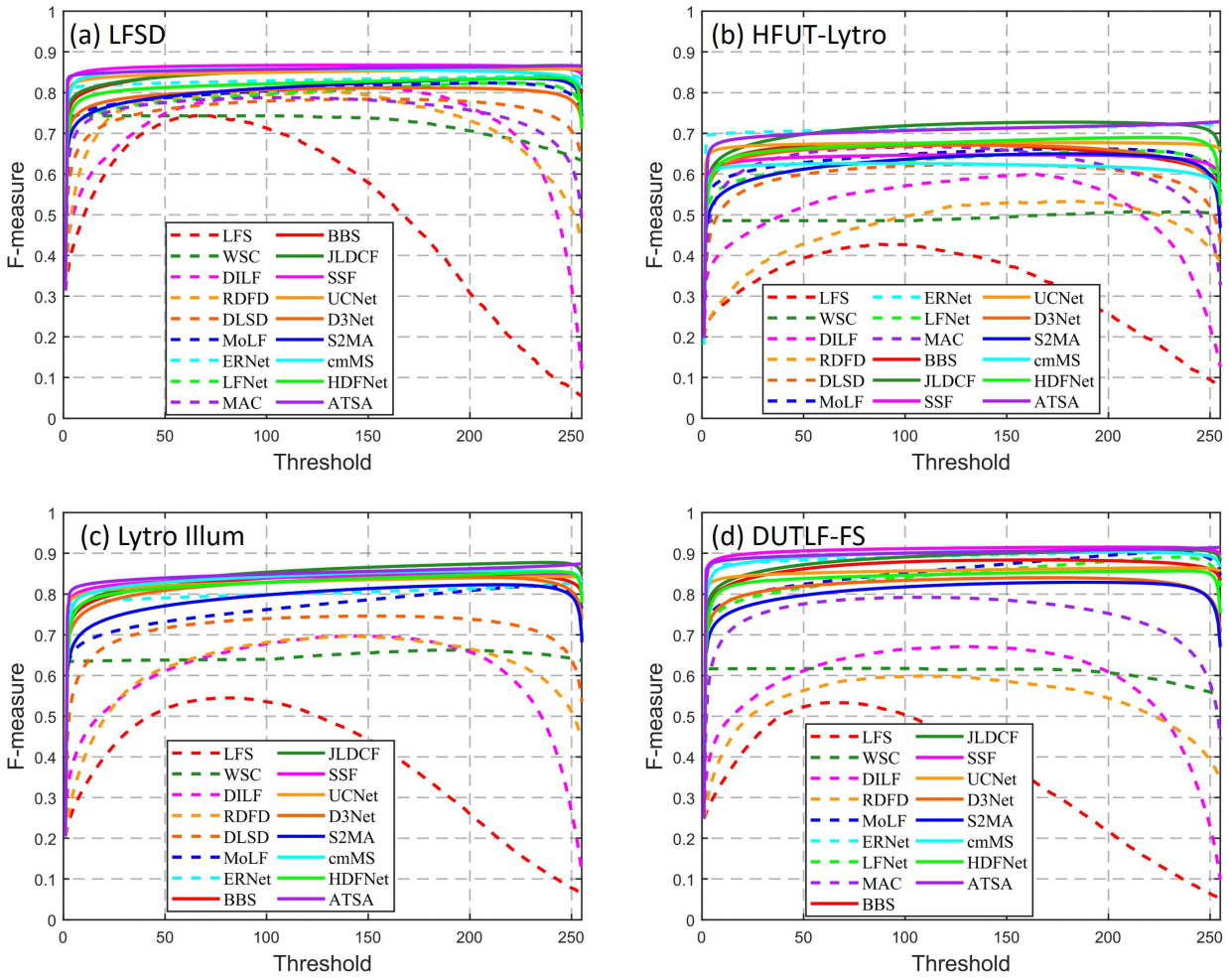

Fig. 13 九个光场 SOD 模型：LFS [42]，WSC [41]，DILF [94]，RDFD [81]，DLSD [56]，MoLF [98]，ERNet [57]，LFNet [97] 和 MAC [93]，以及九个 RGB-D SOD 模型：BBS [22]，JLDCF [25, 26] ，SSF [99]，UCNet [92]，D3Net [20]，S2MA [44]，cmMS [38]，HDFNet [51] 和 ATSA [96] 在四个数据集（(a) LFSD [42]，(b) HFUT-Lytro [95]，(c) Lytro Illum [93] 和 (d)DUTLF-FS [78]）上的 F 指标曲线。实线和虚线分别代表了 *RGB-D SOD* 模型和光场 *SOD* 模型的 F 指标曲线。

