开放场景下的显著性物体

范登平，张静，许刚，程明明，邵岭

Abstract—本文阐明并解决了一个现有显著性物体检测（SOD）数据集中存在的严重设计偏差，即不切实际地假设每张图像中都至少应包含一个清晰整洁的显著性物体。这种设计偏差导致当前最先进的 SOD 模型在现有数据集上进行评估时出现性能上的饱和。然而，在将这些模型应用于真实场景时，效果仍远不能令人满意。基于我们的分析，本文提出了一个新的高质量数据集，并更新了以前的显著性评测。具体而言，本文将该数据集称为开放场景下的显著性物体 (**SOC**)，它包含了从多种常见对象类别中选出的带有显著性物体和非显著性物体的图像。除了对象类别的标注外，每一张显著性图像还带有可以反映真实场景中常见挑战的属性，这些属性为更深入地了解 SOD 问题提供了支持。此外，给定显著性编码器，例如，在一些骨干网络中，现有的显著性模型是为了实现从训练图像集到真值图集的映射而设计的。因此，本文认为，与仅专注于解码器设计相比，改善数据集可获得更多的性能提升。考虑到这一点，本文研究了几种数据集增强策略，包括使用标签平滑隐式地强调显著边界，使用随机图像增强策略以使显著性模型适应各种情况，以及将自监督学习作为一种可从小型数据集中学习的正则化策略。大量的结果证明了这些技巧的有效性。我们还提供了一个 SOD 的全面评测，可在代码仓库中找到：http://dpfan.net/SOCBenchmark。

Index Terms—显著性物体检测，SOD，SOC，综述，数据集，基准

1 引言

本文考虑了显著性物体检测（SOD）的任务，该任务旨在检测场景中最引人注意的物体，然后为其提取精确到像素级的轮廓。SOD 的优点在于其有众多的应用，包括前景图评估 [3], [4], [5]，视觉追踪 [6], [7], [8]，动作识别 [9]，图像检索 [10], [11]，信息发现 [12], [13]，图像对比度增强 [14]，行人重识别 [15]，图像分割 [16], [17]，视频分割 [18]，照片组合 [19]，面向内容的图像编辑 [20]，图像摘要 [21]，以及视频压缩 [22], [23]，风格转换 [24], [25]，图像匹配 [26]，自主水下机器人 [27]，伪装物体检测 [28]，美学评分 [29]，自动驾驶汽车 [30]，植物种类鉴别 [31]，虚拟现实/增强现实 [32][1]，索尼的 BRAVIA XR 电视[2] 等。然而，现有的 SOD 数据集 [33], [34], [35], [36], [37], [38], [39], [40], [41], [42], [43] 在数据收集程序或数据质量方面存在缺陷。具体来讲就是，大多数数据集都假设一张图像内包含至少一个显著性物体，因此，它们会丢弃不包含任何显著性物体的图像。本文将其称为**数据选择偏差** [44]。

此外，现有数据集通常包含具有单个对象或几个整洁对象的图像。这些数据集无法充分反映现实世界图像的复杂性，因为现实场景通常为包含多个对象的情况。这种情况造成的结果就是，在现有的大规模数据集上 (例如, DUTS [42]) 训练的所有性能最高的模型都具有近乎饱和的性能（例如, SCRN [45] 在 ECSSD [38] 上取得了 *S-measure* > 0.9 的性能），但在真实图像上仍然无法获得令人满意的结果（例如, SOC [1] 上 *S-measure* < 0.8）。由于当前的 SOD 模型偏向理想条件，因此将其应用于实际场景后，其有效性可能会受到损害。为了解决此问题，有必要引入具有更实际条件的数据集。

Figure 1. SOC 数据集中的样本图像，包括非显著物体（第一行）和显著物体图像（第 2 到 4 行）。对于显著物体图像，本文提供了实例级的真值图（不同颜色表示不同实例）、物体属性和类别标签

RGB SOD 社区面临的另一个问题是，使用现有数据集只能分析模型的整体性能。这是因为没有数据集包含反映真实世界中不同挑战的属性。引入属性会对解决该问题有以下帮助：i）更深入地了解 SOD 问题，ii）研究 SOD 模型的利

- 这项工作的初始版本发表在 ECCV [1] 中。
- 这项工作的主要部分在南开大学完成。
- 本文是论文 [2] 的中文翻译版本。
- 通讯作者：程明明（cmm@nankai.edu.cn）。

1. 增强现实切割 & 粘贴：https://www.youtube.com/watch?v=-N-podTAY9Y.
2. https://www.youtube.com/watch?v=4LnCuTAlVno&feature=youtu.be.

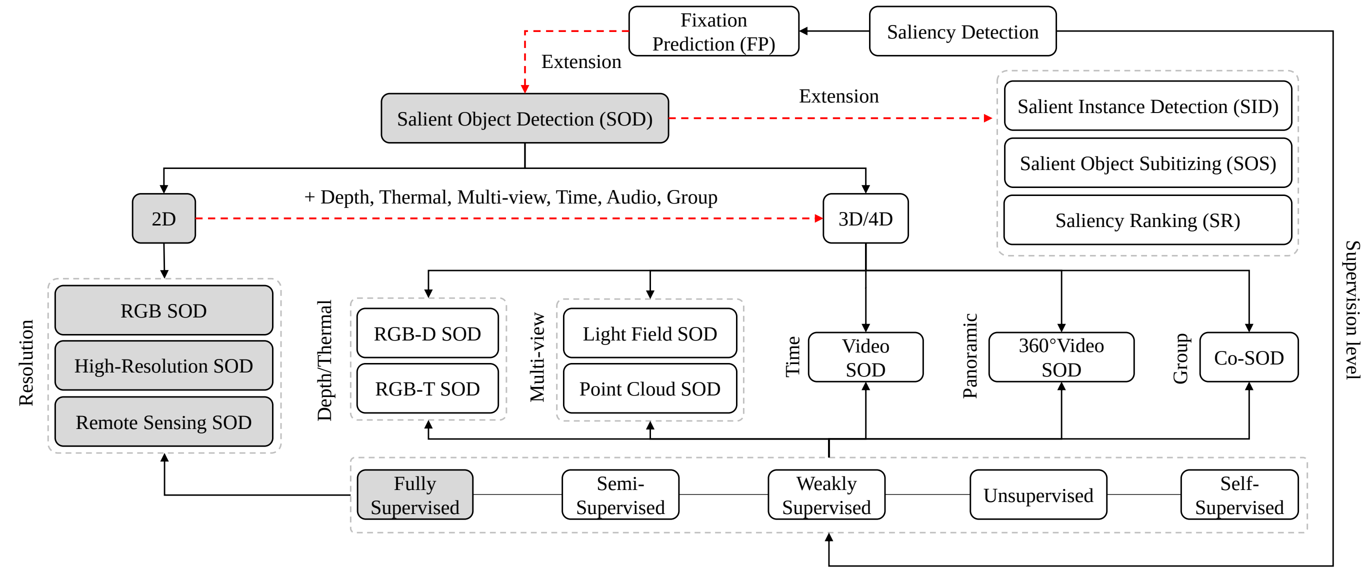


Figure 2. 显著性检测任务的分类。本文用灰色突出显示了这项研究的范围。更多细节详见章节2。

弊，iii）从不同角度客观地评估模型性能等。最后，在给定显著性编码器（例如，一些骨干网络中）的情况下，现有的显著性模型致力于建立从训练图像集到训练真值图集的映射。因此，本文认为，与仅专注于解码器设计相比，在数据集上进行改进，例如，修正数据偏差问题，可以获得更高的性能增益。

为此，本文研究了几种数据集增强策略，包括标签平滑以突出显著边界，随机图像增强以使显著性模型适应各种情况，以及将自监督学习作为从小数据集学习的一种正则化形式。大量的实验验证了这些技巧的有效性。

本文的贡献总结如下：

2) **数据集**。本文收集了一个新的高质量 SOD 数据集，名为“开放场景下的显著性物体”或 SOC。SOC 是迄今为止最大的实例级 SOD 数据集，包含 6,000 张来自 80 多个常见类别的图像。它与现有数据集在三个方面不同：i）显著性物体具有类别标注，可用于研究新问题，如，弱监督 SOD。ii）包含非显著性图像和对象使得该数据集比现有数据集更具现实性和挑战性。iii）显著性对象的属性反映了现实世界中遇到的各种情况，例如运动模糊，遮挡以及杂乱背景。从而 SOC 缩小了现有数据集与真实场景之间的距离。

2) **回顾 & 评测**。本文呈现了最大规模的 RGB SOD 研究，回顾了 201 个经典模型，包括 84 种使用手工特征的算法和 117 种基于深度学习的模型。此外，本文还维护了一个在线基准评测平台（即：http://dpfan.net/SOCBenchmark），以动态追踪该领域的发展。此外，本文提供了**前** 100 个 SOD 模型中最全面的基准评测。为了评估模型，本文不仅首次呈现了总体性能，而且还展示了一个基于属性的评测结果。这样可以更深入地了解模型从而提供更完整的评测。

3) **策略**。本文调研了有偏差的数据集的问题并引入了三种数据集增强策略；即使用标签平滑以隐式强调显著边界，使用随机图像增强以使显著性模型适应各种情况，以及将自监督学习作为一种可从小型数据集中学习基础上再策略。尽管本文的策略看似简单，但本文仍然可以在五个现有的最先进模型基础上再平均提高 S_α 指标 1.1%。

4) **讨论 & 未来方向**。基于本文的 SOC，我们呈现了现有 SOD 算法的优缺点，讨论了几个未被充分研究的开放性问题，并在六个层次上，例如，数据集层面，任务层面，模型层面，监督层面，评估层面和应用层面。提供了未来潜在的研究方向，

这项工作在以下几个方面扩展了本文以前的会议版本 [1]。首先，本文提供了本文 SOC 中的更多细节，包括没有显著性物体的样本图像，带有属性的图像以及属性的统计信息。第二，本文研究了三种与训练数据集相关的新颖策略，以充分利用非显著性对象数据并刷新了最先进的性能。第三，本文在 SOC 上进行了最大规模的 SOD 模型基准评测（46 个传统模型和 54 个深度学习模型）。最后，根据基准评测结果，本文重点介绍了 SOD 中的一些基础的研究方向和挑战。

2 相关工作

2.1 范围

显著性物体检测源自注视点预测（FP）[46], [47] 任务，随后将注意力区域转换为准确的对象级别区域。SOD 中可以追溯到的开创性工作有 [48], [49]。目前已经开发了用于有限分辨率（宽或高 < 500 像素），高分辨率（比如, 1080p, 4K）[50], [51]，甚至远程遥感数据 [52] 的 2D 图像算法。根据监督策略分类，SOD 模型有五种类型：全监督 [53]，半监督 [54]，弱监督 [55], [56], [57]，无监督 [58], [59], [60] 和自监督 [60], [61]。

Table 1
知名 SOD 数据集总结。本文的 SOC 是唯一一个满足所有要求的。根据 [77]，这些数据集分为三类：早期 (▲)，知名/先进 (♦)，和特殊 (♢)。可以在章节2.2中找到更多细节。

#	数据集名称	年份	出版商	High-Quality	$\geq 5k$	Non-Salient	Attribute	Category	Bounding Box	Object	Instance
1	MSRA-A, -B [33] ▲	2007	CVPR	✓	✓	-	-	-	✓	✓	-
2	SED1, SED2 [34] ▲	2007	CVPR	✓	-	-	-	-	-	✓	-
3	ASD [83] ▲	2009	CVPR	✓	-	-	-	-	-	✓	-
4	SOD [84] ♦	2010	CVPRW	✓	-	-	-	-	-	✓	-
5	MSRA10K [85] ♦	2011	CVPR	✓	✓	-	-	-	-	✓	-
6	Judd-A [37] ▲	2012	ECCV	✓	-	-	-	-	-	✓	-
7	DUT-O [39] ♦	2013	CVPR	✓	✓	-	-	-	✓	✓	-
8	ECSSD [38] ♦	2013	CVPR	✓	-	-	-	-	-	✓	-
9	PASCAL-S [40] ♦	2014	CVPR	✓	-	-	-	-	-	✓	-
10	HKU-IS [41] ♦	2015	CVPR	✓	-	-	-	-	-	✓	-
11	SOS [64] ♢	2015	CVPR	✓	✓	-	-	-	✓	-	-
12	MSO [64] ♢	2015	CVPR	✓	-	-	-	-	✓	-	-
13	XPIE [86] ♢	2017	CVPR	✓	✓	-	-	-	-	✓	-
14	ILSO [62] ♢	2017	CVPR	-	-	-	-	-	-	✓	✓
15	JOT [87] ♢	2017	FCS	✓	✓	✓	-	-	-	✓	-
16	DUTS [42] ♦	2017	CVPR	✓	✓	-	-	-	-	✓	-
17	SOC (OUR) ♦	2021		✓	✓	✓	✓	✓	✓	✓	✓

最近，还出现了一些 SOD 有趣的拓展，例如显著性实例检测（SID）[62], [63]，显著性物体感数（SOS）[64], [65], [66]，以及显著性排名 [67], [68]。图 2展示了显著性检测任务的分类方法。与以往的 SOD 回顾工作 [69], [70], [71], [72], [73], [74], [75], [76], [77] 不同，本文主要关注全监督条件下的 2D 显著性检测。本文用灰色突出显示了这项研究的范围。对于其它紧密相关的 3D / 4D SOD 任务，请读者参考最近的调查和基准评测工作，例如 RGB-D SOD [78], [79], Event-RGB SOD (ERSOD) [3], Light Field SOD [80], Co-SOD [81], 360°Video SOD [82], 和 Video SOD [18]。

2.2 SOD 数据集

在本节中，我们简要讨论现有的针对 SOD 任务设计的数据集，主要专注于以下方面：标注类型，每个图像的显著性对象数量，图像数量和图像质量。这些数据集在表 1中列出。

早期的数据集要么局限于图像数量，要么受限于显著性物体的标注质量。例如，MSRA-A [33] 和 MSRA-B [33] 中的显著性对象仅以边界框的形式进行标注。ASD [83], SED1 [34] 和 MSRA10K [36] 在大多数图像中仅包含一个显著性物体，同时，SED2 [34] 数据集在每张图中提供了两个物体但是仅包含 100 张图片。为了提高数据集的质量，近年来，研究人员已经开始在相对复杂和开放的背景下收集具有多个对象的图像。这些新的数据集包括 ECSSD [38], DUT-O [39], Judd-A [37], 和 PASCAL-S [40]。与以前的数据集相比，这些数据集在标注质量和图像数量方面都得到了改善。为了解决现阶段仍然存在的缺点，一些数据集 (例如, HKU-IS [41], XPIE [86], 和 DUTS [42]) 提供了大量的按像素标记的图像 (图 3.b)，每个图像有一个以上的显著性对象。然而，它们都忽略了非显著性物体 (图 1中的第一行) 并且没有提供实例级的标注 (图 3.c)。Jiang 等人 [87] 收集了大约 6K 张简单的背景图像（它们大多数是纯纹理图像）来覆盖非显著性场景。

然而，这个名为 JOT 的数据集，并不能表达现实场景的复杂性。如图 7中所示，ILSO [62] 中的数据集包含实例级的显著性检测标注，但是仅有大致标记的边界。除了“标准的” SOD 数据集，还有其它一些针对新任务的特殊数据集，例如显著目标感数比如, SOS [64] 和它的子集 MSO [64]) 等。

综上所述，现有的数据集主要集中在具有清晰显著性对象和简单背景的图像上。考虑到现有数据集的上述局限性，需要一个包含非显著性对象的、“自然”纹理和具有属性的显著性对象的更贴近实际的数据集，以供将来在该领域进行研究。这样的数据集可以更深入地了解 SOD 模型的优缺点，并有助于克服性能饱和问题。本文的 SOC 的独特之处在于它提供了各种高质量的标注，如表 1中所示。

(a) 图像 (b) 先前的方法 (c) 本文的方法(d) 图像分割数据集

Figure 3. 以前的 SOD 数据集仅通过在显著性物体周围绘制像素精确的轮廓来对图像进行标注 (b)。与物体分割数据集 [88] (d) 不同，本文的 SOC 提供了显著性实例 (c)。本文提供了一个高质量的、大规模标注并包含了能更好地呈现现实世界场景特性图像的数据集。

2.3 SOD 模型

我们注意到, 从 1998 年到 2021 年 2 月底, 已发表了 10,000 多篇有关显著性检测或相关领域的论文。在本节中，本文将尽力总结发表在顶级会议 (例如, NeurIPS, CVPR, ICCV, AAAI) 和期刊 (例如, TPAMI, TIP, TMM) 的文章，以及一些高质量的开源 (比如, arXiv) 工作。本文不采用描述每个模型的方式，而是总结这些模型的关键组件从而呈现出该领域的全局视图。

如表 2中所示，该领域的研究者已经设计了许多不同的方法，在不同级别的监督（例如，无监督，半监督和全监督）下使用超像素，目标提取或边缘/边界标注的方式来解决 SOD 问题。使用通用的聚合策略 (例如, 线性, 非线性)，这些方法主要关注像素，区域和块来设计功能更强大的模型。此外，本文注意到在这些方法中经常使用某些先验 (例如, 中心环绕先验, 局部/全局对比度先验, 前景/背景先验, 和边界先验)。一些模型还利用不同的后处理步骤 (例如, 条件随机场, 形态学, 分水岭, 和最大流策略) 来进一步提高性能。

3. ERSOD: https://github.com/jxr326/ERSOD-Net.

Table 2
使用手工特征的知名 SOD 模型总结。**Agg.:** 聚合策略，例如，LN = linear, NL = non-linear, HI = hierarchical, BA = Bayesian, AD = adaptive, LS = least-square solver, EM = energy minimization, and GMRF = Gaussian MRF. **SL.:** 监督等级，例如，无监督 (★)，半监督 (•)，弱监督 (☾)，全监督 (○)，主动学习 (A). **Sp.:** 是否使用超像素过分割技术。**Pr.:** 是否使用 Proposal 方法。**Ed.:** 是否使用边缘线索。**Post-Pros.:** 是否使用后处理方法 (例如，CRF [89], graph-cut [90], GrabCut [91], Ncut [92]), morphology, max-flow (MF) [93] 或仅使用阈值化。