对于未来的研究至关重要。我们敦促研究人员在构建新的数据集时考虑这个问题。此外，收集完整的数据形式，包括原始数据、焦点堆栈、多视角图像、深度图和微透镜图像阵列，定会促进和推动这一领域的研究。此外，由于原始光场数据相当耗费存储空间（例如，Lytro Illum 的 640 个光场占用 32.8 GB），因此，尤其对于大规模数据而言，在数据存储和传输方面相当具有挑战。数据集的规模使其难以传播。在这种情况下，如果任何数据形式的子集都可供公众使用将是一件好事。

4.2 发展光场 SOD

如上所述，与显著性领域的其他任务相比，目前对光场 SOD 的研究较少。因此，该领域仍处于探索阶段。从第 3.3节中的评测结果可知，前沿方法的性能仍然远不能令人满意，尤其是在 HFUT-Lytro 数据集上。光场 SOD 算法和模型还存在很大进步空间。此外，我们注意到，从 2019 年至 2020 年，仅有七种基于深度学习的光场 SOD 模型出现。我们将光场 SOD 研究的缺乏归因于上述数据问题，以及缺乏对该主题现有方法和数据集的全面调查。

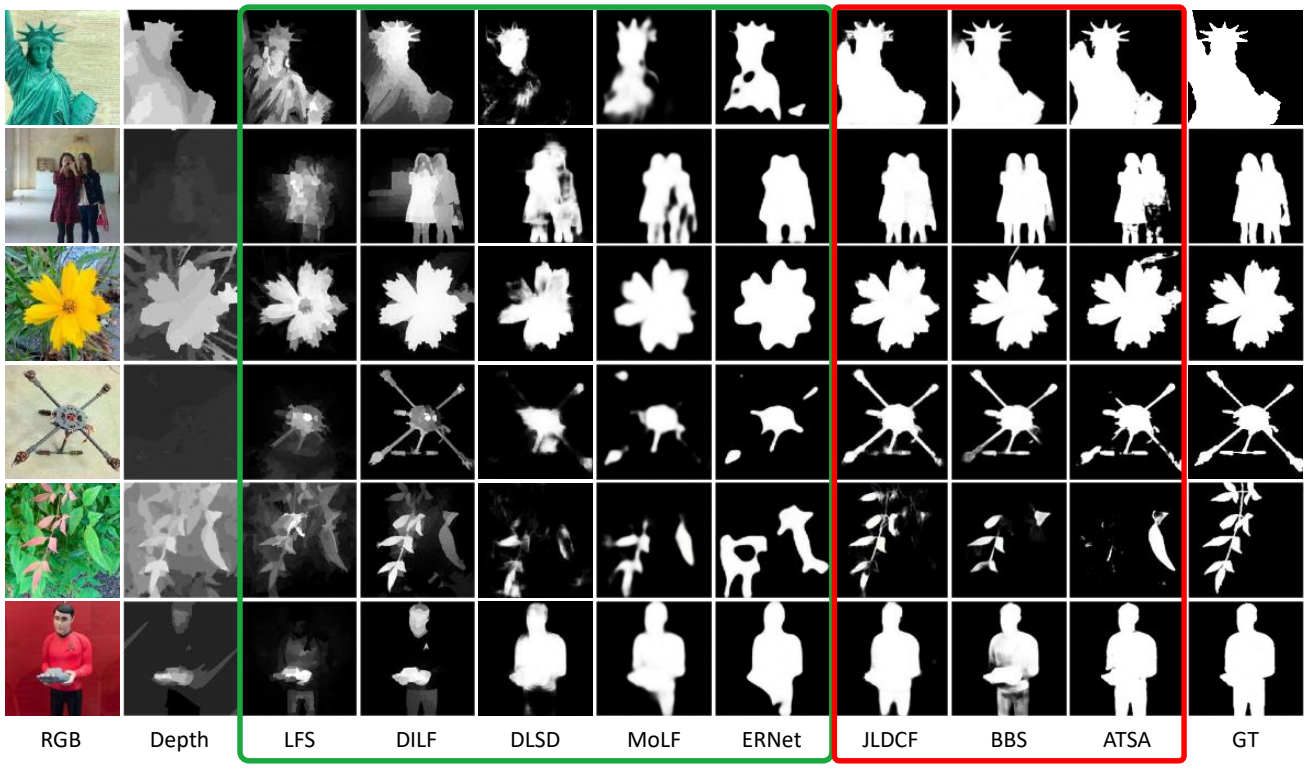

Fig. 14 五个光场 SOD 模型（绿色框）：LFS [42]，DILF [94]，DLSD [56]，MoLF [98] 和 ERNet [57] 以及三个 RGB-D SOD 模型（红色框）：JLDCF [25, 26]，BBS [22] 和 ATSA [96] 的可视化比较。

4.3 多视角图像和微透镜图像阵列