	#	模型	出版商	谷歌	先验知识	独特性	组件	Agg.	SL.	Sp.	Pr.	Ed.	Post-Pros
2010 - 1998	1	Itti [46]	TPAMI	link	center-surround	pixel	Color, Intensity, Orientation	LN	★	-	-	-	-
	2	GBVS [94]	NeurIPS	link	-	pixel	Markovian	-	★	-	-	-	-
	3	FT [83]	CVPR	link	frequency domain	pixel	Color, Luminance	-	★	-	-	-	-
	4	SR [95]	CVPR	link	spectral residual	pixel	Log Spectrum	-	★	-	-	-	-
	5	AIM [96]	NeurIPS	link	maximizing information	patch	Shannon's Self-information	-	★	-	-	-	-
	6	SUN [97]	JOV	link	self-information	pixel	DoG, ICA-derived features	-	★	-	-	-	-
	7	FG [98]	MM	link	local contrast	pixel	Fuzzy Growing	-	★	-	-	-	-
	8	AC [99]	ICVS	link	local contrast	multi-patch	Color, Luminance	LN	★	-	-	-	-
	9	SEG [100]	ECCV	link	local contrast	pixel	Conditional Probabilistic	-	★	-	-	-	CRF
	10	MSSS [101]	ICIP	link	symmetric surround	pixel	Color, Luminance	-	★	-	-	-	graph-cut
	11	ICC [102]	ICCV	link	isophote	global structure	curvedness, isocenters, color	LN	★	-	-	-	graph-cut
	12	EDS [103]	PR	link	-	pixel	threshold, distance, multi-DoG	-	★	-	-	✓	-
	13	RE [104]	ICME	link	local contrast	pixel/patch	Contrast pyramid	-	★	-	-	-	-
	14	RSA [105]	MM	link	global contrast	patch	Polar transfer, NN-GPCA [105]	-	★	-	-	-	-
	15	RU [106]	TMM	link	rule based	pixel	denoising, geometric	-	★	-	-	-	-
	16	CSM [107]	MM	link	frequency&contrast	pixel	Envelope, Skeleton	-	★	-	-	-	-
2014 - 2011	17	LSSC [108]	TIP	link	bayesian	pixel/region	convex hull, subspace clustering	NL	★	✓	-	-	-
	18	COV [109]	JOV	link		pixel/patch	covariance matrices	NL	★	-	-	-	-
	19	GR [110]	SPL	link	contrast, center, smooth	-	convex hull, continuous pair	NL	★	✓	-	-	-
	20	MSS [111]	SPL	link	local, integrity, center	-	various gaussian, convex hull	NL	★	✓	-	-	-
	21	LSMD [112]	AAAI	link	texture, edge, color	pixel/region	hierarchical clustering, gaussian	-	★	✓	✓	-	threshold
	22	BSF [113]	ICIP	link	boundary-based	region	convex hull, soft-segmentation	-	★	✓	-	-	-
	23	HC [85]	CVPR	link	global contrast	region	Histogram-based Contrast	-	★	-	-	-	graph-cut
	24	RC [85]	CVPR	link	global contrast	region	Region-based Contrast	-	★	-	-	-	graph-cut
	25	CA [85]	CVPR	link	context-aware	patch	Four principles	-	★	-	-	-	-
	26	MR [39]	CVPR	link	fore/back-ground	pixel/region	graph-based manifold ranking	-	★	✓	-	-	-
	27	SF [114]	CVPR	link	element contrast	region	uniqueness, spatial	NL	★	-	-	-	-
	28	HS [38]	CVPR	link	global contrast	hi-region	Region-scale, Location heuristic	HI	★	-	-	-	-
	29	DRFI [115]	CVPR	link	background descriptor	region	region vector, multi-level	LN	○	✓	-	-	-
	30	RBD [116]	CVPR	link	background weighted	region	background connectivity	LS	★	✓	-	-	-
	31	LR [117]	CVPR	link	location, semantic, color	pixel/region	Low rank matrix	NL	○	✓	-	-	threshold
	32	PCA [118]	CVPR	link	center-bias priors	patch	color, pattern, gaussian	NL	★	✓	-	-	-
	33	HDCT [119]	CVPR	link	high-dimensional color	pixel	Trimap, color transform	LN	★	✓	-	-	-
	34	CRFM [120]	CVPR	link	aggregation	pixel	GIST descriptor	NL	○	-	-	-	CRF
	35	STD [121]	CVPR	link	statistical textural	region	Graph, sparse texture	-	★	-	-	-	GrabCut
	36	PDE [122]	CVPR	link	representative elements	region	color, background, center	-	★	✓	-	-	-
	37	SUB [123]	CVPR	link	Submodular	region	color, spatial, center	-	○	✓	-	-	threshold
	38	PISA [124]	CVPR	link	spatial	pixel/region	color, structure, orientation	NL	★	-	-	✓	-
	39	DSR [125]	ICCV	link	reconstruction errors	multi-region	background, obj./centerGaussian	BA	★	✓	-	-	-
	40	MC [126]	ICCV	link	markov random walks	region	Markov Chain	-	★	✓	-	-	-
	41	GC [127]	ICCV	link	global cue	region	GMM, appearance, spatial	AD	★	-	-	-	-
	42	SVO [128]	ICCV	link	center-surround	patch/region	Graph, Obj.	EM	★	✓	✓	-	-
	43	CSD [129]	ICCV	link	center-surround	multi-patch	color, orientation, intensity	LN	★	-	-	-	-
	44	UFO [130]	ICCV	link	focus, objectness	pixel/region	Uniqueness, Focusness, Obj.	NL	★	✓	✓	✓	threshold
	45	CHM [131]	ICCV	link	center-surround, local	mRegion/patch	SVM, hyperedge	LN	•	✓	-	✓	threshold
	46	CIO [132]	ICCV	link	objectness	Region	Graph, frequency, Obj.	GMRF	★	✓	-	-	-
	47	CC [133]	ICCV	link	convexity context	mRegion	concavity, bounding box	-	★	✓	-	-	graph-cut
	48	GS [134]	ECCV	link	boundary, connectivity	patch/region	Geodesic distance transform	-	★	✓	-	✓	-
	49	CB [135]	BMVC	link	context, shape, center	mRegion	Iterative energy minimization	LN	★	✓	✓	-	-
	50	SLMR [136]	BMVC	link	low-rank matrix	Region	sparse noise	-	★	✓	-	-	-
2018 - 2015	51	SMD [137]	TPAMI	link	texture, edge, color	pixel/region	hierarchical clustering, gaussian	-	★	✓	✓	-	threshold
	52	RS [138]	TPAMI	link	fore/back-ground	region	manifold ranking, grouping cue	-	★	✓	-	-	-
	53	BFS [139]	NC	link	fore/back-ground seed	region	Gaussian falloff, threshold	NL	★	✓	-	-	-
	54	GLC [140]	PR	link	global/local contrast	region	HOG, LBP, codebook,graph-cut	LN	★	✓	-	-	-
	55	DSP [141]	PR	link	propagation	region	sink points, chi-square distance	NL	★	✓	-	✓	-
	56	LPS [142]	TIP	link	label propagation-base	pixel/region	three-cue-center, affinity matrix	NL	★	✓	-	-	-
	57	MAPM [143]	TIP	link	background	region	Markov absorption probability		★	✓	-	-	-
	58	MIL [144]	TIP	link	instance	region	multi-instance learning, SVM	-	•	✓	✓	-	-
	59	RCRR [145]	TIP	link	reversion correction	pixel/region	regular-random walks ranking	-	★	✓	-	-	-
	60	FCB [146]	TIP	link	fore/back-ground, center	region	color difference, color volume	NL	★	✓	-	-	-
	61	NCS [147]	TIP	link	center bias	pixel/region	Ncut, merging scheme	EM	★	✓	-	✓	Ncut
	62	MDC [148]	TIP	link	direction contrast	pixel	OTSU, morphological filter	NL	★	-	-	-	watershed
	63	HCCH [149]	TIP	link	closure completeness & reliability	object	hierarchical segmentation	NL	★	-	-	✓	-
	64	JLSE [150]	TIP	link	exemplar-aided	region	joint latent space embedding	-	○	✓	-	-	-
	65	IFC [151]	TMM	link	boundary homogeneity	pixel/region	linear feedback control system	-	★	✓	-	-	-
	66	NIO [152]	TNNLS	link	smoothness, boundary	region	graph, iterative optimization	BA	•	✓	-	-	-
	67	MBS [153]	ICCV	link	barrier distance	pixel	backgroundness cue	-	★	-	-	-	morphology
	68	GP [154]	ICCV	link	diffusion based	region/pixel	diffusion/laplacian matrix	-	★	✓	-	-	-
	69	BSCA [155]	CVPR	link	color/space contrast	region/pixel	cellular automata, bayesian	-	★	✓	-	-	OTSU [156]
	70	BL [157]	CVPR	link	image prior	mRegion	SVM, MKB [158], LBP	LN	○	✓	-	-	-
	71	MST [159]	CVPR	link	geometry information	pixel	minimum spanning tree	-	★	✓	-	-	morphology
	72	RRWR [160]	CVPR	link	error-boundary removal	pixel/region	regular-random walks ranking	-	★	✓	-	-	-
	73	TLLT [161]	CVPR	link	propagation,boundary	region	convex hull, teach-to-learn	-	★	✓	-	-	-
	74	WSC [162]	CVPR	link	weighted sparse coding	region	color histogram, dictionary	NL	★	✓	-	-	-
	75	PM [163]	ECCV	link	propagation	region	extended random walk	LN	★	✓	-	-	-
2021 - 2019	76	TSG [164]	TCSVT	link	regionally spatial consistency	region	Sparse Representation, graph	LN	★	✓	-	-	MF
	77	LFCS [54]	TCSVT	link	smoothness, boundary	region	Discrete Linear Control System	LN	•	✓	-	-	-
	78	AIGC [165]	TCSVT	link	contrast, object	region	irregular graph	-	★	✓	-	-	-
	79	FTOE [166]	TMM	link	contrast, center, distribute	pixel/region	fuzzy theory, object enhancement	LN	★	✓	✓	-	-
	80	MSGC [167]	TMM	link	fore/back-ground seed	region	multi-scale, global cue	NL	★	✓	-	-	-
	81	SIA [168]	TMM	link	boundary, dhs [169]	-	Cellular Automation	BA	★	✓	-	-	-
	82	KSR [170]	TIP	link	trained on [33]	region	R-CNN, Rank-SVM, subspace	-	A	-	✓	-	-
	83	MSR [171]	TIP	link	boundary connectivity	region	MBD [172]	-	★	✓	-	-	OTSU
	84	LRR [173]	TIP	link	background	pixel/region	Celluar Automata [155], FCN32	Metric	★	✓	-	-	-

Table 3

使用基于深度学习的知名 SOD 模型总结。可以从表 2中获得更详尽的描述。MB = MSRA-B 数据集 [33]。M10K = MSRA-10K [36] 数据集。P-VOC2010 = PASCAL VOC 2010 语义分割数据集 [174]。CRF = 条件随机场。**点击学术链接将链接到特定作者的谷歌学术**。

	#	模型	出版商	谷歌学术	训练数量	训练集	骨干网络	SL.	Sp.	Pr.	Ed.	CRF
2015	1	SupCNN [175]	IJCV	link	800	ECSSD [38]	-	○	✓	-	-	-
	2	LEGS [176]	CVPR	link	340+3,000	PASCAL-S [40]+MB [33]	-	○	-	✓	-	-
	3	MDF [41]	CVPR	link	2,500	MB [33]	-	○	✓	-	✓	-
	4	MC [177]	CVPR	link	8,000	M10K [36]	GoogLeNet [178]	○	✓	-	-	-
2016	5	DSL [179]	TCSVT	link	(5,168+10,000)*80%	DUT-O [39]+M10K [36]	LeNet [180]/VGGNet16	○	✓	-	-	-
	6	DISC [181]	TNNLS	link	9,000	M10K [36]	-	○	✓	-	-	-
	7	DS [182]	TIP	link	10,000	M10K [36]	VGGNet [183]	○	✓	-	✓	✓
	8	SSD [184]	ECCV	link	2,500	MB [33]	AlexNet [185]	○	✓	✓	-	-
	9	CRPSD [186]	ECCV	link	10,000	M10K [36]	VGGNet	○	✓	-	-	-
	10	RFCN [187]	ECCV	link	10,103+10,000	P-VOC2010 [174]+M10K [36]	VGGNet	○	✓	-	✓	-
	11	MAP [188]	CVPR	link	∼5,500	SOS [64]	VGGNet	○	-	✓	-	-
	12	SU [189]	CVPR	link	15,000+10,000	SALI [190]+M10K [36]	VGGNet	○	-	-	-	✓
	13	RACD [191]	CVPR	link	10,565	DUT-O [39]+NJU [192]+NLP [193]	VGGNet	○	-	-	-	-
	14	ELD [194]	CVPR	link	9,000	M10K [36]	VGGNet	○	✓	-	-	-
	15	DHS [169]	CVPR	link	3,500+6,000	DUT-O [39]+M10K [36]	VGGNet	○	-	-	-	-
	16	DCL [195]	CVPR	link	2,500	MB [33]	VGGNet	○	✓	-	-	✓
2017	17	DLS [196]	CVPR	link	10,000	M10K [36]	VGGNet	○	✓	-	-	-
	18	MSRNet [62]	CVPR	link	(500+)2,500+2,500	(ILSO [62]+)MB [33]+HKU-IS [41]	VGGNet	○	-	✓	✓	✓
	19	SRM [197]	CVPR	link	10,553	DUTS [42]	ResNet50 [198]	○	-	-	-	-
	20	NLDF [199]	CVPR	link	2,500	MB [33]	VGGNet	○	-	-	✓	✓
	21	WSS [42]	CVPR	link	456K	ImageNet [200]	VGGNet	○	✓	-	✓	✓
	22	DSS [201]	CVPR	link	2,500	HKU-IS [41]+MB [33]	VGGNet	○	-	-	✓	✓
	23	FSN [202]	ICCV	link	10,000	M10K [36]	VGGNet	○	-	-	-	-
	24	SVF [203]	ICCV	link	10,000	M10K [36]	VGGNet	☾	✓	-	-	-
	25	UCF [204]	ICCV	link	10,000	M10K [36]	VGGNet	○	-	-	-	-
	26	AMU [205]	ICCV	link	10,000	M10K [36]	VGGNet	○	-	-	✓	-
2018	27	EAR [206]	TCYB	link	2,500+2,500	HKU-IS [41]+MB [33]	VGGNet16	○	-	-	-	-
	28	Refinet [207]	TMM	link	3,000	MB [33]	VGGNet16	○	✓	-	✓	✓
	29	LICNN [208]	AAAI	link	456K	ImageNet [200]	VGGNet	○	-	-	-	-
	30	ASMO [55]	AAAI	link	82,783+2,500+2,500	MsCO [88]+HKU-IS [41]+MB [33]	ResNet101	○	-	-	-	✓
	31	RADF [209]	AAAI	link	10,000	M10K [36]	VGGNet	○	-	-	-	✓
	32	R3Net [210]	IJCAI	link	10,000	M10K [36]	ResNeXt [211]	○	-	-	-	✓
	33	C2SNet [212]	ECCV	link	20,000+10,000	Web [212]+M10K [36]	VGGNet	○	✓	✓	-	-
	34	RAS [213]	ECCV	link	2,500	MB [33]	VGGNet	○	-	-	-	-
	35	LPSNet [214]	CVPR	link	10,553	DUTS [42]	VGGNet16	○	-	-	-	-
	36	RSOD [215]	CVPR	link	425	PASCAL-S [40]	ResNet101	○	-	✓	-	-
	37	DUS [59]	CVPR	link	2,500	MB [33]	ResNet101	☾	-	-	-	-
	38	ASNet [216]	CVPR	link	15,000+10,000+5,168	SALI [190]+M10K [36]+DUT-O [39]	VGGNet	○	-	-	-	-
	39	BMPM [217]	CVPR	link	10,553	DUTS [42]	VGGNet	○	-	-	-	-
	40	DGRL [218]	CVPR	link	10,553	DUTS [42]	ResNet50	○	-	-	-	-
	41	PiCA [219]	CVPR	link	10,553	DUTS [42]	VGGNet16/ResNet50	○	-	-	-	✓
	42	PAGRN [220]	CVPR	link	10,553	DUTS [42]	VGGNet19	○	-	-	-	-
2019	43	SE2Net [221]	arXiv	link	10,553	DUTS [42]	VGGNet/ResNeXt101	○	-	-	-	-
	44	DRMC [222]	arXiv	link	10,533	DUTS [42]	VGGNet/ResNet101	○	-	-	-	✓
	45	RDSNet [223]	arXiv	link	10,000+10,553	M10K [36]+DUTS [42]	VGGNet/ResNet-152	○	-	-	-	✓
	46	AADF [224]	TCSVT	link	10,553	DUTS [42]	DenseNet161 [225]	○	-	-	-	-
	47	CCAL [226]	TMM	link	9,000	M10K [36]	VGGNet	○	-	-	-	-
	48	DeepUSPS [60]	NeurIPS	link	2,500	MB [33]	DRN-network [227]	☾	-	-	-	-
	49	FBG [228]	TIP	link	2,500	MB [33]	VGGNet16	○	-	-	✓	-
	50	SPA [229]	TIP	link	4,000	HKU-IS [41]	-	○	✓	-	-	✓
	51	ConnNet [230]	TIP	link	2,500+2,500	MB [33]+HKU-IS [41]	ResNet50	○	-	-	-	-
	52	LFRWS [231]	TIP	link	10,000	M10K [36]	VGGNet16	○	-	-	✓	-
	53	RSR [67]	TPAMI	link	425	Extended of PASCAL-S [40]	ResNet101	○	-	-	-	-
	54	SSNet [232]	TPAMI	link	10,000	M10K [36]	VGGNet16	☾	✓	-	-	-
	55	LVNet [233]	TGRS	link	600	ORSSD [233]	-	○	-	-	-	-
	56	Deepside [234]	NC	link	2,500+10,553	MB [33]+DUTS [42]	VGGNet16	○	✓	-	-	-
	57	SuperVAE [235]	AAAI	link	-	-	VGGNet19	☾	✓	-	-	-
	58	DEF [236]	AAAI	link	10,553	DUTS [42]	ResNet101	○	-	-	-	-
	59	CapSal [56]	CVPR	link	82,783+5,265	MsCO [88]+COCO-CapSal [56]	ResNet101	☾	-	-	-	-
	60	MWS [237]	CVPR	link	300,000+10,553	ImageNet [200]+DUTS [42]	-	☾	✓	-	-	✓
	61	MLMS [238]	CVPR	link	10,553	DUTS [42]	VGGNet16	○	-	-	✓	-
	62	ICNet [239]	CVPR	link	10,000	M10K [36]	VGGNet16/ResNet50	○	-	-	-	✓
	63	AFNet [240]	CVPR	link	10,533	DUTS [42]	VGGNet16	○	-	-	✓	-
	64	PFANet [241]	CVPR	link	10,553	DUTS [42]	VGGNet16	○	-	-	✓	-
	65	PAGE [242]	CVPR	link	10,000	M10K [36]	VGGNet16	○	-	-	✓	✓
	66	CPD [243]	CVPR	link	10,533	DUTS [42]	VGGNet/ResNet50	○	-	-	-	-
	67	PoolNet [244]	CVPR	link	10,533	DUTS [42]	VGGNet/ResNet	○	-	-	✓	-
	68	BASNet [245]	CVPR	link	10,553	DUTS [42]	ResNet34/Xavier [246]	○	-	-	✓	-
	69	JDF [247]	ICCV	link	2,500	MB [33]	VGGNet16	○	-	-	✓	-
	70	DPOR [248]	ICCV	link	10,533	DUTS [42]	VGGNet16	○	-	-	-	-
	71	JLNet [249]	ICCV	link	10,582+10,533	P-VOC2010 [174]+DUTS [42]	DenseNet169	○	-	-	-	✓
	72	GLFN [51]	ICCV	link	1,600+10,533	HRSOD [51]+DUTS [42]	VGGNet	○	-	-	-	✓
	73	SIBA [250]	ICCV	link	10,533	DUTS [42]	ResNet50	○	-	-	✓	-
	74	SCRNet [45]	ICCV	link	10,533	DUTS [42]	ResNet50	○	-	-	✓	-
	75	EGNet [251]	ICCV	link	10,533	DUTS [42]	VGGNet/ResNet	○	-	-	✓	-
2020	76	HUAN [252]	TIP	link	10,553	DUTS [42]	VGGNet/ResNet/ResNetXt	○	-	-	-	✓
	77	ALM [253]	TIP	link	10,000+4,447	M10K [36]+HKU-IS [41]	DenseNet	○	✓	-	-	-
	78	HFFNet [254]	TIP	link	10,553	DUTS [42]	VGGNet16	○	-	-	✓	-
	79	DFI [255]	TIP	link	10,553	DUTS [42]	ResNet50	○	-	-	✓	-
	80	R2Net [256]	TIP	link	10,553	DUTS [42]	VGGNet16	○	-	-	-	-
	81	MRNet [257]	TIP	link	10,553	DUTS [42]	ResNet50	○	-	-	-	-
	82	CIG [258]	TIP	link	10,000	M10K [36]	VGGNet16	○	-	-	✓	-
	83	RASNet [259]	TIP	link	2,500	MB [33]	VGGNet16	○	-	-	-	-
	84	ASNet [260]	TPAMI	link	15,000+10,000+5,168	SALI [190]+M10K [36]+DUT-O [39]	VGGNet	○	-	-	-	-
	85	DNNet [261]	TCYB	link	2,500+2,500	MB [33]+HKU-IS [41]	-	○	-	-	-	-
	86	CAANet [262]	TCYB	link	10,553	DUTS [42]	VGGNet16	○	-	-	-	-
	87	ROSA [263]	TCYB	link	2,500+5,168+2,500	HKU-IS [41]+DUT-O [39]+MB [33]	FCN [264]	○	✓	-	-	-
	88	DSRNet [265]	TCSVT	link	10,553	DUTS [42]	DenseNet	○	-	-	-	-
	89	EGNL [266]	TCSVT	link	2,500	MB [33]	VGGNet16	○	-	-	✓	-
	90	SACNet [267]	TCSVT	link	10,553	DUTS [42]	ResNet101	○	-	-	-	-
	91	FLGC [268]	TMM	link	10,553	DUTS [42]	VGGNet16	○	-	-	-	-
	92	TSNet [269]	TMM	link	4,000	MD4K [269]	ResNet50/VGGNet16	○	-	-	-	-

Table 4
基于深度学习的知名 SOD 模型的总结。可以在表2 & 3中找到更详尽的描述。

	#	模型	出版商	谷歌学术	训练数量	训练集	骨干网络	SL.	Sp.	Pr.	Ed.	CRF
2020	93	SUCA [270]	TMM	link	10,553	DUTS [42]	ResNet50	○	-	-	-	-
	94	MIJR [271]	TMM	link	2,500+5,000	MB [33]+DUTS [42]	VGGNet16	○	✓	✓	-	✓
	95	CAGVgg [272]	PR	link	10,553	DUTS [42]	VGGNet/ResNet/NASNet [273]	○	-	-	-	-
	96	U2Net [274]	PR	link	10,553	DUTS [42]	UNet	○	-	-	-	-
	97	SalGAN [275]	TII	link	10,000	M10K [36]	VGGNet16	○	-	-	-	-
	98	ADA [276]	AAAI	link	2,500+780	MB [33]+NIR [276]	VGGNet16	○	-	-	-	-
	99	PFPNet [277]	AAAI	link	10,553	DUTS [42]	ResNet101	○	-	-	-	-
	100	GCPANet [278]	AAAI	link	10,553	DUTS [42]	ResNet50	○	-	-	-	-
	101	F3Net [279]	AAAI	link	10,553	DUTS [42]	ResNet50	○	-	-	✓	-
	102	LDF [280]	CVPR	link	10,553	DUTS [42]	ResNet50	○	-	-	✓	-
	103	ITSD [281]	CVPR	link	10,553	DUTS [42]	VGGNet16/ResNet50	○	-	-	✓	-
	104	SANet [57]	CVPR	link	10,553	DUTS [42]	VGGNet16	☾	-	-	✓	✓
	105	MINet [282]	CVPR	link	10,553	DUTS [42]	VGGNet16/ResNet50	○	-	-	-	-
	106	ABPNet [283]	ECCV	link	10,553	DUTS [42]	VGGNet16	○	-	-	✓	-
	107	CSNet [284]	ECCV	link	10,553	DUTS [42]	-	○	-	-	-	-
	108	GateNet [285]	ECCV	link	10,553	DUTS [42]	VGGNet16	○	-	-	-	✓
2021	109	DNA [286]	TCYB	link	10,553	DUTS [42]	VGGNet16/ResNet50	○	-	-	-	-
	110	DAFNet [52]	TIP	link	1,400	EORSSD [52]	VGGNet16	○	-	-	✓	-
	111	HGA [287]	TIP	link	10,553	DUTS [42]	VGGNet16	○	-	-	✓	-
	112	HIRN [288]	TIP	link	10,553	DUTS [42]	VGGNet16	○	-	-	✓	-
	113	SCWS [289]	AAAI	link	10,553	SDUTS [57]	ResNet50	☾	-	-	-	-
	114	PFS [290]	AAAI	link	10,553	DUTS [42]	ResNet50	○	-	-	✓	-
	115	KRNet [291]	AAAI	link	10,553	DUTS [42]	ResNet50	○	-	-	✓	-
	116	BAS [32]	arXiv	link	10,553	DUTS [42]	ResNet34	○	-	-	✓	-
	117	ICON [53]	arXiv	link	10,553	DUTS [42]	ResNet50	○	-	-	-	-

最近，研究者已经提出了许多基于不同网络体系结构的深度学习 SOD 模型，例如多层感知器，全卷积网络（FCN），混合网络以及胶囊网络，它们的性能要高于传统方法。根据学习范式，大多数深度 SOD 模型可以大致分为两种类型：单任务学习和多任务学习方法。本文在表3 和表 4中总结了训练数据，骨干网络和其它组件。

本文主要关注宏观统计，而非微观描述。敬请读者参考最近的架构回顾 [77]。本文希望这份全面的综述[4]可以为这个快速发展的领域中后来的研究人员提供指导。

2.4 针对深度模型的数据集增强策略

现有的深度 SOD 模型专注于设计有效的解码器 [45], [52], [261], [262], [265], [279], [286] ，以聚合来自骨干网络 [198], [211], [292] 不同层次的特征。本文认为，当它们使用从输入训练图像集到输出训练真值图集的映射函数时，深层模型也应该关注数据集增强策略以提高模型泛化能力。这三种不同的策略已经被广泛地研究，包括标签平滑 [293]，图像增强 [294], [295] 和自监督学习 [296]。

与直接采用一位有效监督学习不同，“标签平滑”技术从平滑的监督中学习，因此可以使用生成的平滑标签 [293] 或干扰标签 [297] 来松弛监督信号。Miyato 等人 [298] 将局部扰动应用于数据点以增加模型分布的平滑度。为了获得一个更鲁棒，泛化性能更好的模型，Xie 等人 [297] 在每次迭代中将标签的一部分随机替换为不正确的值。另外，Wager 等人 [299] 证明了用指数族中已知分布的噪声破坏训练示例可以为判别模型注入适当的生成假设，从而减少了泛化误差。Peterso 等人提出了一个软标签数据集 (CIFAR10H [300]) ，旨在通过提供跨类别的标签分布而不是硬独热编码标签来反映人类感知的不确定性。

4. 研究小组: https://github.com/DengPingFan/Saliency-Authors.

图像增强 [294] 是一种可以扩展训练数据集的多样性，从而提高模型的泛化能力的有效技术。现有的图像增强技术可以大致分为两类：1）人工设计的策略，例如，旋转或尺度转换，和 2) 学习得到的策略 [301], [302]。对于第一类，其将预定义的数据扩充策略应用于数据集。除了广泛使用的旋转和尺度变换外，该类别中其它被广泛研究的方法是擦除技术 [303], [304]，该技术通过随机擦除部分图像块来实现数据增强。此外，混合方法 [305], [306] 利用混合数据增强策略从现有训练数据集中生成新样本，以减轻预测中的不确定性。对于第二类 [301]，网络会根据图像条件学习数据增强策略，该策略通常由深度神经网络进行参数设置。以这种方式，将图像输入数据增强网络，生成具有超参数控制数据增强程度的增强样本。

自监督学习 [296], [307]，也称为一致性学习，它定义了一种无标注的辅助任务，以提供用于特征学习的替代监督信号。传统上，自监督学习用于无监督表示学习，用于学习图像或视频的特征嵌入。最近，有工作将自监督学习定义为辅助任务，并在弱监督 [289] 或半监督 [308] 学习框架中使用它。最新的几个代表性工作可以在 [309], [310], [311] 中找到。

据本文所知，现有的显著性物体检测工作没有在探索数据集增强策略的过程中关注数据集偏差问题。在本文中，我们认为在通过挖掘数据集的增强策略也可以带来显著的性能提升。而且，这些解决方案是通用的，可以轻松地应用于现有的显著性检测网络。

3 SOC 数据集

在本节中，将介绍本文新的、旨在详细反应真实世界场景的、具有挑战性的 SOC 数据集。来自 SOC 样例图像如图 1所示，此外，关于 SOC 的类别和属性信息的统计信息分别如图 4 (a) 和图 6所示。基于现有数据集的优点和缺点，本文确定了全面和平衡的数据集应该满足的七个关键因素。

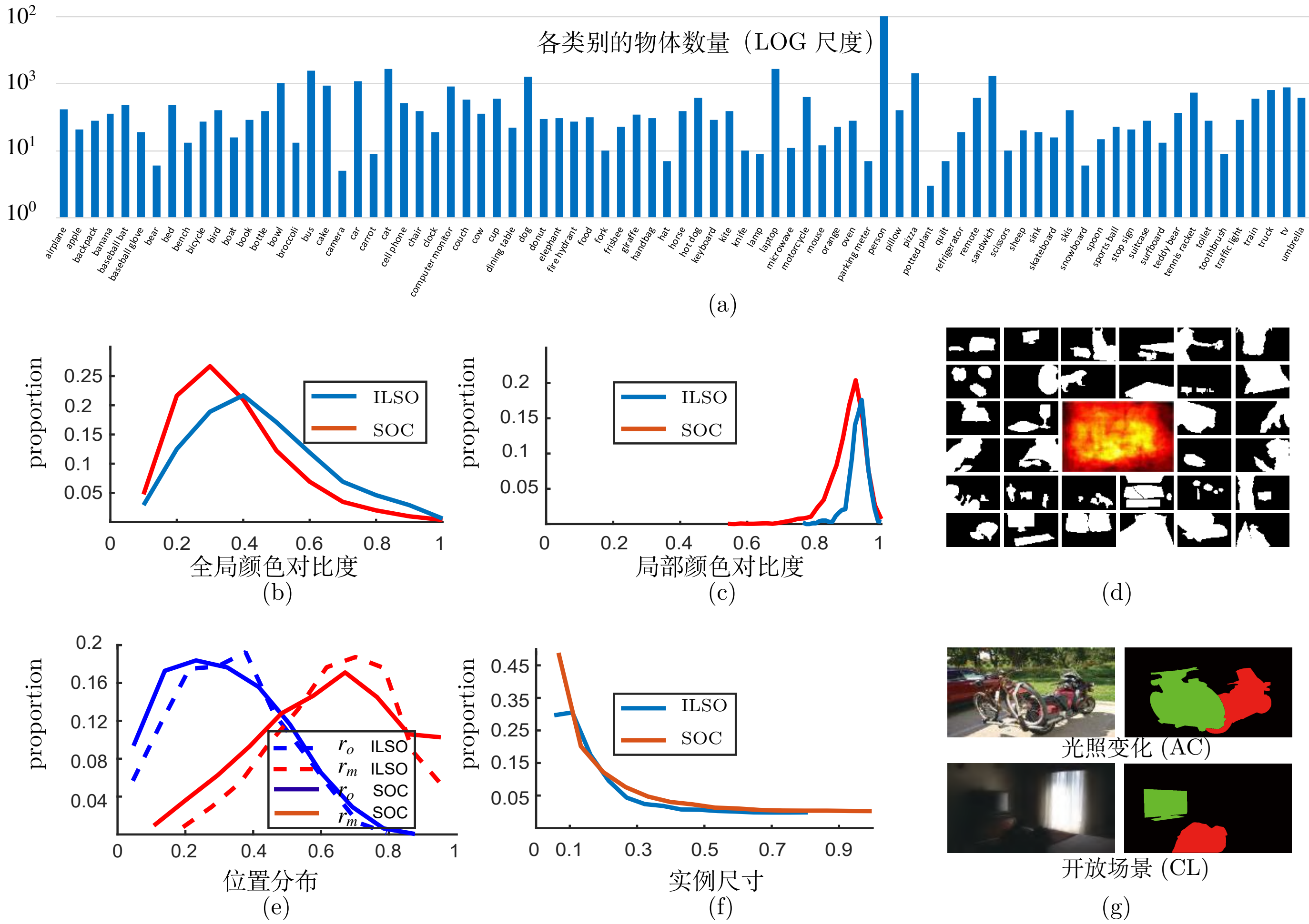


Figure 4. (a) 本文的 SOC 数据集中每个类别标注的实例数量。(b, c) 全局颜色对比度和局部颜色对比度的统计数据。(d) 一组来自本文数据集的显著图及其叠加图。(e) SOC 中的显著物体的位置分布。(f) SOC 和 ILSO [62] 的实例大小分布。(g) 不同属性的可视化例子。

1) **非显著物体的存在**。几乎所有的现有的 SOD 数据集都假设图像包含至少一个显著物体并丢弃了不包含显著物体的图像 [87]。然而，这种假设是一种会导致*数据选择偏差*的过于理想化的设定。在真实场景的设定中，图像并不总是包含显著物体。例如，一些无定型的背景图像，如天空，草地和纹理等场景中根本不包含显著物体 [312]。非显著物体或背景“元素”可能占据整个场景，因此严重限制了显著物体的可能位置。Xia 等人 [86] 通过判断什么是显著物体和什么不是显著物体，提出了先进的 SOD 模型，这说明非显著物体对推理显著物体至关重要。这也表明非显著物体和显著物体在 SOD 中应受到同等的重视。包含一定数量的非显著物体的图像会使得数据集更接近真实场景，同时也使得 SOD 任务变得更有挑战性。本文将“非显著物体”定义为没有显著物体的图像或具有“东西”性质的图像。如 [86], [312] 中所述，“东西”类别包括（a）密集分布的相似物体，（b）形状模糊，和（c）没有语义的区域，分别如图 5 (a)-(c) 所示。

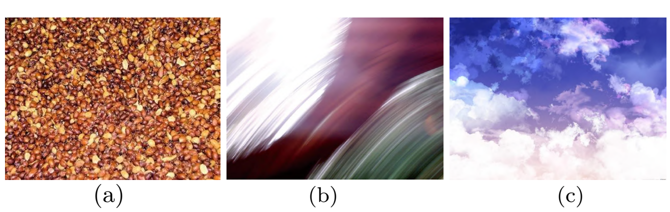


Figure 5. 一些非显著图像的示例 a) 拥挤场景，b) 运动模糊，c) 没有感兴趣区域的背景.

为了防止数据选择偏差，本文与 Torralba 和 Efros [44] 的提议一样，自动随机地来选择图像。基于非显著物体的定义，本文从 DTD [313] 数据集中收集了 783 个纹理图像。为了丰富多样性，又从互联网和其它数据集中收集了 2217 幅图像，包括极光，天空，人群，商店以及许多其它类型的真实场景 [35], [40], [87], [88]。

2) **图像的数量和类别**。相当数量的图像对于捕捉现实世界场景的多样性和丰富性至关重要。此外，大量的数据可以让 SOD 模型避免过拟合并增强泛化性能。为此，本文首先从 MS-COCO 数据集 [88] 中随机采集了 3,000 张图片，其中包含“自然环境中常见对象的日常场景”。随后，本文为 80 个对象类别进行了标注（参见补充材料）。请注意，和 [44] 中讨

Table 5
显著性物体图像属性及其对应描述列表。通过观察现有数据集的特征，本文总结了这些属性。可以在图 1 和 图 4(g) 中找到一些可视化示例。有关更多示例，请参阅补充材料。

属性	描述
AC (**光照变化**)	物体区域中明显的光照变化。
BO (**大物体**)	物体面积和图像面积的比值大于 0.5。
CL (**开放环境**)	物体周围的前景和背景区域具有相似的颜色，本文将全局颜色对比度大于 0.2，局部颜色对比度小于 0.9 的图像标记为开放环境图像。(章节3)。
HO (**异构物体**)	由视觉上独特或不相似的部分组成的物体。
MB (**运动模糊**)	由于相机或运动的抖动使得物体具有模糊的边界。
OC (**遮挡**)	物体被部分或全部遮挡。
OV (**超出视野**)	物体的部分区域超出了图像边界。
SC (**形状复杂**)	物体有复杂的边界，如纤细的组件 (例如，动物的脚) 和洞等。
SO (**小物体**)	物体面积和图像面积的比值小于 0.1。

论的那样，本文将数据选择与标记的过程分开，以避免出现数据选择偏差。请参考小节“7) 高质量的显著对象标签”获取更多信息。图 4 (a) 展示了每个类别的显著物体的数量。它表明“人”类别占很大比例，这是合理的，因为人们通常与其它对象一起出现在日常场景中。本文将数据集按照 6:2:2 的比例分为训练集，验证集和测试集。

3) **显著物体的全局/局部颜色对比**。如 [40] 中所述，术语“显著”与前景和背景的全局/局部对比度有关。因此，检查显著物体是否易于检测是非常重要的。对于每个物体，本文分别计算前景和背景的 RGB 颜色直方图。然后，利用 χ^2 距离来测量两个直方图之间的距离。全局和局部颜色对比度分布分别如图 4 (b) 和 (c) 所示。与 ILSO 相比，本文的 SOC 中低全局和局部颜色对比度的物体占据更大的比例。

4) **显著物体的位置**。*中心偏差*被认为是显著性检测数据集中影响最大的偏差之一 [40], [70], [314]。图 4 (d) 展示出了一组图像及其叠加图（比如，平均掩码图）。可以看出，虽然显著的物体位于不同的位置，但是叠加图仍然表明这组图像是存在中心偏差的。不幸的是，以前的基准评测通常采用这种不准确的方式来分析显著物体的位置分布。为了避免这种误导现象，本文绘制了图 4 (e) 中两个量 r_o 和 r_m 的统计情况，其中 r_o 和 r_m 分别表示物体中心和物体中最远（边缘）点离图像中心的距离。将 r_o 和 r_m 除以图像对角线长度的一半以进行归一化，使得 $r_o, r_m \in [0,1]$。从这些统计数据中，本文可以观察到数据集中的显著物体不受中心偏差的影响。

5) **显著物体的大小**。每个显著物体实例的大小被定义为物体面积占图像总面积的比例 [40]。如图 4 (f) 所示，与仅有的实例级数据集 ILSO [62] 相比，SOC 中的显著物体的大小变化范围更广泛。此外，SOC 中的中型物体具有更高的比例。

6) **具有属性的显著对象**。在数据集中，图像的属性信息有助于研究者客观评估模型在不同参数和变量下的性能。研究者还可以对模型失败的情况进行检查。为此，本文定义了一组属性来表示真实场景中面临的特定情况，如*运动模糊*，*遮挡* and *开放背景*等 (在 表 5中总结)。请注意，因为这些属性不是互斥的，所以一个图像可以使用多个属性进行标注。

受到 [315] 的启发，本文在图 6左边展示了数据集图片属

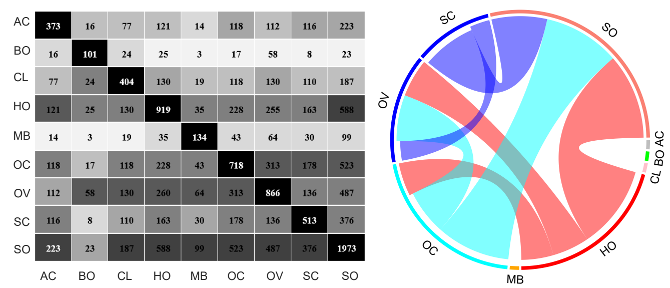
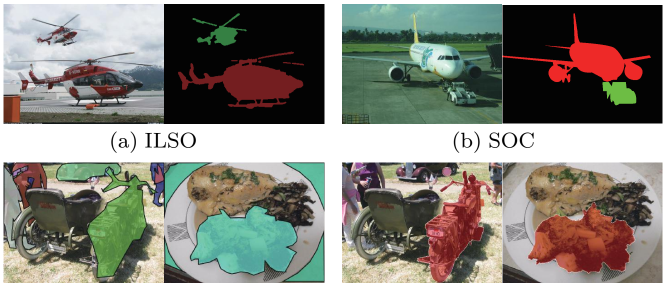


Figure 6. **左：SOC 数据集中显著性图像的属性分布。网格中的每个数字表示图像的出现次数。右：基于出现频率绘制的属性之间的主要依赖关系。宽度较大的连接表示该属性对其它属性的依赖较高。**

性的分布情况。SO 类型因为本文精确的实例级标注（比如，图 3中的网球拍）而占有最大的比例。HO 属性因为现实世界的场景由不同视觉特色的材料组成，也占有很大比例。*运动模糊*（MB 在视频帧中比静态图像更常见，但是偶尔也会在静态图像中出现。因此，MB 类型在本文的数据集中占有相对较小的比例。由于真实图像通常包含多个属性，为此本文在图 6的右侧根据出现频率展示了属性之间的依赖关系。例如，包含很多异构物体的场景可能具有大量的彼此间的遮挡，从而形成复杂的空间结构。因此，HO 类型和 OC 类型，OV 类型和 SO 类型之间具有强依赖性。

7) **高质量的显著对象标签**。正如 [316] 中提到的，在 ECSSD 数据集（包含 1000 张图像）上的训练比使用其它数据集（例如，MSRA10K，包含 10000 张图像）获得了更好的效果。因为除了规模以外，数据集质量也是一个重要因素。为了获得大量高质量的图像，本文从 MS-COCO 数据集 [88]，上随机选择图像，MS-COCO 是一个大型真实世界数据集，其中的物体用多边形标注（比如，粗略标注）。高质量标注在提高 SOD 模型的准确性方面也起到了关键作用 [83]。为此，本文使用逐像素的标注来重新标注数据集。类似于著名 SOD 任务导向的评测数据集 [33], [34], [35], [36], [38], [41], [42], [62], [83], [86], [87]，本文没有使用眼动仪设备。本文采用了两个步骤来保证高质量的标注：(i) 要求 5 个观众使用标定框标记他们认为在每个图像中较为显著的物体。(ii) 保留大多数（≥ 3）



(a) ILSO (b) SOC (c) MS-COCO (d) SOC

Figure 7. 与最近提出的数据集的比较，(a) 实例级数据集 ILSO [62]（用不连续的粗略边界标注），(c) MS-COCO 数据集 [88]，(b，d) 本文的 SOC 数据集，标注边界更平滑，质量更高。

观众在显著性上意见相同的物体（IOU of the bbox > 0.8）。第一阶段后，我们得到了 3000 个用标定框标注的显著性物体图像。第二阶段，本文根据标定框的提示进一步手工标记显著物体的逐像素轮廓。请注意，本文共有 10 名志愿者参与了整个步骤以交叉检验标注的质量。最后，本文保留了 3000 张具有高质量，实例级标记的显著物体的图像。如图 7 (b & d) 所示，本文的物体边界的标注是精确、清晰和平滑的。在标注过程中，本文还添加了一些未在 MS-COCO 数据集 [88] 中标记的新类别（例如，电脑显示器，帽子，枕头等）。

4 本文的数据增强策略

与致力于设计强大的用于特征聚合的解码器策略不同，本文引入了三种简单的数据集增强策略来实现更好的模型泛化能力。我们认为，本文提出的数据增强策略很容易在现有的全监督的 SOD 模型中实施，并且仅需少量修改即可获得良好的性能。本文将 RGB 显著性训练数据集定义为 $D = \{x_i, y_i\}_{i=1}^{N}$，其中 x_i, y_i 为输入的 RGB 图像以及对应的显著性真值图，记 i 为训练图像编号，N 是训练数据集的大小。由于 SOD 是一个二分类预测任务，显著性真值图 y 往往是一个二值结果图，并且大多数 SOD 技术使用二值（或带权）交叉熵损失函数来验证显著性预测结果。与将显著性真值图看作一个二值分割结果图不同，本文首次引入了“标签平滑” [293] 作为实现模型高效训练和高模型性能的有效技术。然后，本文采用随机图像增强来生成各种样本，以获得更好的模型泛化能力。最后，作为一种在半监督或无监督学习中被广泛研究的技术 [296], [307]，本文将自监督学习解决方案扩展到全监督的 SOD，从而建立了一个鲁棒的模型。

4.1 标签平滑

标签平滑与知识蒸馏。应用标签平滑时，最重要的方案之一便是使用用于知识蒸馏的师生网络 [317] 。通常，在一个师生网络中，教师模型具有很强的学习能力，而学生模型则具有较低的学习能力。然后，教师模型通过为学生模型提供“软目标”的方式来教授学生模型。正如在 [318] 中讨论的，“软目标”包含丰富的数据相似性结构，这对于产生增强的学生模型至关重要。此外，标签平滑可被视为输出分配正则化的一种形式，可以防止网络过度拟合。正如在 [293] 中指出的，硬标签可能会导致过拟合，因为模型将为每个类别分配完全概率，因此不能保证泛化性能很好。通过使用软标签，模型可以学习数据的结构，从而防止数据高估。遵循相同的数据设置，例如，使用标签平滑 [319]，引入在线标签平滑解决方案，从而根据模型的预测逐渐更新软标签。

常规设置。给定输入图像 x 和它对应的显著性真值图 y，传统的深度显著性模型 f_θ 通过最小化交叉熵损失来训练模型：$\mathcal{L}_{ce}(y,s) = -\sum_{i=1}^{N}\sum_{u,v} y_i^{u,v}\log s_i^{u,v}$ 来获得显著性预测结果 $s = f_\theta(x)$，其中 (u,v) 为像素坐标。针对基于硬标签的框架，通常设定 $y \in \{0,1\}$，其中，1 表示显著性前景，0 表示背景。