如表 1所示，大多数现有模型使用焦点堆栈和深度图，而多视角图像和微透镜图像阵列作为另外两种类型的光场数据表示很少被考虑（仅在五种模型中被使用）。第 3.3节中的评测结果表明，后者的性能不如使用其他数据形式的模型，所以这两种数据形式的使用尚未被充分探讨。因此，需要对光场 SOD 模型进行更多的研究，以探索多视角图像和微透镜图像阵列的有效性。或者，这两种数据表示的信息量可能不如焦点堆栈和深度图——场景深度信息可能被更含蓄地表示。这可能会令使用深度神经网络查找有效映射和挖掘潜在规则变得困难，尤其是在训练数据稀疏的情况下。比较不同数据表示对显著性检测的有效性和冗余性是很有趣的。

4.4 结合高质量深度估计

研究表明，精确的深度图有助于从复杂背景中发现显著物体。不幸的是，在现有数据集中，深度图的质量差异很大，因为从光场估计深度是一项具有挑战性的任务 [32, 54, 58, 69, 70, 77]。该挑战源于一个事实：尽管光场可以通过数字重聚焦技术合成聚焦在任何深度的图像，但每个场景点的深度分布是未知的。此外，确定图像区域是否聚焦本身就是一个难题 [52, 105]。不完善的深度图通常会对使用深度图的模型的检测精度产生负面影响。因此，结合来自光场的高质量深度估计算法有益于光场 SOD。

4.5 边缘感知光场 SOD

SOD 作为一项像素级分割任务 [5]，精确的目标边界对于高质量的显著图至关重要。在 RGB SOD 领域，边缘感知 SOD 模型正在引起越来越多的研究关注 [24, 60, 86]。如实验结果所示，现有的深度光场 SOD 模型很少考虑该问题，导致显著图的粗糙边界和边缘。因此，边缘感知光场 SOD 是一个未来的研究方向。

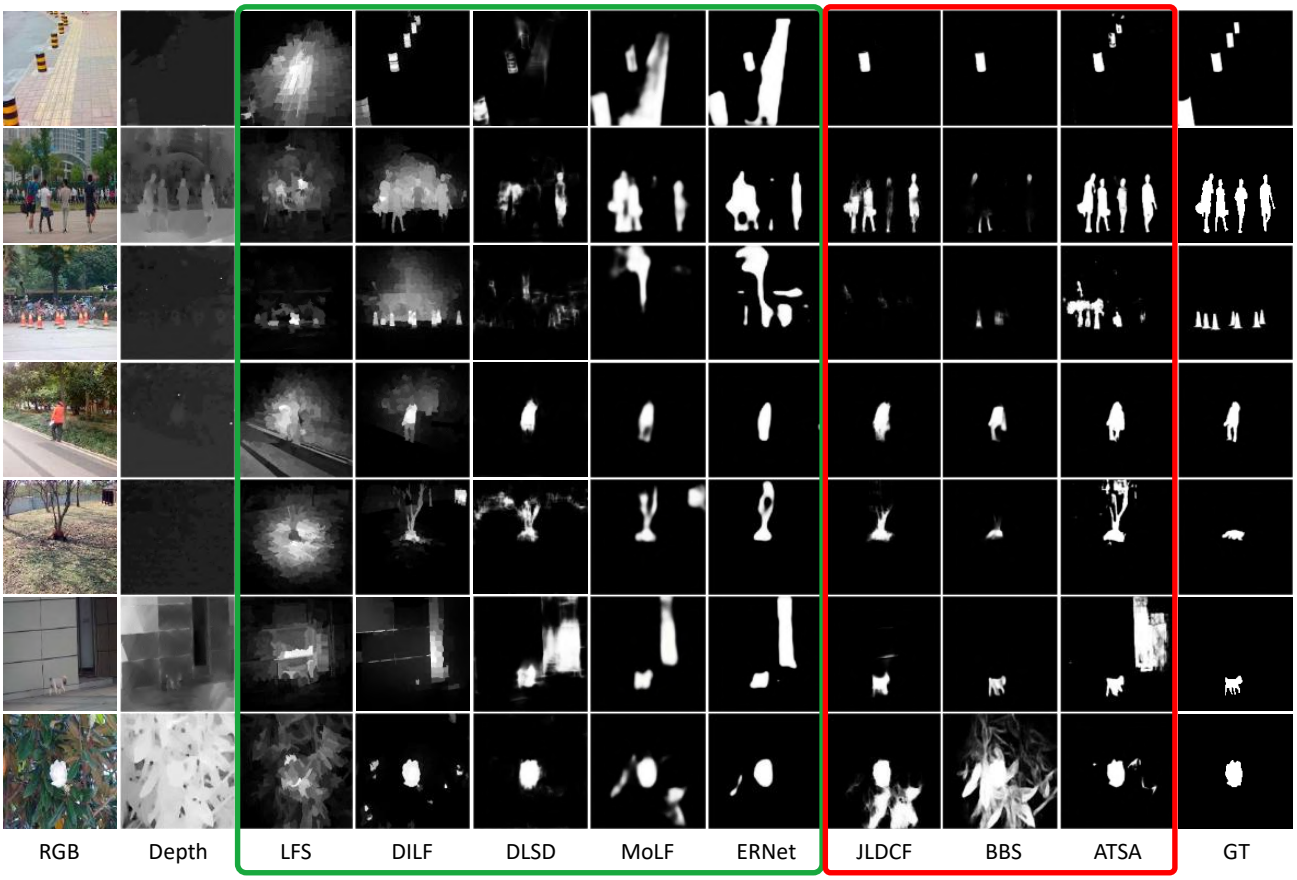

Fig. 15 当检测小物体或者多物体场景时，五个光场 SOD 模型（绿色框）：LFS [42]，DILF [94]，DLSD [56]，MoLF [98] 和 ERNet [57] 以及三个 RGB-D SOD 模型（红色框）：JLDCF [25, 26]，BBS [22] 和 ATSA [96] 的可视化比较。

4.6 采集技术和硬件

第一代光场相机 Lytro 于 2011 年发明，而其下一代 Lytro Illum 于 2014 年推出。后者功能更强大，但比前者体积更大，并且价格也更昂贵。然而总的来说，光场采集技术和硬件的发展速度比电脑、手机等要慢。从 2014 年开始，商用光场相机稀缺。因此发展光场摄影的采集技术和硬件技术成为迫切的需要。目前，光场相机在图像质量、价格和便携性方面还远不能取代传统的 RGB 相机。如果未来光场相机变得便宜且小巧，它们能很容易地集成到手机中，让每个人都可以在日常生活中尝试光场摄影。这将大大增加用户数据和后处理应用程序的需求，为光场 SOD 的发展铺平道路。

4.7 监督策略

现有的基于深度学习的光场模型使用完全监督的方式来学习分割显著物体，这需要足够的标注的训练数据。不幸的是，现有数据集的大小有限：DUTLF-FS 和 DUTLF-MV 分别提供 1000 个和 1100 个训练样本，而其他数据集包含的光场样本都少于 640 个。一方面，少量的训练数据限制了模型的泛化能力。而另一方面，获取大量带标注的数据需要大量的人工成本来进行数据收集和标记。最近，弱监督和半监督学习策略引起了广泛的研究关注，该策略大大减少了标注工作。由于数据友好，它们已被引入 RGB SOD，并进行了一些令人鼓舞的尝试 [59, 90, 91]。因此，未来的一个方向是将这些监督策略扩展到光场 SOD，以克服训练数据不足的问题。此外，几项工作 [12, 16] 已经表明，以自监督的方式预训练模型可以有效提高性能，所以未来也可以将其