标签平滑设置。与上述硬标签设定不同，标签平滑正则化（LSR [293] 使用平滑后的标签 y' 而非 y，公式为：

$$y' = (1-\varepsilon)y + \varepsilon u(x). \tag{1}$$

这里，ε 是平滑参数，$u(x)$ 是通常被定义为均匀分布的固定分布。具有均匀分布的平滑标签 $u(x)$ 此时的定义为：

$$y' = (1-\varepsilon)y + \frac{\varepsilon}{K}, \tag{2}$$

其中 K 是类别的数量。

损失函数。给定平滑后的标签 y' 和硬标签 y，LSR 的损失函数定义为：

$$\mathcal{L}_{ls} = (1-\alpha)\mathcal{L}_{ce}(y,s) + \alpha\mathcal{L}_{ce}(y',s), \tag{3}$$

其中 α 用于平衡平滑标签和硬标签的贡献，平滑标签的相关损失定义为：$\mathcal{L}_{lsr} = \mathcal{L}_{ce}(y',s)$。注意，如果存在其它损失函数，则平滑标签只能用于交叉熵损失。

标签平滑到底起了什么作用？通常的交叉熵损失可以被重写为：

$$\mathcal{L}_{ce} = -\log s. \tag{4}$$

其中，s 是经过 sigmoid 激活后的模型预测值（针对二分类），其定义为：

$$s_j = e^{z_j} / \sum_{k=1}^{K} e^{z_k} = 1/(1 + \sum_{k \neq j} e^{z_k - z_j}). \tag{5}$$

将公式式 (4) 中的 s 代入后得到：

$$\mathcal{L}_{ce} = \log(1 + \sum_{k \neq j} e^{z_k - z_j}). \tag{6}$$

可将正确类别和其它类之间的差距定义为：$M = z_k - z_j$。我们就可以得出传统的交叉熵损失旨在最大化二者差距的结论。

针对标签平滑设置，如式 (2) 中所示，可重写平滑标签相关损失 $\mathcal{L}_{lsr}$ 为：

$$\begin{aligned}\mathcal{L}_{lsr} &= -((1-\varepsilon)y + \varepsilon/K)\log s - (1-(1-\varepsilon)y - \varepsilon/K)\log(1-s)\\ &= -(y\log s + (1-y)\log(1-s)) + (\varepsilon y - \frac{\varepsilon}{K})\log(\frac{s}{1-s}).\end{aligned} \tag{7}$$

在式 (5) 中使用 s 的定义可得：

$$\frac{s_j}{1-s_j} = \frac{1}{\sum_{k=1}^{K} e^{z_k - z_j} - 1}. \tag{8}$$

结合式 (8) 和式 (7) 可得：

$$\mathscr{L}_{lsr} = \mathscr{L}_{ce}(y,s) + (\varepsilon y - \frac{\varepsilon}{K}) * \frac{1}{\sum_{k=1}^{K} e^{z_k - z_j} - 1}. \tag{9}$$

式 (9) 中的第一个部分旨在最小化正确类别与其它类别之间的差距，这与式 (6) 中常规的二值交叉熵损失目标相同。第二部分朝相反（相对式 (6) 而言）方向缩小这种差距。通过这种方式，与标签平滑相关的损失可以平衡正确类与其它类之间的差距，这是防止模型过自信的正则化方法。

4.2 数据增强

作为一种有效的数据预处理技术，数据增强旨在从现有数据集中生成新样本，从而产生具有良好泛化能力的模型。给定训练数据集 $D = \{x_i, y_i\}_{i=1}^{N}$，数据增强产生一个新的数据集 $D' = \{x'_i, y'_i\}_{i=1}^{N'}$。如前文所述，两种主要的数据增强类型已受到特别关注。它们包括人工制定的策略和学习得到的策略 [301], [302]。对于学习的策略，本文观察到增强后的数据可能会根据上下文而发生巨大变化，这对于图像分类而言可能不是问题，但会改变图像的显著性属性。因此，我们仅关注手工制定的策略。

对于手工制定的数据增强策略，现有工作 [303], [304], [305], [306] 主要集中在三个方向：1）图像转换，例如，尺度或旋转变换；2）混合以生成新样本，这些样本是现有样本的近似；以及 3）在真值图上增加噪声。与学习的策略类似，混合策略会更改图像的上下文信息，这对于基于上下文的任务（例如显著性对象检测）有害。因此，本文将重点放在两种非常简单的数据增强技术上，即图像变换和向真值图中添加噪声。针对图像转换，本文随机放缩，旋转和裁剪部分图像（保留原始图片的 85% 的上下文信息）。针对增加噪声的解决方案，本文遵循 $\mathscr{N}(0.1, 0.3)$ 分布随机向显著性真值图中添加高斯噪声，从而得到一个有噪声的真值图。请注意，对于图像变换，本文同时变换图像和真值图，而在向真值图添加噪声时，本文仅处理显著性真值图。

4.3 自监督学习

自监督学习在不了解任务本身或真值图的情况下从图像中学习，这使其成为一种无监督的特征学习技术。按照惯例，对于有监督的学习环境，损失函数定义为 $\mathscr{L}_{ce}(y,s)$，其中 s 为模型预测，y 是真值图。针对自监督学习，最终的损失函数包含两个主要的部分：传统的交叉熵损失 $\mathscr{L}_{ce}(y,s)$ 和一个作为正则器的无监督损失，比如，$\mathscr{L}(g(x),s)$，其中 $g(x)$ 是原始输入 x 的变形。研究 [296], [308] 介绍了一种以旋转估计为辅助任务的自监督损失。

类似地，本文引入了缩放/旋转一致性损失函数来实现缩放/旋转不变性预测。具体而言，给定输入图像 x，本文将其预测定义为 s。随后，采用图像转换（缩放或旋转变换）可得到 x^t。然后，对预测 s 执行相同的变换，得到 s'。将 x^t 送入相同的显著性检测网络获得显著性预测结果并记作 s^t。我们假设 s' 和 s^t 应该相似。采用单尺度结构相似性指标（SSIM）[320], [321] 作为相似指标，则自监督损失可定义为：

$$\mathscr{L}_{ss} = 1 - SSIM(s', s^t). \tag{10}$$

4.4 利用了本文策略的损失函数

通过引入三种数据增强的策略，本文首先将随机数据增强方法用于训练图像集和训练真值图集，如节4.2中所示。之后可根据式 (1) 生成了平滑标签，在本文中，设置 $K = 2$ 表示显著的前景和背景区域。除了式 (3) 中的损失函数，本文还引入了自监督损失 $\mathscr{L}_{ss}$。最终本文的损失函数定义为：

$$\mathscr{L} = \mathscr{L}_{ls} + \gamma \mathscr{L}_{ss} \tag{11}$$

其中，引入的 γ 用于平衡自监督损失，根据经验设定 $\gamma = 0.3$。

5 SOC 评测

基于三个标准（比如，典型的框架，开源以及最先进的性能），本文从调研的 201 个方法中选择了 46 个传统 SOD 方法和 54 个深度学习模型（见章节2）来进行后续的基准评测。据我们所知，该评测是 RGB SOD 领域中最全面的研究。

5.1 实验设置

5.1.1 评估指标

请注意，本文的 SOC 数据集中非显著图像的真值图是全零矩阵，因此直接使用传统的 F 度量 [83] 将导致非常低且不准确的得分。因此，本文采用三个黄金指标（比如，MAE [322]，最大 E 度量 [5] 和 S 度量 [4]）来避免上述问题的出现，从而提供一个更可靠的评估。本文的 python 评估工具箱已开源。[5]

- MAE（M）是平均绝对误差度量，被广泛用于测量预测值和真值之间的像素级差异。
- **E 度量**（E_{ξ}^{max}）是一种新的感知指标，同时考虑了局部和全局相似性。
- **S 度量**（S_{α}）是一个在区域和对象级别量化结构相似性的标准度量。

Table 6
基准评测实验中使用的 SOC 数据集。

	SOC_train	SOC_val	SOC_test	合计
显著图 (Sal)	1,800	600	600	3,000
非显著图 (NonSal)	1,800	600	600	3,000
合计	3,600	1,200	1,200	6,000

5. https://github.com/mczhuge/SOCToolbox.

Table 7

在 SOC 测试集（1,200 张图像）上在以下方面比较传统 SOD 算法：$S_\alpha \uparrow$, $E_\xi^{max} \uparrow$, 和 $M \downarrow$。前三名的结果分别用红色，蓝色 和 绿色高亮表示。每个分数的上标是相应的排名。这些方法的细节在表 2中总结。总体排名指数表示三个指标的平均排名。这些结果可在Google Drive中获取。

	#	模型	代码	$S_\alpha \uparrow$	$E_\xi^{max} \uparrow$	$M \downarrow$	排名
2014-before	1	SUN [97]	Matlab	0.475^{46}	0.688^{44}	0.436^{46}	46
	2	LSSC [108]	Matlab + C	0.552^{45}	0.714^{43}	0.365^{45}	45
	3	BSF [113]	Matlab	0.554^{44}	0.728^{38}	0.353^{44}	44
	4	GR [110]	Matlab + C	0.588^{41}	0.715^{42}	0.332^{42}	43
	5	HS [38]	EXE	0.601^{40}	0.729^{37}	0.321^{41}	42
	6	Itti [46]	Matlab	0.587^{42}	0.736^{30}	0.311^{39}	41
	7	AIM [96]	Matlab	0.605^{39}	0.670^{45}	0.250^{24}	39
	8	GBVS [94]	Matlab	0.615^{36}	0.733^{35}	0.293^{37}	39
	9	LR [117]	Matlab	0.642^{31}	0.723^{40}	0.253^{27}	36
	10	CA [323]	Matlab + C	0.606^{38}	0.750^{22}	0.291^{36}	35
	11	MR [39]	Matlab + C	0.645^{29}	0.734^{33}	0.259^{31}	32
	12	SEG [100]	Matlab + C	0.576^{43}	0.765^{7}	0.352^{43}	32
	13	FT [83]	C	0.626^{34}	0.738^{29}	0.236^{20}	28
	14	MC [126]	Matlab + C	0.656^{23}	0.736^{30}	0.251^{25}	26
	15	CB [135]	Matlab + C	0.653^{25}	0.758^{13}	0.268^{33}	23
	16	SR [95]	Matlab/C++	0.658^{21}	0.661^{46}	0.156^{4}	23
	17	PCA [118]	Matlab + C	0.670^{18}	0.741^{28}	0.209^{13}	17
	18	MSS [111]	Matlab	0.682^{12}	0.776^{4}	0.231^{19}	10
	19	SF [114]	C	0.699^{6}	0.747^{26}	0.130^{1}	8
	20	DSR [125]	Matlab + C	0.702^{5}	0.751^{20}	0.184^{8}	8
	21	MSSS [101]	C	0.683^{11}	0.757^{14}	0.164^{5}	7
	22	HDCT [119]	Matlab	0.696^{7}	0.774^{5}	0.201^{12}	6
	23	DRFI [115]	C	0.709^{4}	0.791^{2}	0.197^{11}	4
	24	COV [109]	Matlab	0.711^{3}	0.761^{9}	0.146^{2}	2
	25	RBD [116]	Matlab	0.716^{2}	0.784^{3}	0.186^{9}	2
2021-2015	26	WMR [324]	Matlab + C	0.640^{32}	0.733^{35}	0.269^{34}	38
	27	MAPM [143]	Matlab + C	0.644^{30}	0.722^{41}	0.256^{29}	37
	28	BL [157]	Matlab + C	0.623^{35}	0.751^{20}	0.296^{38}	32
	29	RRWR [160]	Matlab	0.647^{27}	0.735^{32}	0.258^{30}	31
	30	WLRR [325]	Matlab + C	0.614^{37}	0.759^{11}	0.312^{40}	30
	31	RCRR [145]	Matlab	0.650^{26}	0.734^{33}	0.255^{28}	29
	32	GP [154]	Matlab + C	0.632^{33}	0.759^{11}	0.287^{35}	27
	33	TLLT [161]	Matlab	0.656^{23}	0.725^{39}	0.214^{15}	25
	34	BSCA [155]	Matlab + C	0.657^{22}	0.755^{16}	0.259^{31}	22
	35	SMD [137]	Matlab	0.662^{20}	0.748^{25}	0.246^{22}	21
	36	MDC [148]	C	0.675^{16}	0.744^{27}	0.219^{17}	20
	37	DSP [141]	Matlab + C	0.664^{19}	0.754^{17}	0.248^{23}	17
	38	MIL [144]	Matlab + C	0.671^{17}	0.750^{22}	0.236^{20}	17
	39	MST [159]	C	0.647^{27}	0.773^{6}	0.251^{25}	16
	40	GLC [140]	Matlab + C	0.676^{15}	0.756^{15}	0.223^{18}	15
	41	MBS [153]	Matlab	0.678^{14}	0.753^{18}	0.214^{15}	14
	42	LPS [142]	Matlab + C	0.694^{9}	0.749^{24}	0.183^{7}	13
	43	WFD [326]	C	0.680^{13}	0.760^{10}	0.213^{14}	12
	44	BFS [139]	Matlab + C	0.696^{7}	0.753^{18}	0.195^{10}	10
	45	WSC [162]	Matlab	0.693^{10}	0.765^{7}	0.179^{6}	5
	46	HCCH [149]	Matlab	0.736^{1}	0.794^{1}	0.149^{3}	1

5.1.2 训练与测试协议

评测中使用的 SOC 数据集的统计信息汇总在表 6中。对于传统算法，本文直接在 SOC 测试集（1,200 张图像）上测试其性能。对于深度学习模型，本文首先在其默认训练数据集下采用预训练的模型及其建议的训练参数设置 (见表 3 & 4) ，之后在 SOC 测试集上验证它们来粗略地得到前 100 的模型 (见表 7 & 8)。最后，本文提供了对 15 种 SOTA 方法的定量比较和详细分析，其中包括排名前 5 的传统方法和排名前 10 的深度学习模型。

5.2 定量比较

为了构建一个标准的排行榜 (比如，相同的分辨率，阈值步骤和评估工具)，本文采用了三个黄金指标，比如，S_α, E_ξ^{max}, M。

表 7中展示了 46 个 SOTA 传统 SOD 算法在 SOC 测试集上的性能。在 S **度量** (比如，S_α) 和最大 E **度量** (E_ξ^{max}) 上，HCCH 方法大大超过了所有竞争对手。RBD，COV 和 DRFI 在 S_α 得分方面获得可观的性能。同时，COV 在 S_α 度量项中排名第三，但是在 E_ξ^{max} 中排名第九。在评估项 MAE (比如，M) 中，表现前五的方法为：SF, COV, HCCH, SR 和 MSSS。值得一提的是，SF 减少了 M，胜过了所有最近的传统 SOD 方法。基于这些综合评分，排名前五的方法为 HCCH, RBD, COV, DRFI 和 WSC。

Table 8

在 SOC 测试集（1200 张图像）上评估 54 种基于深度学习的 SOD 模型。本文在表 3 和 表 4列出了采用的默认实现方法，使用默认实现来测试它们的泛化性能。结果可在Google Drive中找到。

	#	模型	代码	$S_\alpha \uparrow$	$E_\xi^{max} \uparrow$	$M \downarrow$	排名
2015	1	LEGS [176]	Caffe	0.679^{53}	0.765^{54}	0.228^{53}	54
	2	MDF [41]	Caffe	0.739^{49}	0.768^{53}	0.144^{43}	49
	3	MC [177]	Caffe	0.757^{47}	0.823^{43}	0.138^{35}	43
2016	4	DSL [179]	Caffe	0.724^{52}	0.810^{47}	0.194^{52}	51
	5	DISC [181]	Caffe	0.735^{51}	0.810^{47}	0.175^{50}	50
	6	DCL [195]	Caffe	0.771^{44}	0.836^{39}	0.157^{48}	45
	7	ELD [194]	Caffe	0.774^{42}	0.836^{39}	0.138^{35}	40
	8	DS [182]	Caffe	0.779^{40}	0.860^{24}	0.155^{46}	37
	9	DHS [169]	Pytorch	0.800^{32}	0.848^{33}	0.122^{30}	33
	10	RFCN [187]	Caffe	0.814^{23}	0.858^{27}	0.113^{23}	25
2017	11	UCF [204]	Caffe	0.654^{54}	0.805^{51}	0.285^{54}	53
	12	AMU [205]	Caffe	0.737^{50}	0.808^{50}	0.185^{51}	51
	13	SVF [203]	Caffe	0.761^{45}	0.816^{45}	0.156^{47}	47
	14	WSS [42]	Caffe	0.778^{41}	0.821^{44}	0.140^{39}	42
	15	DSS [201]	Caffe	0.807^{30}	0.858^{27}	0.111^{20}	27
	16	SRM [197]	Caffe	0.822^{16}	0.859^{26}	0.111^{20}	21
	17	MSRNet [62]	Caffe	0.816^{19}	0.871^{16}	0.117^{25}	20
	18	NLDF [199]	Tensorflow	0.816^{19}	0.860^{24}	0.104^{13}	16
2018	19	RAS [213]	Pytorch	0.759^{46}	0.813^{46}	0.151^{44}	46
	20	R3Net [210]	Pytorch	0.773^{43}	0.825^{42}	0.138^{35}	41
	21	LPSNet [214]	Pytorch	0.795^{35}	0.838^{38}	0.143^{42}	39
	22	DGRL-GLN [218]	Caffe	0.794^{36}	0.845^{36}	0.141^{40}	38
	23	C2SNet [212]	Caffe	0.791^{37}	0.845^{36}	0.138^{35}	36
	24	PiCA-Res [219]	Pytorch	0.810^{28}	0.858^{27}	0.128^{31}	31
	25	BMPM [217]	Tensorflow	0.810^{28}	0.853^{30}	0.119^{27}	29
	26	ASNet [216]	Keras	0.817^{18}	0.865^{20}	0.111^{20}	17
2019	27	MWS [237]	Pytorch	0.757^{47}	0.828^{41}	0.172^{49}	47
	28	AFNet [240]	Caffe	0.812^{24}	0.850^{32}	0.120^{29}	29
	29	SIBA [250]	Caffe	0.800^{32}	0.884^{10}	0.130^{33}	26
	30	Deepside [234]	Caffe	0.815^{21}	0.861^{23}	0.119^{27}	24
	31	PFANet [241]	Tensorflow	0.815^{21}	0.846^{35}	0.101^{8}	22
	32	PoolNet [244]	Pytorch	0.829^{13}	0.868^{18}	0.106^{16}	14
	33	SCRNet [45]	Pytorch	0.833^{11}	0.872^{15}	0.105^{14}	13
	34	CPDVgg [243]	Pytorch	0.856^{3}	0.889^{6}	0.079^{2}	2
	35	EGNet [251]	Pytorch	0.858^{1}	0.896^{2}	0.078^{1}	1
2020	36	ABPNet [283]	Pytorch	0.783^{38}	0.810^{47}	0.153^{45}	44
	37	U2Net [274]	Pytorch	0.780^{39}	0.795^{52}	0.105^{14}	35
	38	GCPANet [278]	Pytorch	0.807^{30}	0.848^{33}	0.133^{34}	34
	39	ITSD [281]	Pytorch	0.798^{34}	0.870^{17}	0.142^{41}	32
	40	MINet [282]	Pytorch	0.819^{17}	0.864^{22}	0.117^{25}	22
	41	SANet [57]	Pytorch	0.812^{24}	0.868^{18}	0.106^{16}	17
	42	GateNetVgg [285]	Pytorch	0.827^{15}	0.865^{20}	0.108^{18}	15
	43	F3Net [279]	Pytorch	0.828^{14}	0.891^{5}	0.109^{19}	12
	44	CSNet [284]	Pytorch	0.834^{10}	0.876^{14}	0.103^{10}	11
	45	LDF [280]	Pytorch	0.835^{9}	0.878^{12}	0.103^{10}	10
	46	RASNet [259]	Pytorch	0.832^{12}	0.887^{8}	0.103^{10}	9
	47	CAGVgg [272]	Keras	0.837^{8}	0.878^{12}	0.088^{4}	8
	48	DFI [255]	Pytorch	0.838^{7}	0.903^{1}	0.101^{8}	5
	49	R2Net [256]	Pytorch	0.857^{2}	0.885^{9}	0.084^{3}	4
2021	50	SCWS [289]	Pytorch	0.811^{26}	0.851^{31}	0.115^{24}	28
	51	ICON [53]	Pytorch	0.811^{26}	0.896^{2}	0.128^{31}	19
	52	BAS [32]	Pytorch	0.842^{5}	0.882^{11}	0.092^{7}	7
	53	ABP [327]	Pytorch	0.842^{5}	0.889^{6}	0.091^{6}	6
	54	CVAE [327]	Pytorch	0.849^{4}	0.892^{4}	0.089^{5}	3

SOC 测试集上的 54 种深度学习 SOD 模型的定量结果在表 8中展示。在指标项 S_α 上，EGNet, R2Net 和 CPDVgg 是排名前三的模型，其得分均超过了 0.85。大约 46%（比如，21/45）的模型得分在 0.650 到 0.800 之间。与传统方法得到 S_α 评分为 0.736 对比，除了四个早期模型 (比如，DISC, DSL, LEGS 和 UCF)，可以看到在过去几年中模型性能的持续提