引入光场 SOD。

4.8 RGB-D 和光场 SOD 的联系

光场 SOD 和 RGB-D SOD 之间存在密切的联系，因为这两项任务都探索用于显著性检测的场景深度信息，而深度信息可以使用各种技术从光场数据中获得。这就是为什么 RGB-D SOD 可以被视为光场 SOD 退化的方案。如表 4所示，将 RGB-D SOD 模型应用于光场很简单，而我们认为反过来也是可能的。例如，直观上看，在一对 RGB 和深度图像上重建光场数据（例如焦点堆栈或多视角图像）是可能的 [56]。如果实现了这个联接，这两个领域的模型之间的相互转换就变得可行了，之后光场模型就可以应用于 RGB-D 数据。在不久的将来，这种联系将是一个值得探讨的有趣问题。

4.9 其他潜在方向

受显著性领域最新进展的启发，未来研究还有其他几个潜在的方向。例如，高分辨率显著物体检测 [89] 旨在处理高分辨率图像的显著物体分割，在光场 SOD 中可以考虑实现高分辨率细节。此外，虽然现有的光场数据集是在对象级别进行标注的，用于分离单个对象的实例级注释和检测 [9, 10, 23, 40, 45] 也可以引入这一领域。实例敏感的应用场景很多，例如图像字幕 [34]、多标签图像识别 [85]，以及各种弱监督/无监督学习场景 [13, 36]。最新的工作试图解决弱监督显著性实例检测 [71]。同样，可以将更多精力放在实例级的真值注释和设计实例级光场 SOD 模型上。此外，人眼关注点预测 [4, 5, 7] 是显著性检测的另一个子领域。到目前为止，还没有使用光场数据进行人眼关注点预测的研究。由于光场提供了丰富的自然场景信息，我们希望光场的各种数据形式可以提供有用线索来帮助消除模糊的人眼关注区域。最后，光场数据有利于其他与 SOD 密切相关的任务，例如伪装物体检测 (COD) [19] 和透明物体分割 [87]，该类物体经常从背景中借用纹理，并与周围环境有相似的外观。最后，还有一个悬而未决的问题：光场信息如何比深度信息更有益于 SOD？深度信息可以由光场数据生成，并且是光场数据的子集。不同形式的光场数据，例如焦点堆栈和多视角图像，在某种程度上隐含了深度信息，这表明现有模型可能在隐式地利用这些深度信息。那么用显式方式（例如 RGB-D SOD 模型）和隐式方式使用深度信息有什么区别呢？这是一个有趣的问题，但遗憾的是，自从 2014 年提出光场 SOD 问题以来，没有任何研究给出任何直接的答案或证据。这值得今后进一步研究和了解。

5 结论

我们为光场 SOD 提供了第一个综合性的综述和评测，同时总结和讨论了现有的研究和相关数据集。我们对具有代表性的光场 SOD 模型进行了评测，并从定性和定量角度将它们与几个前沿的 RGB-D SOD 模型进行了比较。由于现有的光场数据集在数据表示上有些不一致，我们为现有数据集生成了补充数据，使其完整和统一。此外，我们还讨论了未来研究的几个潜在方向，并概述了一些尚未解决的问题。尽管光场 SOD 在过去几年取得了进展，但关于该问题的基于深度学习的工作仍然只有七项，这为设计更强大的网络架构留下了很大的空间，其中包含例如边缘感知设计和自上而下的细化等有效模块，以此提高 SOD 性能。我们希望这项调查将成为推动这一领域发展的催化剂，并在未来促进一些有趣的工作。

References

[1] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk. Frequency-tuned salient region detection. In *IEEE conference on computer vision and pattern recognition*, pages 1597–1604, 2009.

[2] E. Adelson and J. Bergen. The plenoptic function and the elements of early vision. In *Computational Models of Visual Processing*. Cambridge: MIT Press, 1991.

[3] A. Agarwala, M. Dontcheva, M. Agrawala, S. Drucker, A. Colburn, B. Curless, D. Salesin, and M. Cohen. Interactive digital photomontage. *ACM Transactions on Graphics*, 2004.

[4] A. Borji. Saliency prediction in the deep learning era: Successes and limitations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43:679–700, 2021.

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

[6] A. Borji, M.-M. Cheng, H. Jiang, and J. Li. Salient

object detection: A survey. *Computational Visual Media*, 5:117–150, 2019.