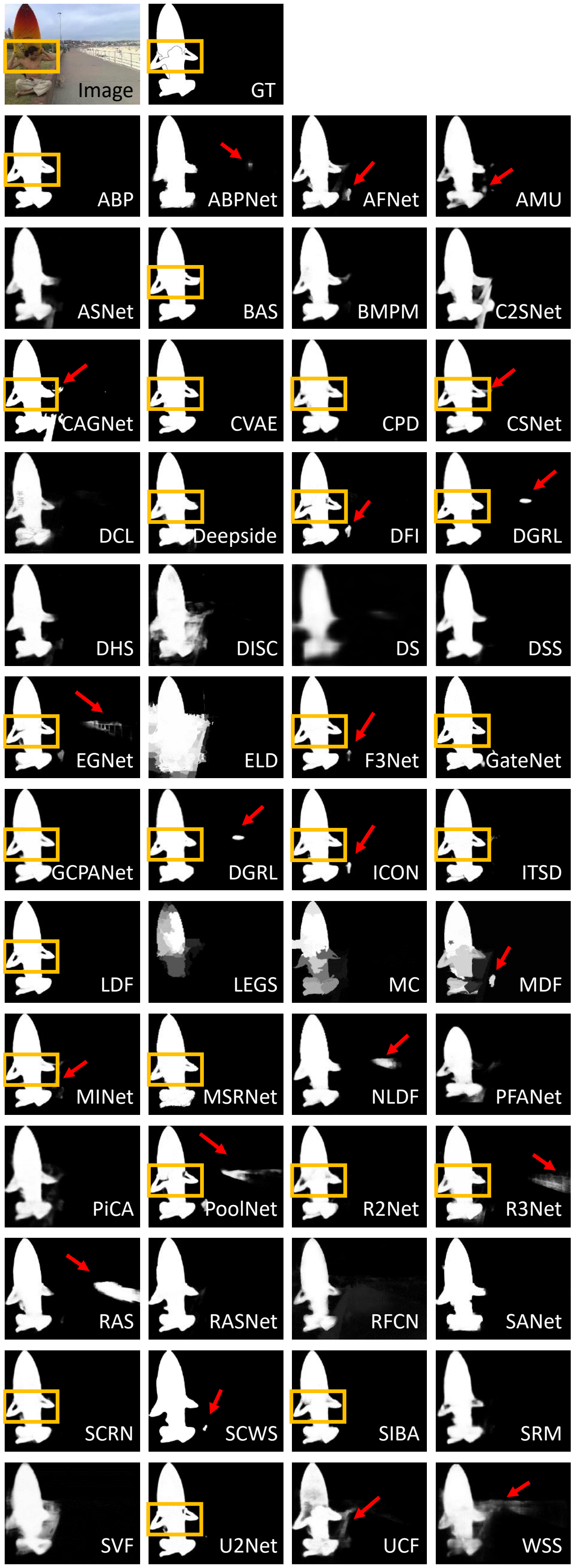


Figure 8. 深度学习模型的可视化结果。

升。与此同时，45 个模型中的 30 个获得了高性能表现 (例如, 0.800≤ S_α ≤0.850)，平均性能接近 0.820。有趣的是，就

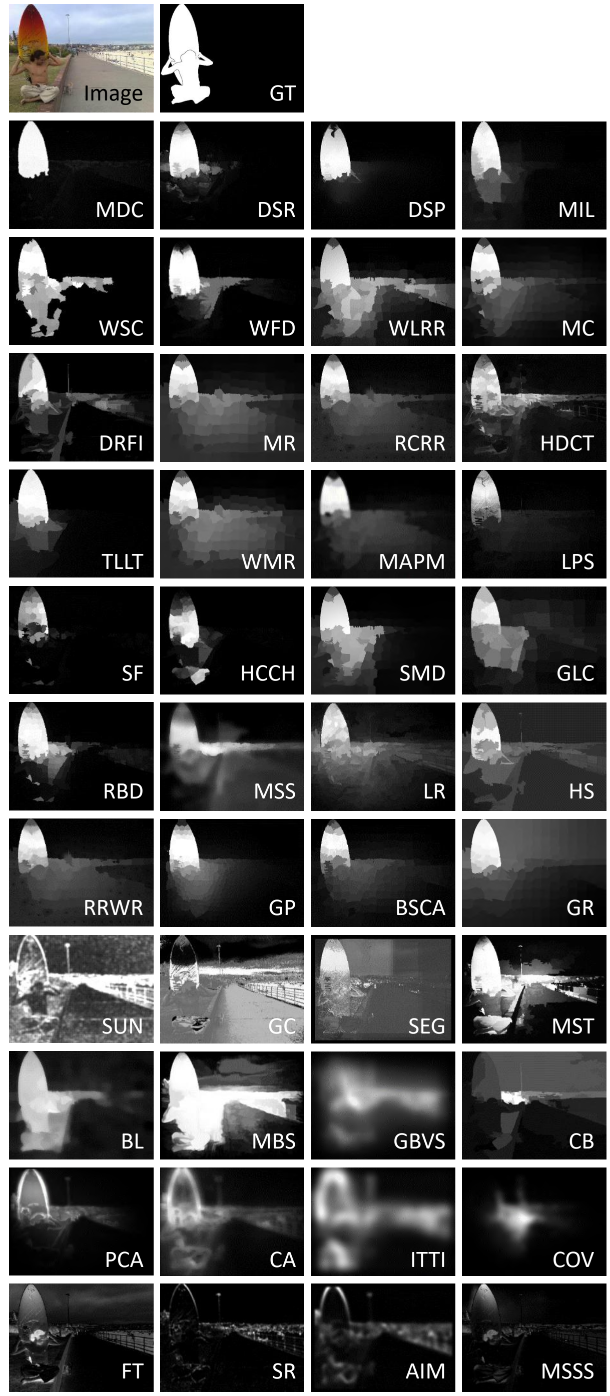


Figure 9. 最先进的传统方法的定性结果。

E_ξ^{max} 而言，多任务学习框架 DFI 和完整性学习模型的最佳和次佳得分分别为 0.903 和 0.896。就 MAE 而言，前三名的模型为 EGNet，CPDVgg 和 R2Net，这与 S 度量的结果一致。从评测的 54 个模型，我们发现在 S 度量方面表现良好的模型在 MAE 中也表现良好。总体而言，排名前 10 位的方法是 EGNet，CPDVgg，CVAE，R2Net，DFI，ABP，BAS，CAGVgg，RASNet 和 LDF。在接下来的小节 (章节6) 中，本文将对这些模型进行更详细的分析。

Table 9

在属性级别性能方面对 14 种最先进方法的比较。对于深度学习模型，本文在 SOC-Sal_train 数据集（比如，1800 张图像）上重新进行了训练。请到2, 3, & 4中寻找更多的细节。这些结果可以在Google Drive中找到。

	属性 / 模型	AC		BO		CL		HO		MB		OC		OV		SC		SO		平均分.	
		$S_\alpha \uparrow$	$M \downarrow$	$S_\alpha \uparrow$	$M \downarrow$	$S_\alpha \uparrow$	$M \downarrow$	$S_\alpha \uparrow$	$M \downarrow$	$S_\alpha \uparrow$	$M \downarrow$	$S_\alpha \uparrow$	$M \downarrow$	$S_\alpha \uparrow$	$M \downarrow$	$S_\alpha \uparrow$	$M \downarrow$	$S_\alpha \uparrow$	$M \downarrow$	$S_\alpha \uparrow$	$M \downarrow$
传统方法	COV [109]	0.505	0.216	0.277	0.577	0.453	0.280	0.508	0.229	0.494	0.219	0.484	0.246	0.423	0.314	0.535	0.174	0.525	0.172	0.467	0.270
	WSC [162]	0.541	0.205	0.356	0.517	0.517	0.252	0.556	0.211	0.536	0.210	0.529	0.227	0.475	0.292	0.567	0.170	0.535	0.181	0.512	0.252
	HCCH [149]	0.585	0.199	0.354	.0.525	0.537	0.254	0.615	0.197	0.547	0.202	0.552	0.225	0.468	0.298	0.595	0.165	0.588	0.162	0.538	0.247
	DRFI [115]	0.598	0.229	0.391	0.513	0.570	0.274	0.618	0.230	0.556	0.230	0.577	0.248	0.527	0.304	0.614	0.188	0.585	0.197	0.560	0.268
	RBD [116]	0.589	0.225	0.429	0.481	0.575	0.260	0.625	0.216	0.557	0.213	0.583	0.235	0.521	0.295	0.602	0.191	0.579	0.192	0.562	0.256
深度学习	ABP [327]	0.767	0.092	0.592	0.315	0.742	0.125	0.787	0.101	0.742	0.095	0.740	0.112	0.746	0.132	0.759	0.083	0.741	0.080	0.735	0.126
	EGNet [251]	0.791	0.088	0.593	0.307	0.739	0.137	0.788	0.110	0.763	0.115	0.743	0.120	0.750	0.138	0.800	0.076	0.753	0.088	0.747	0.131
	CPDVgg [243]	0.806	0.076	0.626	0.278	0.765	0.118	0.808	0.096	0.786	0.097	0.765	0.103	0.760	0.127	0.801	0.070	0.765	0.076	0.765	0.116
	CAGVgg [272]	0.795	0.080	0.700	0.208	0.782	0.115	0.808	0.098	0.764	0.102	0.751	0.120	0.763	0.127	0.795	0.081	0.744	0.093	0.767	0.114
	RASNet [259]	0.821	0.066	0.626	0.276	0.785	0.106	0.816	0.087	0.788	0.086	0.776	0.096	0.779	0.113	0.810	0.066	0.774	0.070	0.772	0.107
	CVAE [327]	0.813	0.075	0.688	0.217	0.790	0.107	0.816	0.092	0.784	0.091	0.771	0.104	0.776	0.115	0.820	0.069	0.767	0.080	0.781	0.106
	LDF [280]	0.819	0.071	0.697	0.212	0.796	0.105	0.824	0.088	0.792	0.085	0.781	0.098	0.790	0.107	0.780	0.073	0.801	0.072	0.787	0.101
	R2Net [256]	0.827	0.071	0.656	0.257	0.802	0.107	0.826	0.092	0.794	0.097	0.789	0.099	0.791	0.112	0.807	0.072	0.788	0.073	0.787	0.109
	BAS [32]	0.831	0.060	0.723	0.166	0.785	0.110	0.814	0.093	0.797	0.072	0.780	0.101	0.781	0.114	0.820	0.072	0.787	0.075	0.791	0.096
	平均分.	0.721	0.125	0.551	0.346	0.688	0.168	0.729	0.139	0.693	0.137	0.687	0.152	0.668	0.185	0.722	0.111	0.693	0.115	-	-

5.3 定性比较

本文在图 8 和 图 9中展示了两组定性比较。正如图 8中所示，深度学习模型得到的显著图在不同程度上与真值图相似。具体来说，对于 ASNet，C2SNet，BMPM，DCL，DHS，DSS，DS，DISC，SVF，RFCN 和 PFANet，它们可以很好地识别对象的位置。但是，所有这些方法都会在对象边界上产生模糊的响应。PFANet，MDF，MC 和 LEGS 甚至几乎破坏了对象的完整性。为了更好地分析这些结果，我们引入了黄色矩形来标记高质量的分割区域，并使用红色箭头指出错误。我们观察到八个模型（ABPNet，AFNet，AMU，NLDF，RAS，SCWS，UCF 和 WSS）可以定位人物对象，但会引入其它噪声。同时还注意到，CAGNet，CSNet，MINet，DGRL，EGNet，F3Net，ICON，PoolNet 和 R3Net 甚至可以捕获人肘部的小结构。此外，与上述方法相比，来自 R2Net，Deepside，SIBA 和 MSRNet 的显著图呈现了更好的结果。令人惊讶的是，BAS，U2Net，ABP，CPD，GateNet，GCPANet，ITSD，LDF，SCRN 和 CAVE 的表现非常接近真值，并且在黄色矩形区域中形成了刀锋状的边界，而没有任何额外的噪声。

与深度学习模型形成鲜明对比的是，传统模型（图 9）都无一例外地失败了。WSC，HCCH 和 RBD 是三种最有希望的方法。但是，它们的结果仍然与真值图相差甚远，因为它们主要基于从颜色，方向，对比度中提取的各种先验特征。此外，由于人位于图像边界附近，因此先验中心偏置在这种情况下不适合，因此对于这些方法，该示例更具挑战性。

6 进一步基准评测

6.1 基于属性的评估

基于7 & 8中展示的排名靠前的模型，本文在 SOC-Sal_train 数据集（1800 张图像）上进一步重新训练了排名前十[6]的深度学习模型（使用它们的默认设置），并在 SOC-Sal 测试集上针对基于属性的评估进行了测试。表 9展示了各种 SOD 模型在特定属性的子数据集上的性能。由于篇幅限制，在接下来的部分，仅选择一些代表性属性进行进一步分析。

大物体（BO）当物体与相机距离很近时，经常会出现大物体（BO）场景，因此在图片中可以清楚地看到微小的文字或图案。在这种情况下，倾向于关注局部信息的模型将被严重误导，与其平均性能对比会出现严重的性能损失（例如，BAS 上 8.6% 的 S_α 损失，CAGVgg 上 8.7% 的损失，LDF 上 11.4% 的损失，以及 COV 上 40.7% 的损失）。在所有属性中，对于传统模型和深度学习模型而言，BO 都是最困难的。

小物体（SOs）对于某些 SOD 模型是棘手的。四个模型（比如, BAS, CVAE, CAGVgg, 和 RASNet）在这种场景下出现性能下降（例如, 从 BAS-0.5% 到 RASNet-3.6%），这是因为 SOs 在 CNN 降采样的过程中容易被忽略。相反，其它模型在 SOs 上具有增强的性能，在 BOs 上却显著降低了性能。

异构物体（HOs）在自然场景中很常见。所有模型在 HOs 上的性能都有一定程度的提高，从 2.9% 波动到 14.3%。本文猜测这是因为 HO 图像在所有数据集中所占的比例很高，如图 6所示，因此模型对这一属性更加熟悉。

遮挡（OC）场景当物体被部分遮挡时发生。因此，SOD 模型必须捕获全局语义以弥补对象的不完整信息。正如所观察到的，传统模型比起其平均情况获得了更高的性能。然而对于深度学习模型，这种情况是相反的。

正如表 9 中最后一行（每个属性的平均性能）所示，MB 和 SO 有相同的 S_α 得分。此外，AC 和 SC 的平均得分也非常相近。似乎现有的基于深度学习的 SOD 模型可以有效地解决外观变化和形状复杂问题。与 OV 和 OC 属性相似，CL 和 MB 对现有方法仍然充满挑战，只能得到中等的（比如, 0.65< S_α <0.70）S 度量得分。

6.2 与基线对比

本文引入了三种数据集增强策略，以防止由于数据集偏差而导致网络高估。这些措施包括标签平滑、随机数据扩充和自监督学习。本文的策略可以很容易地在现有的显著性对象检测框架中作为通用的数据处理技术。因此，我们将本文的策略引

6. DFI 的作者仅放出了测试代码，因此本文仅在 9 个模型上进行验证。

Table 10
本文的数据集增强策略的贡献。

方法 \ 指标	$S_\alpha \uparrow$	$E_\xi^{max} \uparrow$	$M \downarrow$
RASNet [259]	0.832	0.887	0.103
Our-RASNet	0.841	0.897	0.096
LDF [280]	0.835	0.878	0.103
Our-LDF	0.845	0.891	0.097
BAS [32]	0.842	0.882	0.092
Our-BAS	0.856	0.895	0.086
R2Net [256]	0.857	0.885	0.084
Our-R2Net	0.868	0.899	0.080
CVAE [327]	0.849	0.892	0.089
Our-CVAE	0.863	0.902	0.086

Table 11
每个数据集增强策略的贡献。

方法 \ 指标	$S_\alpha \uparrow$	$E_\xi^{max} \uparrow$	$M \downarrow$
CVAE [327]	0.849	0.892	0.089
LS	0.851	0.895	0.088
SS	0.852	0.894	0.088
RDA	0.855	0.896	0.086
Our-CVAE	0.863	0.902	0.086

Table 12
章节6中跨数据集泛化结果。在一个数据集上训练 UC-Net (CVPR'20) [328] 并在其它所有数据集上测试。"Sel."：对角线分数（在同一数据集上进行训练和测试）。"Oth."：除自身外，所有方面平均得分。

指标	$S_\alpha \uparrow$ [4]									下降 $\downarrow$
Train \ Test	SOC	M10K	DU-O	DUTS	ECC	HKU	ILSO	Sel.	Oth.	
SOC [1]	.884	.768	.686	.834	.749	.774	.841	.884	.775	12%
M10K [36]	.800	.921	.784	.894	.881	.882	.884	.921	.854	7%
DU-O [39]	.833	.898	.854	.877	.862	.867	.886	.854	.871	-2%
DUTS [42]	.795	.882	.793	.910	.890	.903	.900	.910	.861	5%
ECC [38]	.791	.886	.800	.901	.901	.898	.903	.901	.863	4%
HKU [41]	.818	.892	.787	.904	.883	.910	.905	.910	.865	5%
ILSO [62]	.841	.888	.790	.898	.882	.896	.920	.920	.866	6%
Oth.	.813	.869	.773	.885	.858	.870	.887			

入到五个基准显著性物体检测模型，并在表 10中展示了其性能，其中"Our-"代表使用了本文数据集增强策略的基准模型。我们进一步调研了每种数据增强策略的贡献，并在表 11中展示了其性能，这里我们选用 CVAE [327] 作为基础模型。

训练 & 测试原则。本文使用对应训练数据集重新训练了表 10中的五个模型，例如, 使用 MB [33] 训练 RASNet [259]，使用 DUTS [42] 训练了其它四个模型。本文遵循它们原来的训练和测试设置，例如，相同的最大轮数，学习率，训练和测试图像大小。

讨论。表 10显示，使用本文提出的策略使模型获得了更好的性能，这说明了本文解决方案的有效性。进一步而言，表 11中，"LS", "RDA", "SS"代表在基础模型上分别加入标签平滑策略，随机数据增强以及自监督学习。它表明，随机数据增强可实现最大的性能提升，而标签平滑和自监督学习则可实现相当的性能提升。主要原因是数据扩充将各种样本引入到初始训练数据集中，这有效地提高了模型的泛化能力。针对自监督学习策略，由于 CVAE 模型 [327] 已经采用多尺度图像作为输入策略，因此本文观察到性能略有提高。但是，总体上更好的性能仍然可以验证所提出策略的有效性。标签平滑 [293] 是为了防止模型过自信而引入的，使得模型能够获得更好的校准。但是，在显著性检测领域还没有评价指标可以刻画这样的校准误差。本文将在预期的校准误差 [329] 中进行调研，并将其扩展到将来的显著性检测任务中，以更好地解释显著性模型的校准误差。

6.3 跨数据集泛化

为了研究现有 SOD 数据集的难度，本文采用 CDA（跨数据分析） [44] 方法。给定 N 个候选数据集 $\{D^n\}_{n=1}^N$，本文首先在 D_i 上训练模型，之后在其它数据集（比如, $\{D^n\}_{n=1,n\neq i}^N$）上进行测试。根据 [77], [330]，本文分别从每个数据集中随机选择 800 张图像和 200 张图像作为训练集和测试集。

本文在现有流行的、图片数量超过 1,000 张的数据集上训练 UC-Net [328]。表 12展示了每个数据集上 S_α 指标的性能，表中每列展示了在特定数据集上测试并在所有其它数据集上进行训练的 UC-Net 得分；每行表示在一个数据集上训练并在所有其它数据集上进行测试的 UC-Net 的性能，从而可以说明用于训练的数据集的泛化能力。观察发现，相比其它数据集，本文的 SOC（例如, Oth. = 0.813）和 DU-O（Oth. = 0.773）数据集更加困难。此外，我们注意到在本文的 SOC 数据集上训练的模型在其它数据集上的效果不佳（例如, Oth. 列: 0.775）。这是因为本文的 SOC 数据集在开放环境中包含许多现实世界中的显著对象，因此该模型无法充分拟合现有数据集中的简单或干净的场景。我们的核心观察结果也印证了这一观点，即数据选择偏差在现有数据集中普遍存在。

7 未来方向

人的注意力会受到四个关键因素的影响：

- **视觉特性**。人们的注意力会被具有独特的视觉特性的物体所吸引 [331]。
- **记忆**。如果一个人对某个物体很熟悉，那么该物体就更容易引起他的注意。
- **目的**。也就是说，带有特定目的观察者和无目的观察者的眼动数据注意力图差异很大。
- **情感**。除上述因素外，人对同一场景的注意力还可能会受到人的情感（例如，幸福，悲伤，愤怒）的影响。

正如 Cave [331] 中所阐明的，注意力控制是由这些因素共同决定的。不幸的是，现有 SOD 数据集的标注时没有明确地描述它们解决的因素。相比之下，本文 SOC 数据集的真值图是基于 salicon（自由视图任务）数据集得到的[7]也称为意义图，意义图在最近的研究中 [331], [332], [333] 经常被使用。正如 Kalash 等人 [67] 所得出的结论，迄今为止，当前的 SOD 工作大部分解决的是一个病态的问题。因此，我们从 6 个研究维度推荐了一些未来的研究方向以便重新思考 SOD 任务：

（1）**数据层面：**最近，使用 2D（RGB SOD）和 3D(比如, RGB-D, RGB-T）输入数据的视觉显著性检测任务引起了人们的极大兴趣。然而，在光场 SOD（4D），LIDAR SOD 以及 360°SOD 任务中，仍然没有很好的研究。为这些类型的数据建立新的数据集将极大地促进该领域的发展。研究显著性检

7. http://salicon.net/

测的另一个有趣途径是研究细粒度的任务，例如显著性实例检测 [62], [63], [334], [335] 和部分-对象视觉显著性检测 [336]。

(2)**任务层面：**多任务学习在最近的工作中表现出色 [337]。现有的方案主要集中在视觉任务上，例如联合显著性物体检测和伪装物体检测 [338]、同时检测突出的物体，边缘和骨架 [255] 以及同时检测，排序和细分多个显著性对象 [65]。随着 Transformer 技术在自然语言处理（NLP）中的成功应用，将多模态学习引入显著性检测领域可能是进一步整合其它类型信息的可行方法，例如 CV+NLP (与 [339] 类似), CV+ 语音 [340], and CV+ 其它模态。

（3）**模型层面：** 目前已经设计了大量算法以提高检测精度。但是，还有一些有希望的方向可以进一步研究，例如数据增强技术 [341]，高效的 SOD 模型（例如, 轻量级模型 [284], [342]），新的损失函数 [287], [343]，基于排名的模型 [65], [138]，以及基于 transformer 的模型 [344], [345]。

（4）**监督层面：** 除了当前 SOD 模型中最常见的全监督学习之外，其它监督策略，例如，弱监督（比如, 涂鸦 [57], 类别 [346], 和多边形等），半监督 [54]，自监督 [61], [347] 和无监督 [59] 策略也值得研究。

（5）**评估层面：** 评估指标对于模型训练，测试和基准评测都很重要。但是，SOD 社区仍然使用经典指标，例如 IoU，F 度量和 MAE。这些度量标准旨在用于通用评估，而不是专门用于评估 SOD 任务。使得它们在某些特定的应用场合（例如那些具有高质量要求的应用程序）中不能很好地工作。本文设想引入一种针对 SOD 任务的新指标 (例如，[348] 中使用的基于梯度或连通性的误差)，例如加权 F 度量 [3] 和 S 度量 [4]，在将来会是另一个重要的研究方向。

（6）**应用层面：** SOD 任务属于一个更通用的任务，称为类无关对象检测（CAOD）。对于简单的场景（例如，仅包含一到两个清晰对象），SOD 与 CAOD 相同。从这个角度来看，尽管目前 SOD 模型的代表性案例数量有限（例如，阿里巴巴的时尚搜索系统 [339]），但它们在现实世界中仍具有许多潜在应用 [31], [32], [201]。

8 结论

在这篇综述中，本文确定并解决了在 SOD 任务中研究人员长期忽略的*数据选择偏差*问题。与以前的研究不同，本文旨在探索开放场景中的 SOD 任务。为了实现这一目标，本文收集了一个新的具有挑战性且密集标注的 SOC 数据集；分析了大量 (~200) 代表性模型；进行了最完整的（前 100 名）基准评测；设计了一系列简单的学习策略，以有效地利用负样本和训练数据；并且指出了当前面临的一些挑战和未来的研究方向。我们希望这些贡献将为 SOD 社区提供一个在开放环境中探索新技术的机会。然而，在这个广阔的领域中彻底研究所有模型是不切实际的，因此本文试图涵盖最重要的工作，我们会继续在本文的网站上持续更新最新技术。

References

[1] 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 Eur. Conf. Comput. Vis., 2018, pp. 186–202.

[2] D.-P. Fan, J. Zhang, G. Xu, M.-M. Cheng, and L. Shao, "Salient objects in clutter," 2021.

[3] R. Margolin, L. Zelnik-Manor, and A. Tal, "How to evaluate foreground maps?" in IEEE Conf. Comput. Vis. Pattern Recog., 2014, pp. 248–255.

[4] D.-P. Fan, M.-M. Cheng, Y. Liu, T. Li, and A. Borji, "Structure-measure: A New Way to Evaluate Foreground Maps," in Int. Conf. Comput. Vis., 2017, pp. 4548–4557.

[5] D.-P. Fan, C. Gong, Y. Cao, B. Ren, M.-M. Cheng, and A. Borji, "Enhanced-alignment Measure for Binary Foreground Map Evaluation," in Int. Joint Conf. Artif. Intell., 2018, pp. 698–704.

[6] P. Zhang, T. Zhuo, W. Huang, K. Chen, and M. Kankanhalli, "Online object tracking based on cnn with spatial-temporal saliency guided sampling," Neurocomputing, vol. 257, pp. 115–127, 2017.

[7] A. Borji, S. Frintrop, D. N. Sihite, and L. Itti, "Adaptive object tracking by learning background context," in IEEE Conf. Comput. Vis. Pattern Recog. Worksh., 2012, pp. 23–30.

[8] V. Mahadevan and N. Vasconcelos, "On the connections between saliency and tracking," in Adv. Neural Inform. Process. Syst., 2012, pp. 1664–1672.

[9] A. Abdulmunem, Y.-K. Lai, and X. Sun, "Saliency guided local and global descriptors for effective action recognition," Comput. Vis. Media, vol. 2, no. 1, pp. 97–106, 2016.

[10] J. He, J. Feng, X. Liu, T. Cheng, T.-H. Lin, H. Chung, and S.-F. Chang, "Mobile product search with bag of hash bits and boundary reranking," in IEEE Conf. Comput. Vis. Pattern Recog., 2012, pp. 3005–3012.

[11] G. Liu and D. Fan, "A model of visual attention for natural image retrieval," in IEEE Int. Conf. Inf. Sci. Cloud Comput. Companion, 2013, pp. 728–733.

[12] J.-Y. Zhu, J. Wu, Y. Xu, E. Chang, and Z. Tu, "Unsupervised object class discovery via saliency-guided multiple class learning," IEEE Trans. Pattern Anal. Mach. Intell., vol. 37, no. 4, pp. 862–875, 2015.

[13] H. Liu, L. Zhang, and H. Huang, "Web-image driven best views of 3d shapes," The Vis. Comput., vol. 28, no. 3, pp. 279–287, 2012.

[14] K. Gu, G. Zhai, X. Yang, W. Zhang, and C. W. Chen, "Automatic contrast enhancement technology with saliency preservation," IEEE Trans. Circ. Syst. Video Technol., vol. 25, no. 9, pp. 1480–1494, 2015.

[15] R. Zhao, W. Ouyang, and X. Wang, "Unsupervised salience learning for person re-identification," in IEEE Conf. Comput. Vis. Pattern Recog., 2013, pp. 3586–3593.

[16] M. Donoser, M. Urschler, M. Hirzer, and H. Bischof, "Saliency driven total variation segmentation," in Int. Conf. Comput. Vis., 2009, pp. 817–824.

[17] S. J. Oh, R. Benenson, A. Khoreva, Z. Akata, M. Fritz, and B. Schiele, "Exploiting saliency for object segmentation from image level labels," in IEEE Conf. Comput. Vis. Pattern Recog., 2017, pp. 4410–4419.

[18] D.-P. Fan, W. Wang, M.-M. Cheng, and J. Shen, "Shifting more attention to video salient object detection," in IEEE Conf. Comput. Vis. Pattern Recog., 2019, pp. 8554–8564.

[19] T. Chen, M.-M. Cheng, P. Tan, A. Shamir, and S.-M. Hu, "Sketch2photo: Internet image montage," ACM Trans. Graph., vol. 28, no. 5, p. 124, 2009.

[20] M.-M. Cheng, F.-L. Zhang, N. J. Mitra, X. Huang, and S.-M. Hu, "Repfinder: finding approximately repeated scene elements for image editing," ACM Trans. Graph., vol. 29, no. 4, p. 83, 2010.

[21] V. Ramanishka, A. Das, J. Zhang, and K. Saenko, "Top-down visual saliency guided by captions," in IEEE Conf. Comput. Vis. Pattern Recog., vol. 1, no. 2, 2017, p. 7.

[22] C. Guo and L. Zhang, "A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression," IEEE Trans. Image Process., vol. 19, no. 1, pp. 185–198, 2010.

[23] H. Hadizadeh and I. V. Bajic, "Saliency-aware video compression," IEEE Trans. Image Process., vol. 23, no. 1, pp. 19–33, 2014.

[24] Y. Liu, Z. Xu, W. Ye, Z. Zhang, S. Weng, C.-C. Chang, and H. Tang, "Image neural style transfer with preserving the salient regions," IEEE Access, vol. 7, pp. 40 027–40 037, 2019.

[25] M.-M. Cheng, X.-C. Liu, J. Wang, S.-P. Lu, Y.-K. Lai, and P. L. Rosin, "Structure-preserving neural style transfer," IEEE Trans. Image Process., vol. 29, pp. 909–920, 2020.

[26] A. Toshev, J. Shi, and K. Daniilidis, "Image matching via saliency region correspondences," in IEEE Conf. Comput. Vis. Pattern Recog., 2007, pp. 1–8.

[27] M. J. Islam, R. Wang, K. de Langis, and J. Sattar, "Svam: Saliency-guided visual attention modeling by autonomous underwater robots," arXiv preprint arXiv:2011.06252, 2020.

[28] D.-P. Fan, G.-P. Ji, G. Sun, M.-M. Cheng, J. Shen, and L. Shao, "Camouflaged object detection," in IEEE Conf. Comput. Vis. Pattern Recog., 2020, pp. 2777–2787.

[29] Y. Tu, L. Niu, W. Zhao, D. Cheng, and L. Zhang, "Image cropping with composition and saliency aware aesthetic score map," in AAAI Conf. Art. Intell., vol. 34, no. 07, 2020, pp. 12 104–12 111.

[30] J. Kim, T. Misu, Y.-T. Chen, A. Tawari, and J. Canny, "Grounding human-to-vehicle advice for self-driving vehicles," in IEEE Conf. Comput. Vis. Pattern Recog., 2019, pp. 10 591–10 599.

[31] N. Kumar, P. N. Belhumeur, A. Biswas, D. W. Jacobs, W. J. Kress, I. C. Lopez, and J. V. Soares, "Leafsnap: A computer vision system for automatic plant species identification," in Eur. Conf. Comput. Vis., 2012, pp. 502–516.

[32] X. Qin, D.-P. Fan, C. Huang, C. Diagne, Z. Zhang, A. C. Sant'Anna, A. Suàrez, M. Jagersand, and L. Shao, "Boundary-aware segmentation network for mobile and web applications," arXiv preprint arXiv:2101.04704, 2021.

[33] T. Liu, J. Sun, N. Zheng, X. Tang, and H.-Y. Shum, "Learning to detect a salient object," in IEEE Conf. Comput. Vis. Pattern Recog., 2007, pp. 1–8.

[34] S. Alpert, M. Galun, R. Basri, and A. Brandt, "Image segmentation by probabilistic bottom-up aggregation and cue integration," in IEEE Conf. Comput. Vis. Pattern Recog., 2007, pp. 1–8.

[35] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in Int. Conf. Comput. Vis., vol. 2, 2001, pp. 416–423.

[36] M.-M. Cheng, N. J. Mitra, X. Huang, P. H. S. Torr, and S.-M. Hu, "Global contrast based salient region detection," IEEE Trans. Pattern Anal. Mach. Intell., vol. 37, no. 3, pp. 569–582, 2015.

[37] A. Borji, D. N. Sihite, and L. Itti, "Salient object detection: a benchmark," in Eur. Conf. Comput. Vis., 2012, pp. 414–429.

[38] Q. Yan, L. Xu, J. Shi, and J. Jia, "Hierarchical saliency detection," in IEEE Conf. Comput. Vis. Pattern Recog., 2013, pp. 1155–1162.

[39] C. Yang, L. Zhang, H. Lu, X. Ruan, and M.-H. Yang, "Saliency detection via graph-based manifold ranking," in IEEE Conf. Comput. Vis. Pattern Recog., 2013, pp. 3166–3173.

[40] Y. Li, X. Hou, C. Koch, J. M. Rehg, and A. L. Yuille, "The secrets of salient object segmentation," in IEEE Conf. Comput. Vis. Pattern Recog., 2014, pp. 280–287.

[41] G. Li and Y. Yu, "Visual saliency based on multiscale deep features," in IEEE Conf. Comput. Vis. Pattern Recog., 2015, pp. 5455–5463.

[42] 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 IEEE Conf. Comput. Vis. Pattern Recog., 2017, pp. 136–145.

[43] Z. Yang, S. Soltanian-Zadeh, and S. Farsiu, "Biconnet: An edge-preserved connectivity-based approach for salient object detection," arXiv preprint arXiv:2103.00334, 2021.

[44] A. Torralba and A. A. Efros, "Unbiased look at dataset bias," in IEEE Conf. Comput. Vis. Pattern Recog., 2011, pp. 1521–1528.

[45] Z. Wu, L. Su, and Q. Huang, "Stacked cross refinement network for edge-aware salient object detection," in IEEE Conf. Comput. Vis. Pattern Recog., 2019, pp. 7264–7273.

[46] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," IEEE Trans. Pattern Anal. Mach. Intell., vol. 20, no. 11, pp. 1254–1259, 1998.

[47] A. Borji, "Saliency prediction in the deep learning era: Successes and limitations," IEEE Trans. Pattern Anal. Mach. Intell., vol. 43, no. 2, pp. 679–700, 2021.

[48] T. Liu, J. Sun, N. Zheng, X. Tang, and H. Shum, "Learning to detect a salient object," in IEEE Conf. Comput. Vis. Pattern Recog., 2007, pp. 1–8.

[49] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, and H.-Y. Shum, "Learning to detect a salient object," IEEE Trans. Pattern Anal. Mach. Intell., vol. 33, no. 2, pp. 353–367, 2010.

[50] P. Zhang, W. Liu, Y. Zeng, Y. Lei, and H. Lu, "Looking for the detail and context devils: High-resolution salient object detection," IEEE Trans. Image Process., 2021.

[51] Y. Zeng, P. Zhang, J. Zhang, Z. Lin, and H. Lu, "Towards high-resolution salient object detection," in Int. Conf. Comput. Vis., 2019, pp. 7234–7243.

[52] Q. Zhang, R. Cong, C. Li, M.-M. Cheng, Y. Fang, X. Cao, Y. Zhao, and S. Kwong, "Dense attention fluid network for salient object detection in optical remote sensing images," IEEE Trans. Image Process., 2020.

[53] M. Zhuge, D.-P. Fan, N. Liu, D. Zhang, D. Xu, and L. Shao, "Salient object detection via integrity learning," arXiv preprint arXiv:2101.07663, 2021.

[54] 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 Trans. Cybern., vol. 49, no. 4, pp. 1173–1185, 2019.

[55] G. Li, Y. Xie, and L. Lin, "Weakly supervised salient object detection using image labels," in AAAI Conf. Art. Intell., vol. 32, no. 1, 2018.

[56] L. Zhang, J. Zhang, Z. Lin, H. Lu, and Y. He, "Capsal: Leveraging captioning to boost semantics for salient object detection," in IEEE Conf. Comput. Vis. Pattern Recog., 2019, pp. 6024–6033.

[57] J. Zhang, X. Yu, A. Li, P. Song, B. Liu, and Y. Dai, "Weakly-supervised salient object detection via scribble annotations," in IEEE Conf. Comput. Vis. Pattern Recog., 2020, pp. 12 546–12 555.

[58] Y. Zeng, M. Feng, H. Lu, G. Yang, and A. Borji, "An unsupervised game-theoretic approach to saliency detection," IEEE Trans. Image Process., vol. 27, no. 9, pp. 4545–4554, 2018.

[59] J. Zhang, T. Zhang, Y. Dai, M. Harandi, and R. Hartley, "Deep unsupervised saliency detection: A multiple noisy labeling perspective," in IEEE Conf. Comput. Vis. Pattern Recog., 2018, pp. 9029–9038.

[60] T. Nguyen, M. Dax, C. K. Mummadi, N. Ngo, T. H. P. Nguyen, Z. Lou, and T. Brox, "Deepusps: Deep robust unsupervised saliency prediction via self-supervision," in Adv. Neural Inform. Process. Syst., vol. 32, 2019, pp. 204–214.

[61] J. Wang, S. Zhu, J. Xu, and D. Cao, "The retrieval of the beautiful: Self-supervised salient object detection for beauty product retrieval," in ACM Int. Conf. Multimedia, 2019, pp. 2548–2552.

[62] G. Li, Y. Xie, L. Lin, and Y. Yu, "Instance-level salient object segmentation," in IEEE Conf. Comput. Vis. Pattern Recog., 2017, pp. 2386–2395.

[63] R. Fan, M.-M. Cheng, Q. Hou, T.-J. Mu, J. Wang, and S.-M. Hu, "S4net: Single stage salient-instance segmentation," in IEEE Conf. Comput. Vis. Pattern Recog., 2019, pp. 6103–6112.

[64] J. Zhang, S. Ma, M. Sameki, S. Sclaroff, M. Betke, Z. Lin, X. Shen, B. Price, and R. Mech, "Salient object subitizing," in IEEE Conf. Comput. Vis. Pattern Recog., 2015, pp. 4045–4054.

[65] M. A. Islam, M. Kalash, and N. D. Bruce, "Revisiting salient object detection: Simultaneous detection, ranking, and subitizing of multiple salient objects," in IEEE Conf. Comput. Vis. Pattern Recog., 2018, pp. 7142–7150.

[66] S. He, J. Jiao, X. Zhang, G. Han, and R. W. Lau, "Delving into salient object subitizing and detection," in IEEE Conf. Comput. Vis. Pattern Recog., 2017, pp. 1059–1067.

[67] M. Kalash, M. A. Islam, and N. D. Bruce, "Relative saliency and ranking: Models, metrics, data and benchmarks," IEEE Trans. Pattern Anal. Mach. Intell., vol. 43, no. 1, pp. 204–219, 2019.

[68] A. Siris, J. Jiao, G. K. Tam, X. Xie, and R. W. Lau, "Inferring attention shift ranks of objects for image saliency," in IEEE Conf. Comput. Vis. Pattern Recog., 2020, pp. 12 133–12 143.

[69] A. Borji and L. Itti, "State-of-the-art in visual attention modeling," IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 1, pp. 185–207, 2012.

[70] A. Borji, M.-M. Cheng, H. Jiang, and J. Li, "Salient object detection: A benchmark," IEEE Trans. Image Process., vol. 24, no. 12, pp. 5706–5722, 2015.

[71] T. V. Nguyen, Q. Zhao, and S. Yan, "Attentive systems: A survey," Int. J. Comput. Vis., vol. 126, no. 1, pp. 86–110, 2018.

[72] A. Borji, M.-M. Cheng, Q. Hou, H. Jiang, and J. Li, "Salient object detection: A survey," Comput. Vis. Media, vol. 5, no. 2, pp. 117–150, 2019.

[73] D. Zhang, H. Fu, J. Han, A. Borji, and X. Li, "A review of co-saliency detection algorithms: Fundamentals, applications, and challenges," ACM Trans. Intell. Syst. Technol., vol. 9, no. 4, pp. 1–31, 2018.

[74] R. Cong, J. Lei, H. Fu, M.-M. Cheng, W. Lin, and Q. Huang, "Review of visual saliency detection with comprehensive information," IEEE Trans. Circ. Syst. Video Technol., vol. 29, no. 10, pp. 2941–2959, 2018.

[75] 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 Process. Mag., vol. 35, no. 1, pp. 84–100, 2018.

[76] A. Borji, "Saliency prediction in the deep learning era: An empirical investigation," arXiv preprint arXiv:1810.03716, vol. 10, 2018.

[77] 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," IEEE Trans. Pattern Anal. Mach. Intell., 2021.

[78] 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 T. Neural Netw. Learn. Syst., 2021.

[79] T. Zhou, D.-P. Fan, M.-M. Cheng, J. Shen, and L. Shao, "Rgb-d salient object detection: A survey," Comput. Vis. Media, pp. 37–69, 2021.

[80] Y. Jiang, T. Zhou, G.-P. Ji, K. Fu, Q. Zhao, and D.-P. Fan, "Light field salient object detection: A review and benchmark," arXiv preprint arXiv:2010.04968, 2020.

[81] 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," IEEE Trans. Pattern Anal. Mach. Intell., 2021.

[82] J. Li, J. Su, C. Xia, and Y. Tian, "Distortion-adaptive salient object detection in 360° omnidirectional images," IEEE J. Sel. Top. Signal. Process., vol. 14, no. 1, pp. 38–48, 2019.

[83] R. Achanta, S. Hemami, F. Estrada, and S. Süsstrunk, "Frequency-tuned salient region detection," in IEEE Conf. Comput. Vis. Pattern Recog., 2009, pp. 1597–1604.

[84] V. Movahedi and J. H. Elder, "Design and perceptual validation of performance measures for salient object segmentation," in IEEE Conf. Comput. Vis. Pattern Recog. Worksh., 2010, pp. 49–56.

[85] M.-M. Cheng, G.-X. Zhang, N. J. Mitra, X. Huang, and S.-M. Hu, "Global contrast based salient region detection," in IEEE Conf. Comput. Vis. Pattern Recog., 2011, pp. 409–416.

[86] 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 IEEE Conf. Comput. Vis. Pattern Recog., 2017, pp. 4142–4150.

[87] H. Jiang, M.-M. Cheng, S.-J. Li, A. Borji, and J. Wang, "Joint salient object detection and existence prediction," Front. Comput. Sci., vol. 13, no. 4, pp. 778–788, 2019.

[88] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: Common objects in context," in Eur. Conf. Comput. Vis., 2014, pp. 740–755.

[89] P. Krähenbühl and V. Koltun, "Efficient inference in fully

connected crfs with gaussian edge potentials," in Adv. Neural Inform. Process. Syst., 2011.

[90] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," IEEE Trans. Pattern Anal. Mach. Intell., vol. 23, no. 11, pp. 1222–1239, 2001.

[91] C. Rother, K. Vladimir, and B. Andrew, "Grabcut: interactive foreground extraction using iterated graph cuts," ACM Trans. Graph., vol. 23, no. 3, pp. 309–314, 2004.