[7] A. Borji and L. Itti. State-of-the-art in visual attention modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35:185–207, 2013.

[8] A. Borji and L. Itti. Defending yarbus: eye movements reveal observers' task. *Journal of vision*, 14 3:29, 2014.

[9] Z. Cai and N. Vasconcelos. Cascade r-cnn: High quality object detection and instance segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43:1483–1498, 2021.

[10] K. Chen, J. Pang, J. Wang, Y. Xiong, X. Li, S. Sun, W. Feng, Z. Liu, J. Shi, W. Ouyang, C. C. Loy, and D. Lin. Hybrid task cascade for instance segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4969–4978, 2019.

[11] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40:834–848, 2018.

[12] 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 696–705, 2020.

[13] X. Chen and A. Gupta. Webly supervised learning of convolutional networks. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1431–1439, 2015.

[14] M.-M. Cheng, Y. Liu, W.-Y. Lin, Z. Zhang, P. L. Rosin, and P. H. Torr. Bing: Binarized normed gradients for objectness estimation at 300fps. *Computational Visual Media*, 5(1):3–20, 2019.

[15] M.-M. Cheng, G.-X. Zhang, N. Mitra, X. Huang, and S. Hu. Global contrast based salient region detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 409–416, 2011.

[16] 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 846–855, 2020.

[17] 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 4558–4567, 2017.

[18] 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*, 2018.

[19] 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 2774–2784, 2020.

[20] 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*, 32(5):2075–2089, 2020.

[21] 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 8546–8556, 2019.

[22] 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 *European Conference on Computer Vision*, pages 275–292, 2020.

[23] R. Fan, M.-M. Cheng, Q. Hou, T.-J. Mu, J. Wang, and S. Hu. S4net: Single stage salient-instance segmentation. *Computational Visual Media*, 6:191–204, 2020.

[24] M. Feng, H. Lu, and E. Ding. Attentive feedback network for boundary-aware salient object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1623–1632, 2019.

[25] 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. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3049–3059, 2020.

[26] K. Fu, D.-P. Fan, G.-P. Ji, Q. Zhao, J. Shen, and C. Zhu. Siamese network for rgb-d salient object detection and beyond. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.

[27] K. Fu, Q. Zhao, I. Gu, and J. Yang. Deepside: A general deep framework for salient object detection. *Neurocomputing*, 356:69–82, 2019.

[28] A. Gershun. The light field. *Studies in Applied Mathematics*, 18(1-4):51–151, 1939.

[29] J. Han, E. J. Pauwels, and P. M. de Zeeuw. Fast saliency-aware multi-modality image fusion. *Neurocomputing*, 111:70–80, 2013.

[30] K. He, J. Sun, and X. Tang. Single image haze removal using dark channel prior. *IEEE transactions on pattern analysis and machine intelligence*, 33(12):2341–2353, 2010.

[31] L. Itti. Automatic foveation for video compression using a neurobiological model of visual attention. *IEEE Transactions on Image Processing*, 13:1304–1318, 2004.

[32] H. Jeon, J. Park, G. Choe, J. Park, Y. Bok, Y.-W. Tai, and I.-S. Kweon. Accurate depth map estimation from a lenslet light field camera. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1547–1555, 2015.

[33] P. Jiang, H. Ling, J. Yu, and J. Peng. Salient region detection by ufo: Uniqueness, focusness and objectness. In *Proceedings of the IEEE International*

Conference on Computer Vision, pages 1976–1983, 2013.

[34] A. Karpathy and F.-F. Li. Deep visual-semantic alignments for generating image descriptions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39:664–676, 2017.

[35] S. Kuthirummal, H. Nagahara, C. Zhou, and S. Nayar. Flexible depth of field photography. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33:58–71, 2011.

[36] B. Lai and X. Gong. Saliency guided dictionary learning for weakly-supervised image parsing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3630–3639, 2016.

[37] M. Levoy and P. Hanrahan. Light field rendering. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, 1996.

[38] C. Li, R. Cong, Y. Piao, Q. Xu, and C. C. Loy. Rgb-d salient object detection with cross-modality modulation and selection. In *European Conference on Computer Vision*, pages 225–241, 2020.

[39] 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.

[40] G. Li, Y. Xie, L. Lin, and Y. Yu. Instance-level salient object segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 247–256, 2017.

[41] 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.

[42] 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.

[43] 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.

[44] 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*, pages 13753–13762, 2020.

[45] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia. Path aggregation network for instance segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8759–8768, 2018.

[46] Y.-F. Ma, X.-S. Hua, L. Lu, and H.-J. Zhang. A generic framework of user attention model and its application in video summarization. *IEEE Transactions on Multimedia*, 7:907–919, 2005.

[47] Y.-F. Ma, L. Lu, H.-J. Zhang, and M. Li. A user attention model for video summarization. In *Proceedings of the tenth ACM international conference on Multimedia*, pages 533–542, 2002.

[48] F. Moosmann, D. Larlus, and F. Jurie. Learning saliency maps for object categorization. In *International Workshop on The Representation and Use of Prior Knowledge in Vision*, pages 1–15, 2006.

[49] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan. Light field photography with a hand-held plenopic camera. *Technical Report CTSR 2005-02*, CTSR, 01 2005.

[50] N. Ouerhani, J. Bracamonte, H. Hugli, M. Ansorge, and F. Pellandini. Adaptive color image compression based on visual attention. In *Proceedings 11th International Conference on Image Analysis and Processing*, pages 416–421, 2001.

[51] 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*, pages 235–252, 2020.

[52] J. Park, Y.-W. Tai, D. Cho, and I.-S. Kweon. A unified approach of multi-scale deep and hand-crafted features for defocus estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2760–2769, 2017.

[53] F. Perazzi, P. Krähenbühl, Y. Pritch, and A. Sorkine-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, 2012.

[54] Y. Piao, X. Ji, M. Zhang, and Y. Zhang. Learning multi-modal information for robust light field depth estimation. *ArXiv*, abs/2104.05971, 2021.

[55] 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.

[56] 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 Conferences on Artificial Intelligence*, pages 904–911, 2019.

[57] 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 AAAI Conference on Artificial Intelligence*, pages 11865–11873, 2020.

[58] Y. Piao, Y. Zhang, M. Zhang, and X. Ji. Dynamic fusion network for light field depth estimation. *ArXiv*, abs/2104.05969, 2021.

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

[60] X. Qin, Z. Zhang, C. Huang, C. Gao, M. Dehghan, and M. Jägersand. Basnet: Boundary-aware salient object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7471–7481, 2019.

[61] Z. Ren, S. Gao, L. Chia, and I. Tsang. Region-based saliency detection and its application in object

recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, 24:769–779, 2014.

[62] U. Rutishauser, D. Walther, C. Koch, and P. Perona. Is bottom-up attention useful for object recognition? In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages II–II, 2004.

[63] E. Shelhamer, J. Long, and T. Darrell. Fully convolutional networks for semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39:640–651, 2017.

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

[65] X. Shi, Z. Chen, H. Wang, D. Yeung, W. Wong, and W. Woo. Convolutional lstm network: a machine learning approach for precipitation nowcasting. In *Proceedings of the Advances in Neural Information Processing Systems*, pages 802–810, 2015.

[66] 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, 2018.

[67] Y. Sugano, Y. Matsushita, and Y. Sato. Calibration-free gaze sensing using saliency maps. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2667–2674, 2010.

[68] J. Sun and H. Ling. Scale and object aware image retargeting for thumbnail browsing. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1511–1518, 2011.

[69] M. W. Tao, S. Hadap, J. Malik, and R. Ramamoorthi. Depth from combining defocus and correspondence using light-field cameras. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 673–680, 2013.

[70] M. W. Tao, P. P. Srinivasan, J. Malik, S. Rusinkiewicz, and R. Ramamoorthi. Depth from shading, defocus, and correspondence using light-field angular coherence. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1940–1948, 2015.

[71] X. Tian, K. Xu, X. Yang, B. Yin, and R. Lau. Weakly-supervised salient instance detection. In *Proceedings of the Conference on British Machine Vision Conference*, volume abs/2009.13898, 2020.

[72] A. Tsiami, P. Koutras, and P. Maragos. Stavis: Spatio-temporal audiovisual saliency network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4765–4775, 2020.

[73] 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*, 46:1083–1094, 2017.

[74] H. Wang, B. Yan, X. Wang, Y. Zhang, and Y. Yang. Accurate saliency detection based on depth feature of 3d images. *Multimedia Tools and Applications*, 77(12):14655–14672, 2018.