[92] J. Shi and M. Jitendra, "Normalized cuts and image segmentation," IEEE Trans. Pattern Anal. Mach. Intell., vol. 22, no. 8, pp. 888–905, 2000.

[93] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," IEEE Trans. Pattern Anal. Mach. Intell., vol. 26, no. 9, pp. 1124–1137, 2004.

[94] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in Adv. Neural Inform. Process. Syst., 2007, pp. 545–552.

[95] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," in IEEE Conf. Comput. Vis. Pattern Recog., 2007, pp. 1–8.

[96] N. Bruce and J. Tsotsos, "Saliency based on information maximization," in Adv. Neural Inform. Process. Syst., 2006, pp. 155–162.

[97] L. Zhang, M. H. Tong, T. K. Marks, H. Shan, and G. W. Cottrell, "SUN: A Bayesian framework for saliency using natural statistics," J. Vis., vol. 8, no. 7, pp. 32–32, 2008.

[98] Y.-F. Ma and H.-J. Zhang, "Contrast-based image attention analysis by using fuzzy growing," in ACM Int. Conf. Multimedia, 2003, pp. 374–381.

[99] R. Achanta, F. Estrada, P. Wils, and S. Süsstrunk, "Salient region detection and segmentation," in Int. Conf. Comput. Vis. Syst., 2008, pp. 66–75.

[100] E. Rahtu, J. Kannala, M. Salo, and J. Heikkilä, "Segmenting salient objects from images and videos," in Eur. Conf. Comput. Vis., 2010, pp. 366–379.

[101] R. Achanta and S. Süsstrunk, "Saliency detection using maximum symmetric surround," in IEEE Int. Conf. Image Process., 2010, pp. 2653–2656.

[102] R. Valenti, N. Sebe, T. Gevers et al., "Image saliency by isocentric curvedness and color." in Int. Conf. Comput. Vis., vol. 1, no. 3, 2009, p. 6.

[103] P. L. Rosin, "A simple method for detecting salient regions," Pattern Recogn., vol. 42, no. 11, pp. 2363–2371, 2009.

[104] F. Liu and M. Gleicher, "Region enhanced scale-invariant saliency detection," in Int. Conf. Multimedia and Expo, 2006, pp. 1477–1480.

[105] Y. Hu, D. Rajan, and L.-T. Chia, "Robust subspace analysis for detecting visual attention regions in images," in ACM Int. Conf. Multimedia, 2005, pp. 716–724.

[106] Z. Yu and H.-S. Wong, "A rule based technique for extraction of visual attention regions based on real-time clustering," IEEE Trans. Multimedia, vol. 9, no. 4, pp. 766–784, 2007.

[107] H. Yu, J. Li, Y. Tian, and T. Huang, "Automatic interesting object extraction from images using complementary saliency maps," in ACM Int. Conf. Multimedia, 2010, pp. 891–894.

[108] Y. Xie, H. Lu, and M.-H. Yang, "Bayesian saliency via low and mid level cues," IEEE Trans. Image Process., vol. 22, no. 5, pp. 1689–1698, 2012.

[109] E. Erdem and A. Erdem, "Visual saliency estimation by nonlinearly integrating features using region covariances," J. Vis., vol. 13, no. 4, pp. 11–11, 2013.

[110] C. Yang, L. Zhang, and H. Lu, "Graph-regularized saliency detection with convex-hull-based center prior," IEEE Signal Process. Let., vol. 20, no. 7, pp. 637–640, 2013.

[111] N. Tong, H. Lu, L. Zhang, and X. Ruan, "Saliency detection with multi-scale superpixels," IEEE Signal Process. Let., vol. 21, no. 9, pp. 1035–1039, 2014.

[112] H. Peng, B. Li, R. Ji, W. Hu, W. Xiong, and C. Lang, "Salient object detection via low-rank and structured sparse matrix decomposition," in AAAI Conf. Art. Intell., 2013.

[113] J. Sun, H. Lu, and S. Li, "Saliency detection based on integration of boundary and soft-segmentation," in IEEE Int. Conf. Image Process., 2012, pp. 1085–1088.

[114] F. Perazzi, P. Krähenbühl, Y. Pritch, and A. Hornung, "Saliency filters: Contrast based filtering for salient region detection," in IEEE Conf. Comput. Vis. Pattern Recog., 2012, pp. 733–740.

[115] H. Jiang, J. Wang, Z. Yuan, Y. Wu, N. Zheng, and S. Li, "Salient object detection: A discriminative regional feature integration approach," in IEEE Conf. Comput. Vis. Pattern Recog., 2013, pp. 2083–2090.

[116] W. Zhu, S. Liang, Y. Wei, and J. Sun, "Saliency optimization from robust background detection," in IEEE Conf. Comput. Vis. Pattern Recog., 2014, pp. 2814–2821.

[117] X. Shen and Y. Wu, "A unified approach to salient object detection via low rank matrix recovery," in IEEE Conf. Comput. Vis. Pattern Recog., 2012, pp. 853–860.

[118] R. Margolin, A. Tal, and L. Zelnik-Manor, "What makes a patch distinct?" in IEEE Conf. Comput. Vis. Pattern Recog., 2013, pp. 1139–1146.

[119] J. Kim, D. Han, Y.-W. Tai, and J. Kim, "Salient region detection via high-dimensional color transform," in IEEE Conf. Comput. Vis. Pattern Recog., 2014, pp. 883–890.

[120] L. Mai, Y. Niu, and F. Liu, "Saliency aggregation: A data-driven approach," in IEEE Conf. Comput. Vis. Pattern Recog., 2013, pp. 1131–1138.

[121] C. Scharfenberger, A. Wong, K. Fergani, J. S. Zelek, and D. A. Clausi, "Statistical textural distinctiveness for salient region detection in natural images," in IEEE Conf. Comput. Vis. Pattern Recog., 2013, pp. 979–986.

[122] R. Liu, J. Cao, Z. Lin, and S. Shan, "Adaptive partial differential equation learning for visual saliency detection," in IEEE Conf. Comput. Vis. Pattern Recog., 2014, pp. 3866–3873.

[123] Z. Jiang and L. S. Davis, "Submodular salient region detection," in IEEE Conf. Comput. Vis. Pattern Recog., 2013, pp. 2043–2050.

[124] K. Shi, K. Wang, J. Lu, and L. Lin, "PISA: Pixelwise image saliency by aggregating complementary appearance contrast measures with spatial priors," in IEEE Conf. Comput. Vis. Pattern Recog., 2013, pp. 2115–2122.

[125] X. Li, H. Lu, L. Zhang, X. Ruan, and M.-H. Yang, "Saliency detection via dense and sparse reconstruction," in Int. Conf. Comput. Vis., 2013, pp. 2976–2983.

[126] B. Jiang, L. Zhang, H. Lu, C. Yang, and M.-H. Yang, "Saliency detection via absorbing markov chain," in Int. Conf. Comput. Vis., 2013, pp. 1665–1672.

[127] M.-M. Cheng, J. Warrell, W.-Y. Lin, S. Zheng, V. Vineet, and

N. Crook, "Efficient salient region detection with soft image abstraction," in Int. Conf. Comput. Vis., 2013, pp. 1529–1536.

[128] K.-Y. Chang, T.-L. Liu, H.-T. Chen, and S.-H. Lai, "Fusing generic objectness and visual saliency for salient object detection," in Int. Conf. Comput. Vis., 2011, pp. 914–921.

[129] D. A. Klein and S. Frintrop, "Center-surround divergence of feature statistics for salient object detection," in Int. Conf. Comput. Vis., 2011, pp. 2214–2219.

[130] P. Jiang, H. Ling, J. Yu, and J. Peng, "Salient region detection by ufo: Uniqueness, focusness and objectness," in Int. Conf. Comput. Vis., 2013, pp. 1976–1983.

[131] X. Li, Y. Li, C. Shen, A. Dick, and A. Van Den Hengel, "Contextual hypergraph modeling for salient object detection," in Int. Conf. Comput. Vis., 2013, pp. 3328–3335.

[132] Y. Jia and M. Han, "Category-independent object-level saliency detection," in Int. Conf. Comput. Vis., 2013, pp. 1761–1768.

[133] Y. Lu, W. Zhang, H. Lu, and X. Xue, "Salient object detection using concavity context," in Int. Conf. Comput. Vis., 2011, pp. 233–240.

[134] Y. Wei, F. Wen, W. Zhu, and J. Sun, "Geodesic saliency using background priors," in Eur. Conf. Comput. Vis., 2012, pp. 29–42.

[135] H. Jiang, J. Wang, Z. Yuan, T. Liu, N. Zheng, and S. Li, "Automatic salient object segmentation based on context and shape prior," in Brit. Mach. Vis. Conf., vol. 6, no. 7, 2011, p. 9.

[136] W. Zou, K. Kpalma, Z. Liu, and J. Ronsin, "Segmentation driven low-rank matrix recovery for saliency detection," in Brit. Mach. Vis. Conf., 2013, pp. 1–13.

[137] H. Peng, B. Li, H. Ling, W. Hu, W. Xiong, and S. J. Maybank, "Salient object detection via structured matrix decomposition," IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 4, pp. 818–832, 2016.

[138] L. Zhang, C. Yang, H. Lu, X. Ruan, and M.-H. Yang, "Ranking saliency," IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 9, pp. 1892–1904, 2017.

[139] J. Wang, H. Lu, X. Li, N. Tong, and W. Liu, "Saliency detection via background and foreground seed selection," Neurocomputing, vol. 152, pp. 359–368, 2015.

[140] N. Tong, H. Lu, Y. Zhang, and X. Ruan, "Salient object detection via global and local cues," Pattern Recogn., vol. 48, no. 10, pp. 3258–3267, 2015.

[141] S. Chen, L. Zheng, X. Hu, and P. Zhou, "Discriminative saliency propagation with sink points," Pattern Recogn., vol. 60, pp. 2–12, 2016.

[142] H. Li, H. Lu, Z. Lin, X. Shen, and B. Price, "Inner and inter label propagation: salient object detection in the wild," IEEE Trans. Image Process., vol. 24, no. 10, pp. 3176–3186, 2015.

[143] J. Sun, H. Lu, and X. Liu, "Saliency region detection based on markov absorption probabilities," IEEE Trans. Image Process., vol. 24, no. 5, pp. 1639–1649, 2015.

[144] F. Huang, J. Qi, H. Lu, L. Zhang, and X. Ruan, "Salient object detection via multiple instance learning," IEEE Trans. Image Process., vol. 26, no. 4, pp. 1911–1922, 2017.

[145] Y. Yuan, C. Li, J. Kim, W. Cai, and D. D. Feng, "Reversion correction and regularized random walk ranking for saliency detection," IEEE Trans. Image Process., vol. 27, no. 3, pp. 1311–1322, 2017.

[146] G.-H. Liu and J.-Y. Yang, "Exploiting color volume and color difference for salient region detection," IEEE Trans. Image Process., vol. 28, no. 1, pp. 6–16, 2019.

[147] K. Fu, C. Gong, I. Y.-H. Gu, and J. Yang, "Normalized cut-based saliency detection by adaptive multi-level region merging," IEEE Trans. Image Process., vol. 24, no. 12, pp. 5671–5683, 2015.

[148] X. Huang and Y.-J. Zhang, "300-fps salient object detection via minimum directional contrast," IEEE Trans. Image Process., vol. 26, no. 9, pp. 4243–4254, 2017.

[149] Q. Liu, X. Hong, B. Zou, J. Chen, Z. Chen, and G. Zhao, "Hierarchical contour closure-based holistic salient object detection," IEEE Trans. Image Process., vol. 26, no. 9, pp. 4537–4552, 2017.

[150] Y. Kong, J. Zhang, H. Lu, and X. Liu, "Exemplar-aided salient object detection via joint latent space embedding," IEEE Trans. Image Process., vol. 27, no. 10, pp. 5167–5177, 2018.

[151] S. Huo, Y. Zhou, J. Lei, N. Ling, and C. Hou, "Iterative feedback control-based salient object segmentation," IEEE Trans. Multimedia, vol. 20, no. 6, pp. 1350–1364, 2017.

[152] S. Huo, Y. Zhou, W. Xiang, and S.-Y. Kung, "Semisupervised learning based on a novel iterative optimization model for saliency detection," IEEE T. Neural Netw. Learn. Syst., vol. 30, no. 1, pp. 225–241, 2019.

[153] J. Zhang, S. Sclaroff, Z. Lin, X. Shen, B. Price, and R. Mech, "Minimum barrier salient object detection at 80 fps," in Int. Conf. Comput. Vis., 2015, pp. 1404–1412.

[154] P. Jiang, N. Vasconcelos, and J. Peng, "Generic promotion of diffusion-based salient object detection," in Int. Conf. Comput. Vis., 2015, pp. 217–225.

[155] Y. Qin, H. Lu, Y. Xu, and H. Wang, "Saliency detection via cellular automata," in IEEE Conf. Comput. Vis. Pattern Recog., 2015, pp. 110–119.

[156] N. Otsu, "A threshold selection method from gray-level histograms," IEEE Trans. Syst. Man Cybern. Syst., vol. 9, no. 1, pp. 62–66, 1979.

[157] N. Tong, H. Lu, X. Ruan, and M.-H. Yang, "Salient object detection via bootstrap learning," in IEEE Conf. Comput. Vis. Pattern Recog., 2015, pp. 1884–1892.

[158] F. Yang, H. Lu, and Y.-W. Chen, "Human tracking by multiple kernel boosting with locality affinity constraints," in Asian Conf. Comput. Vis., 2010, pp. 39–50.

[159] W.-C. Tu, S. He, Q. Yang, and S.-Y. Chien, "Real-time salient object detection with a minimum spanning tree," in IEEE Conf. Comput. Vis. Pattern Recog., 2016, pp. 2334–2342.

[160] C. Li, Y. Yuan, W. Cai, Y. Xia, and D. Dagan Feng, "Robust saliency detection via regularized random walks ranking," in IEEE Conf. Comput. Vis. Pattern Recog., 2015, pp. 2710–2717.

[161] C. Gong, D. Tao, W. Liu, S. J. Maybank, M. Fang, K. Fu, and J. Yang, "Saliency propagation from simple to difficult," in IEEE Conf. Comput. Vis. Pattern Recog., 2015, pp. 2531–2539.

[162] N. Li, B. Sun, and J. Yu, "A weighted sparse coding framework for saliency detection," in IEEE Conf. Comput. Vis. Pattern Recog., 2015, pp. 5216–5223.

[163] Y. Kong, L. Wang, X. Liu, H. Lu, and X. Ruan, "Pattern mining saliency," in Eur. Conf. Comput. Vis., 2016, pp. 583–598.

[164] Y. Liu, J. Han, Q. Zhang, and L. Wang, "Salient object detection via two-stage graphs," IEEE Trans. Circ. Syst. Video Technol., vol. 29, no. 4, pp. 1023–1037, 2019.

[165] Y. Zhou, T. Zhang, S. Huo, C. Hou, and S.-Y. Kung, "Adaptive irregular graph construction-based salient object detection,"

IEEE Trans. Circ. Syst. Video Technol., vol. 30, no. 6, pp. 1569–1582, 2020.

[166] Y. Zhou, A. Mao, S. Huo, J. Lei, and S.-Y. Kung, “Salient object detection via fuzzy theory and object-level enhancement,” IEEE Trans. Multimedia, vol. 21, no. 1, pp. 74–85, 2019.

[167] X. Lin, Z.-J. Wang, L. Ma, and X. Wu, “Saliency detection via multi-scale global cues,” IEEE Trans. Multimedia, vol. 21, no. 7, pp. 1646–1659, 2019.

[168] Y. Xu, X. Hong, F. Porikli, X. Liu, J. Chen, and G. Zhao, “Saliency integration: An arbitrator model,” IEEE Trans. Multimedia, vol. 21, no. 1, pp. 98–113, 2019.

[169] N. Liu and J. Han, “Dhsnet: Deep hierarchical saliency network for salient object detection,” in IEEE Conf. Comput. Vis. Pattern Recog., 2016, pp. 678–686.

[170] L. Zhang, J. Sun, T. Wang, Y. Min, and H. Lu, “Visual saliency detection via kernelized subspace ranking with active learning,” IEEE Trans. Image Process., vol. 29, pp. 2258–2270, 2020.

[171] X. Huang, Y. Zheng, J. Huang, and Y.-J. Zhang, “50 fps object-level saliency detection via maximally stable region,” IEEE Trans. Image Process., vol. 29, pp. 1384–1396, 2020.

[172] R. Strand, K. C. Ciesielski, F. Malmberg, and P. K. Saha, “The minimum barrier distance,” Comput. Vis. Image Underst., vol. 117, no. 4, pp. 429–437, 2013.

[173] Y.-Y. Zhang, S. Zhang, P. Zhang, H.-Z. Song, and X.-G. Zhang, “Local regression ranking for saliency detection,” IEEE Trans. Image Process., vol. 29, pp. 1536–1547, 2020.

[174] M. Everingham, L. Van Gool, C. Williams, J. Winn, and A. Zisserman, “The pascal visual object classes challenge 2012 (voc2012) results (2012),” in URL http://www. pascal-network. org/challenges/VOC/voc2011/workshop/index. html, 2011.

[175] S. He, R. W. Lau, W. Liu, Z. Huang, and Q. Yang, “Supercnn: A superpixelwise convolutional neural network for salient object detection,” Int. J. Comput. Vis., vol. 115, no. 3, pp. 330–344, 2015.

[176] L. Wang, H. Lu, X. Ruan, and M.-H. Yang, “Deep networks for saliency detection via local estimation and global search,” in IEEE Conf. Comput. Vis. Pattern Recog., 2015, pp. 3183–3192.

[177] R. Zhao, W. Ouyang, H. Li, and X. Wang, “Saliency detection by multi-context deep learning,” in IEEE Conf. Comput. Vis. Pattern Recog., 2015, pp. 1265–1274.

[178] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, “Going deeper with convolutions,” in IEEE Conf. Comput. Vis. Pattern Recog., 2015, pp. 1–9.

[179] Y. Yuan, C. Li, J. Kim, W. Cai, and D. D. Feng, “Dense and sparse labeling with multidimensional features for saliency detection,” IEEE Trans. Circ. Syst. Video Technol., vol. 28, no. 5, pp. 1130–1143, 2016.

[180] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” IEEE Inst. Electr. Electron. Eng., vol. 86, no. 11, pp. 2278–2324, 1998.

[181] T. Chen, L. Lin, L. Liu, X. Luo, and X. Li, “DISC: Deep image saliency computing via progressive representation learning,” IEEE T. Neural Netw. Learn. Syst., vol. 27, no. 6, pp. 1135–1149, 2016.

[182] X. Li, L. Zhao, L. Wei, M.-H. Yang, F. Wu, Y. Zhuang, H. Ling, and J. Wang, “Deepsaliency: Multi-task deep neural network model for salient object detection,” IEEE Trans. Image Process., vol. 25, no. 8, pp. 3919–3930, 2016.

[183] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” in Int. Conf. Learn. Represent., 2015.

[184] J. Kim and V. Pavlovic, “A shape-based approach for salient object detection using deep learning,” in Eur. Conf. Comput. Vis., 2016, pp. 455–470.

[185] A. Krizhevsky, S. Ilya, and G. E Hinton, “Imagenet classification with deep convolutional neural networks,” in Adv. Neural Inform. Process. Syst., 2012.

[186] Y. Tang and X. Wu, “Saliency detection via combining region-level and pixel-level predictions with cnns,” in Eur. Conf. Comput. Vis., 2016, pp. 809–825.

[187] L. Wang, L. Wang, H. Lu, P. Zhang, and X. Ruan, “Saliency detection with recurrent fully convolutional networks,” in Eur. Conf. Comput. Vis., 2016, pp. 825–841.

[188] J. Zhang, S. Sclaroff, Z. Lin, X. Shen, B. Price, and R. Mech, “Unconstrained salient object detection via proposal subset optimization,” in IEEE Conf. Comput. Vis. Pattern Recog., 2016, pp. 5733–5742.

[189] S. S. Kruthiventi, V. Gudisa, J. H. Dholakiya, and R. Venkatesh Babu, “Saliency unified: A deep architecture for simultaneous eye fixation prediction and salient object segmentation,” in IEEE Conf. Comput. Vis. Pattern Recog., 2016, pp. 5781–5790.

[190] M. Jiang, S. Huang, J. Duan, and Q. Zhao, “SALICON: Saliency in context,” in IEEE Conf. Comput. Vis. Pattern Recog., 2015, pp. 1072–1080.

[191] J. Kuen, Z. Wang, and G. Wang, “Recurrent attentional networks for saliency detection,” in IEEE Conf. Comput. Vis. Pattern Recog., 2016, pp. 3668–3677.

[192] R. Ju, L. Ge, W. Geng, T. Ren, and G. Wu, “Depth saliency based on anisotropic center-surround difference,” in IEEE Int. Conf. Image Process., 2014, pp. 1115–1119.

[193] H. Peng, B. Li, W. Xiong, W. Hu, and R. Ji, “Rgbd salient object detection: a benchmark and algorithms,” in Eur. Conf. Comput. Vis., 2014, pp. 92–109.

[194] G. Lee, Y.-W. Tai, and J. Kim, “Deep saliency with encoded low level distance map and high level features,” in IEEE Conf. Comput. Vis. Pattern Recog., 2016, pp. 660–668.

[195] G. Li and Y. Yu, “Deep contrast learning for salient object detection,” in IEEE Conf. Comput. Vis. Pattern Recog., 2016, pp. 478–487.