[75] L. Wang, H. Lu, Y. Wang, M. Feng, D. Wang, B. Yin, and X. Ruan. Learning to detect salient objects with image-level supervision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3796–3805, 2017.

[76] 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.

[77] T. Wang, A. A. Efros, and R. Ramamoorthi. Occlusion-aware depth estimation using light-field cameras. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3487–3495, 2015.

[78] T. Wang, Y. Piao, H. Lu, X. chun Li, and L. Zhang. Deep learning for light field saliency detection. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 8837–8847, 2019.

[79] W. Wang, J. Shen, L. Shao, and F. Porikli. Correspondence driven saliency transfer. *IEEE Transactions on Image Processing*, 25:5025–5034, 2016.

[80] W. Wang, J. Shen, R. Yang, and F. Porikli. Saliency-aware video object segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40:20–33, 2018.

[81] X. Wang, Y. Dong, Q. Zhang, and Q. Wang. Region-based depth feature descriptor for saliency detection on light field. *Multimedia Tools and Applications*, pages 1–18, 2020.

[82] X. Wang, S. You, X. Li, and H. Ma. Weakly-supervised semantic segmentation by iteratively mining common object features. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1354–1362, 2018.

[83] Y. Wei, J. Feng, X. Liang, M.-M. Cheng, Y. Zhao, and S. Yan. Object region mining with adversarial erasing: A simple classification to semantic segmentation approach. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6488–6496, 2017.

[84] Y. Wei, X. Liang, Y. Chen, X. Shen, M.-M. Cheng, Y. Zhao, and S. Yan. Stc: A simple to complex framework for weakly-supervised semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39:2314–2320, 2017.

[85] Y. Wei, W. Xia, J. Huang, B. Ni, J. Dong, Y. Zhao, and S. Yan. Cnn: Single-label to multi-label. *ArXiv*, abs/1406.5726, 2014.

[86] Z. Wu, L. Su, and Q. Huang. Stacked cross refinement

network for edge-aware salient object detection. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 7263–7272, 2019.

[87] Y. Xu, H. Nagahara, A. Shimada, and R. Taniguchi. Transcut: Transparent object segmentation from a light-field image. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3442–3450, 2015.

[88] 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.

[89] Y. Zeng, P. Zhang, J. Zhang, Z. Lin, and H. Lu. Towards high-resolution salient object detection. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 7233–7242, 2019.

[90] Y. Zeng, Y.-Z. 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 6067–6076, 2019.

[91] D. Zhang, D. Meng, L. Zhao, and J. Han. Bridging saliency detection to weakly supervised object detection based on self-paced curriculum learning. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 3538–3544, 2016.

[92] 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*, pages 8579–8588, 2020.

[93] 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.

[94] 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 Conferences on Artificial Intelligence*, pages 2212–2218, 2015.

[95] 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.

[96] 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 *European Conference on Computer Vision*, pages 374–390, 2020.

[97] 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.

[98] M. Zhang, J. Li, W. Ji, Y. Piao, and H. Lu. Memory-oriented decoder for light field salient object detection. In *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, pages 898–908, 2019.

[99] 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*, pages 3469–3478, 2020.

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

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

[102] Q. Zhang, S. Wang, X. Wang, Z. Sun, S. Kwong, and J. Jiang. A multi-task collaborative network for light field salient object detection. *IEEE Transactions on Circuits and Systems for Video Technology*, 31(5):1849–1861, 2020.

[103] X. Zhang, Y. Wang, J. Zhang, L. Hu, and M. Wang. Light field saliency vs. 2d saliency: A comparative study. *Neurocomputing*, 166:389–396, 2015.

[104] J. Zhao, Y. Cao, D.-P. Fan, M.-M. Cheng, X. yi 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 3922–3931, 2019.

[105] W. Zhao, F. Zhao, D. Wang, and H. Lu. Defocus blur detection via multi-stream bottom-top-bottom network. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42:1884–1897, 2020.

[106] T. Zhou, D.-P. Fan, M.-M. Cheng, J. Shen, and L. Shao. Rgb-d salient object detection: A survey. *Computational Visual Media*, pages 1–33, 2021.

[107] W. Zhu, S. Liang, Y. Wei, and J. Sun. Saliency optimization from robust background detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2814–2821, 2014.