[196] P. Hu, B. Shuai, J. Liu, and G. Wang, “Deep level sets for salient object detection,” in IEEE Conf. Comput. Vis. Pattern Recog., 2017, pp. 2300–2309.

[197] T. Wang, A. Borji, L. Zhang, P. Zhang, and H. Lu, “A stagewise refinement model for detecting salient objects in images,” in IEEE Conf. Comput. Vis. Pattern Recog., 2017, pp. 4019–4028.

[198] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in IEEE Conf. Comput. Vis. Pattern Recog., 2016, pp. 770–778.

[199] Z. Luo, A. Mishra, A. Achkar, J. Eichel, S. Li, and P.-M. Jodoin, “Non-local deep features for salient object detection,” in IEEE Conf. Comput. Vis. Pattern Recog., 2017, pp. 6609–6617.

[200] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “Imagenet: A large-scale hierarchical image database,” in IEEE Conf. Comput. Vis. Pattern Recog., 2009, pp. 248–255.

[201] Q. Hou, M.-M. Cheng, X. Hu, A. Borji, Z. Tu, and P. H. Torr, "Deeply supervised salient object detection with short connections," in IEEE Conf. Comput. Vis. Pattern Recog., 2017, pp. 3203–3212.

[202] X. Chen, A. Zheng, J. Li, and F. Lu, "Look, perceive and segment: Finding the salient objects in images via two-stream fixation-semantic cnns," in Int. Conf. Comput. Vis., 2017, pp. 1050–1058.

[203] D. Zhang, J. Han, and Y. Zhang, "Supervision by fusion: Towards unsupervised learning of deep salient object detector," in Int. Conf. Comput. Vis., 2017, pp. 4048–4056.

[204] P. Zhang, D. Wang, H. Lu, H. Wang, and B. Yin, "Learning uncertain convolutional features for accurate saliency detection," in Int. Conf. Comput. Vis., 2017, pp. 212–221.

[205] P. Zhang, D. Wang, H. Lu, H. Wang, and X. Ruan, "Amulet: Aggregating multi-level convolutional features for salient object detection," in Int. Conf. Comput. Vis., 2017, pp. 202–211.

[206] S. Chen, B. Wang, X. Tan, and X. Hu, "Embedding attention and residual network for accurate salient object detection," IEEE Trans. Cybern., 2018.

[207] K. Fu, Q. Zhao, and I. Y.-H. Gu, "Refinet: A deep segmentation assisted refinement network for salient object detection," IEEE Trans. Multimedia, vol. 21, no. 2, pp. 457–469, 2018.

[208] C. Cao, Y. Huang, Z. Wang, L. Wang, N. Xu, and T. Tan, "Lateral inhibition-inspired convolutional neural network for visual attention and saliency detection," in AAAI Conf. Art. Intell., 2018.

[209] X. Hu, L. Zhu, J. Qin, C.-W. Fu, and P.-A. Heng, "Recurrently aggregating deep features for salient object detection," in AAAI Conf. Art. Intell., 2018.

[210] Z. Deng, X. Hu, L. Zhu, X. Xu, J. Qin, G. Han, and P.-A. Heng, "R3net: Recurrent residual refinement network for saliency detection," in AAAI Conf. Art. Intell., 2018, pp. 684–690.

[211] S. Xie, R. Girshick, P. Dollar, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in IEEE Conf. Comput. Vis. Pattern Recog., 2017, pp. 1492–1500.

[212] X. Li, F. Yang, H. Cheng, W. Liu, and D. Shen, "Contour knowledge transfer for salient object detection," in Eur. Conf. Comput. Vis., 2018, pp. 355–370.

[213] S. Chen, X. Tan, B. Wang, and X. Hu, "Reverse attention for salient object detection," in Eur. Conf. Comput. Vis., 2018, pp. 234–250.

[214] Y. Zeng, H. Lu, L. Zhang, M. Feng, and A. Borji, "Learning to promote saliency detectors," in IEEE Conf. Comput. Vis. Pattern Recog., 2018, pp. 1644–1653.

[215] M. Amirul Islam, M. Kalash, and N. D. Bruce, "Revisiting salient object detection: Simultaneous detection, ranking, and subitizing of multiple salient objects," in IEEE Conf. Comput. Vis. Pattern Recog., 2018, pp. 7142–7150.

[216] W. Wang, J. Shen, X. Dong, and A. Borji, "Salient object detection driven by fixation prediction," in IEEE Conf. Comput. Vis. Pattern Recog., 2018, pp. 1711–1720.

[217] L. Zhang, J. Dai, H. Lu, Y. He, and G. Wang, "A bi-directional message passing model for salient object detection," in IEEE Conf. Comput. Vis. Pattern Recog., 2018, pp. 1741–1750.

[218] T. Wang, L. Zhang, S. Wang, H. Lu, G. Yang, X. Ruan, and A. Borji, "Detect globally, refine locally: A novel approach to saliency detection," in IEEE Conf. Comput. Vis. Pattern Recog., 2018, pp. 3127–3135.

[219] N. Liu, J. Han, and M.-H. Yang, "Picanet: Learning pixel-wise contextual attention for saliency detection," in IEEE Conf. Comput. Vis. Pattern Recog., 2018, pp. 3089–3098.

[220] X. Zhang, T. Wang, J. Qi, H. Lu, and G. Wang, "Progressive attention guided recurrent network for salient object detection," in IEEE Conf. Comput. Vis. Pattern Recog., 2018, pp. 714–722.

[221] S. Zhou, J. Wang, F. Wang, and D. Huang, "Se2net: Siamese edge-enhancement network for salient object detection," arXiv preprint arXiv:1904.00048, 2019.

[222] Z. Li, C. Lang, Y. Chen, J. Liew, and J. Feng, "Deep reasoning with multi-scale context for salient object detection," arXiv preprint arXiv:1901.08362, 2019.

[223] S. Jia and N. D. Bruce, "Richer and deeper supervision network for salient object detection," arXiv preprint arXiv:1901.02425, 2019.

[224] L. Zhu, J. Chen, X. Hu, C.-W. Fu, X. Xu, J. Qin, and P.-A. Heng, "Aggregating attentional dilated features for salient object detection," IEEE Trans. Circ. Syst. Video Technol., vol. 30, no. 10, pp. 3358–3371, 2019.

[225] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in IEEE Conf. Comput. Vis. Pattern Recog., 2017, pp. 4700–4708.

[226] Y. Tang and X. Wu, "Salient object detection using cascaded convolutional neural networks and adversarial learning," IEEE Trans. Multimedia, vol. 21, no. 9, pp. 2237–2247, 2019.

[227] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," IEEE Trans. Pattern Anal. Mach. Intell., vol. 40, no. 4, pp. 834–848, 2018.

[228] Y. Wang, X. Zhao, X. Hu, Y. Li, and K. Huang, "Focal boundary guided salient object detection," IEEE Trans. Image Process., vol. 28, no. 6, pp. 2813–2824, 2019.

[229] T. V. Nguyen, K. Nguyen, and T.-T. Do, "Semantic prior analysis for salient object detection," IEEE Trans. Image Process., vol. 28, no. 6, pp. 3130–3141, 2019.

[230] M. Kampffmeyer, N. Dong, X. Liang, Y. Zhang, and E. P. Xing, "Connnet: A long-range relation-aware pixel-connectivity network for salient segmentation," IEEE Trans. Image Process., vol. 28, no. 5, pp. 2518–2529, 2019.

[231] P. Zhang, W. Liu, H. Lu, and C. Shen, "Salient object detection with lossless feature reflection and weighted structural loss," IEEE Trans. Image Process., vol. 28, no. 6, pp. 3048–3060, 2019.

[232] D. Zhang, J. Han, Y. Zhang, and D. Xu, "Synthesizing supervision for learning deep saliency network without human annotation," IEEE Trans. Pattern Anal. Mach. Intell., vol. 42, no. 7, pp. 1755–1769, 2019.

[233] C. Li, R. Cong, J. Hou, S. Zhang, Y. Qian, and S. Kwong, "Nested network with two-stream pyramid for salient object detection in optical remote sensing images," IEEE Trans. Geosci. Remote. Sens., vol. 57, no. 11, pp. 9156–9166, 2019.

[234] K. Fu, Q. Zhao, I. Y.-H. Gu, and J. Yang, "Deepside: A general deep framework for salient object detection," Neurocomputing, vol. 356, pp. 69–82, 2019.

[235] B. Li, Z. Sun, and Y. Guo, "Supervae: Superpixelwise variational autoencoder for salient object detection," in AAAI Conf. Art. Intell., vol. 33, no. 01, 2019, pp. 8569–8576.

[236] Y. Zhuge, Y. Zeng, and H. Lu, "Deep embedding features for salient object detection," in AAAI Conf. Art. Intell., 2019.

[237] Y. Zeng, Y. Zhuge, H. Lu, L. Zhang, M. Qian, and Y. Yu, “Multi-source weak supervision for saliency detection,” in IEEE Conf. Comput. Vis. Pattern Recog., 2019, pp. 6074–6083.

[238] R. Wu, M. Feng, W. Guan, D. Wang, H. Lu, and E. Ding, “A mutual learning method for salient object detection with intertwined multi-supervision,” in IEEE Conf. Comput. Vis. Pattern Recog., 2019, pp. 8150–8159.

[239] W. Wang, J. Shen, M.-M. Cheng, and L. Shao, “An iterative and cooperative top-down and bottom-up inference network for salient object detection,” in IEEE Conf. Comput. Vis. Pattern Recog., 2019, pp. 5968–5977.

[240] M. Feng, H. Lu, and E. Ding, “Attentive feedback network for boundary-aware salient object detection,” in IEEE Conf. Comput. Vis. Pattern Recog., 2019, pp. 1623–1632.

[241] T. Zhao and X. Wu, “Pyramid feature attention network for saliency detection,” in IEEE Conf. Comput. Vis. Pattern Recog., 2019, pp. 3085–3094.

[242] W. Wang, S. Zhao, J. Shen, S. C. Hoi, and A. Borji, “Salient object detection with pyramid attention and salient edges,” in IEEE Conf. Comput. Vis. Pattern Recog., 2019, pp. 1448–1457.

[243] Z. Wu, L. Su, and Q. Huang, “Cascaded partial decoder for fast and accurate salient object detection,” in IEEE Conf. Comput. Vis. Pattern Recog., 2019, pp. 3907–3916.

[244] J.-J. Liu, Q. Hou, M.-M. Cheng, J. Feng, and J. Jiang, “A simple pooling-based design for real-time salient object detection,” in IEEE Conf. Comput. Vis. Pattern Recog., 2019, pp. 3917–3926.

[245] X. Qin, Z. Zhang, C. Huang, C. Gao, M. Dehghan, and M. Jagersand, “BASNet: Boundary-Aware Salient Object Detection,” in IEEE Conf. Comput. Vis. Pattern Recog., 2019, pp. 7479–7489.

[246] G. Xavier and B. Yoshua, “Understanding the difficulty of training deep feedforward neural networks,” in Int. Conf. Artif. Intell. Stat., 2010, pp. 249–256.

[247] Y. Xu, D. Xu, X. Hong, W. Ouyang, R. Ji, M. Xu, and G. Zhao, “Structured modeling of joint deep feature and prediction refinement for salient object detection,” in Int. Conf. Comput. Vis., 2019, pp. 3789–3798.

[248] Y. Liu, Q. Zhang, D. Zhang, and J. Han, “Employing deep part-object relationships for salient object detection,” in Int. Conf. Comput. Vis., 2019, pp. 1232–1241.

[249] Y. Zeng, Y. Zhuge, H. Lu, and L. Zhang, “Joint learning of saliency detection and weakly supervised semantic segmentation,” in Int. Conf. Comput. Vis., 2019, pp. 7223–7233.

[250] J. Su, J. Li, Y. Zhang, C. Xia, and Y. Tian, “Selectivity or invariance: Boundary-aware salient object detection,” in Int. Conf. Comput. Vis., 2019, pp. 3799–3808.

[251] 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 Int. Conf. Comput. Vis., 2019, pp. 8779–8788.

[252] S. Zhou, J. Wang, J. Zhang, L. Wang, D. Huang, S. Du, and N. Zheng, “Hierarchical u-shape attention network for salient object detection,” IEEE Trans. Image Process., vol. 29, pp. 8417–8428, 2020.

[253] Y. Cai, L. Dai, H. Wang, L. Chen, and Y. Li, “A novel saliency detection algorithm based on adversarial learning model,” IEEE Trans. Image Process., vol. 29, pp. 4489–4504, 2020.

[254] X. Li, D. Song, and Y. Dong, “Hierarchical feature fusion network for salient object detection,” IEEE Trans. Image Process., vol. 29, pp. 9165–9175, 2020.

[255] J.-J. Liu, Q. Hou, and M.-M. Cheng, “Dynamic feature integration for simultaneous detection of salient object, edge, and skeleton,” IEEE Trans. Image Process., vol. 29, pp. 8652–8667, 2020.

[256] M. Feng, H. Lu, and Y. Yu, “Residual learning for salient object detection,” IEEE Trans. Image Process., vol. 29, pp. 4696–4708, 2020.

[257] L. Zhang, J. Wu, T. Wang, A. Borji, G. Wei, and H. Lu, “A multistage refinement network for salient object detection,” IEEE Trans. Image Process., vol. 29, pp. 3534–3545, 2020.

[258] Y. Liu, J. Han, Q. Zhang, and C. Shan, “Deep salient object detection with contextual information guidance,” IEEE Trans. Image Process., vol. 29, pp. 360–374, 2020.

[259] S. Chen, X. Tan, B. Wang, H. Lu, X. Hu, and Y. Fu, “Reverse attention-based residual network for salient object detection,” IEEE Trans. Image Process., vol. 29, pp. 3763–3776, 2020.

[260] W. Wang, J. Shen, X. Dong, A. Borji, and R. Yang, “Inferring salient objects from human fixations,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 42, no. 8, pp. 1913–1927, 2020.

[261] H. Li, G. Li, B. Yang, G. Chen, L. Lin, and Y. Yu, “Depthwise nonlocal module for fast salient object detection using a single thread,” IEEE Trans. Cybern., 2020.

[262] J. Li, Z. Pan, Q. Liu, Y. Cui, and Y. Sun, “Complementarity-aware attention network for salient object detection,” IEEE Trans. Cybern., 2020.

[263] H. Li, G. Li, and Y. Yu, “Rosa: Robust salient object detection against adversarial attacks,” IEEE Trans. Cybern., vol. 50, no. 11, pp. 4835–4847, 2019.

[264] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” in IEEE Conf. Comput. Vis. Pattern Recog., 2015, pp. 3431–3440.

[265] L. Wang, R. Chen, L. Zhu, H. Xie, and X. Li, “Deep sub-region network for salient object detection,” IEEE Trans. Circ. Syst. Video Technol., 2020.

[266] Z. Tu, Y. Ma, C. Li, J. Tang, and B. Luo, “Edge-guided non-local fully convolutional network for salient object detection,” IEEE Trans. Circ. Syst. Video Technol., 2020.

[267] X. Hu, C.-W. Fu, L. Zhu, T. Wang, and P.-A. Heng, “Sac-net: Spatial attenuation context for salient object detection,” IEEE Trans. Circ. Syst. Video Technol., 2020.

[268] Q. Ren, S. Lu, J. Zhang, and R. Hu, “Salient object detection by fusing local and global contexts,” IEEE Trans. Multimedia, 2020.

[269] Z. Wu, S. Li, C. Chen, A. Hao, and H. Qin, “A deeper look at image salient object detection: Bi-stream network with a small training dataset,” IEEE Trans. Multimedia, 2020.

[270] J. Li, Z. Pan, Q. Liu, and Z. Wang, “Stacked u-shape network with channel-wise attention for salient object detection,” IEEE Trans. Multimedia, 2020.

[271] G. Ma, C. Chen, S. Li, C. Peng, A. Hao, and H. Qin, “Salient object detection via multiple instance joint re-learning,” IEEE Trans. Multimedia, vol. 22, no. 2, pp. 324–336, 2020.

[272] S. Mohammadi, M. Noori, A. Bahri, S. G. Majelan, and M. Havaei, “Cagnet: Content-aware guidance for salient object detection,” Pattern Recogn., vol. 103, p. 107303, 2020.

[273] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, “Learning transferable architectures for scalable image recognition,” in IEEE Conf. Comput. Vis. Pattern Recog., 2018, pp. 8697–8710.

[274] X. Qin, Z. Zhang, C. Huang, M. Dehghan, O. R. Zaiane, and M. Jagersand, “U2-net: Going deeper with nested u-structure

for salient object detection," Pattern Recogn., vol. 106, p. 107404, 2020.

[275] C. Wang, S. Dong, X. Zhao, G. Papanastasiou, H. Zhang, and G. Yang, "Saliencygan: Deep learning semisupervised salient object detection in the fog of iot," IEEE Trans. Industr. Inform., vol. 16, no. 4, pp. 2667–2676, 2020.

[276] S. Song, H. Yu, Z. Miao, J. Fang, K. Zheng, C. Ma, and S. Wang, "Multi-spectral salient object detection by adversarial domain adaptation," in AAAI Conf. Art. Intell., vol. 34, no. 07, 2020, pp. 12 023–12 030.

[277] B. Wang, Q. Chen, M. Zhou, Z. Zhang, X. Jin, and K. Gai, "Progressive feature polishing network for salient object detection," in AAAI Conf. Art. Intell., vol. 34, no. 07, 2020, pp. 12 128–12 135.

[278] Z. Chen, Q. Xu, R. Cong, and Q. Huang, "Global context-aware progressive aggregation network for salient object detection," in AAAI Conf. Art. Intell., vol. 34, no. 07, 2020, pp. 10 599–10 606.

[279] J. Wei, S. Wang, and Q. Huang, "F^3net: Fusion, feedback and focus for salient object detection," in AAAI Conf. Art. Intell., vol. 34, no. 07, 2020, pp. 12 321–12 328.

[280] J. Wei, S. Wang, Z. Wu, C. Su, Q. Huang, and Q. Tian, "Label decoupling framework for salient object detection," in IEEE Conf. Comput. Vis. Pattern Recog., 2020, pp. 13 025–13 034.

[281] H. Zhou, X. Xie, J.-H. Lai, Z. Chen, and L. Yang, "Interactive two-stream decoder for accurate and fast saliency detection," in IEEE Conf. Comput. Vis. Pattern Recog., 2020, pp. 9141–9150.

[282] Y. Pang, X. Zhao, L. Zhang, and H. Lu, "Multi-scale interactive network for salient object detection," in IEEE Conf. Comput. Vis. Pattern Recog., 2020, pp. 9413–9422.

[283] J. Zhang, J. Xie, and N. Barnes, "Learning noise-aware encoder-decoder from noisy labels by alternating back-propagation for saliency detection," in Eur. Conf. Comput. Vis., 2020.

[284] S.-H. Gao, Y.-Q. Tan, M.-M. Cheng, C. Lu, Y. Chen, and S. Yan, "Highly efficient salient object detection with 100k parameters," in Eur. Conf. Comput. Vis., 2020, pp. 702–721.

[285] X. Zhao, Y. Pang, L. Zhang, H. Lu, and L. Zhang, "Suppress and balance: A simple gated network for salient object detection," in Eur. Conf. Comput. Vis., 2020, pp. 35–51.

[286] Y. Liu, M.-M. Cheng, X. Zhang, G.-Y. Nie, and M. Wang, "Dna: Deeply-supervised nonlinear aggregation for salient object detection," IEEE Trans. Cybern., 2021.

[287] Z. Chen, H. Zhou, J. Lai, L. Yang, and X. Xie, "Contour-aware loss: Boundary-aware learning for salient object segmentation," IEEE Trans. Image Process., vol. 30, pp. 431–443, 2020.

[288] S. Zhou, J. Wang, L. Wang, J. Zhang, F. Wang, D. Huang, and N. Zheng, "Hierarchical and interactive refinement network for edge-preserving salient object detection," IEEE Trans. Image Process., vol. 30, pp. 1–14, 2020.

[289] S. Yu, B. Zhang, J. Xiao, and E. G. Lim, "Structure-consistent weakly supervised salient object detection with local saliency coherence," in AAAI Conf. Art. Intell., 2021.

[290] M. Ma, C. Xia, and J. Li, "Pyramidal feature shrinking for salient object detection," in AAAI Conf. Art. Intell., 2021.

[291] B. Xu, H. Liang, R. Liang, and P. Chen, "Locate globally, segment locally: A progressive architecture with knowledge review network for salient object detection," in AAAI Conf. Art. Intell., 2021.

[292] S.-H. Gao, M.-M. Cheng, K. Zhao, X.-Y. Zhang, M.-H. Yang, and P. Torr, "Res2net: A new multi-scale backbone architecture," IEEE Trans. Pattern Anal. Mach. Intell., vol. 43, no. 2, pp. 652–662, 2021.

[293] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in IEEE Conf. Comput. Vis. Pattern Recog., 2016, pp. 2818–2826.

[294] K. K. Singh and Y. J. Lee, "Hide-and-seek: Forcing a network to be meticulous for weakly-supervised object and action localization," in Int. Conf. Comput. Vis., 2017, pp. 3544–3553.

[295] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "mixup: Beyond empirical risk minimization," in Int. Conf. Learn. Represent., 2017.

[296] Z. Feng, C. Xu, and D. Tao, "Self-supervised representation learning by rotation feature decoupling," in IEEE Conf. Comput. Vis. Pattern Recog., 2019, pp. 10 364–10 374.

[297] L. Xie, J. Wang, Z. Wei, M. Wang, and Q. Tian, "Disturblabel: Regularizing cnn on the loss layer," in IEEE Conf. Comput. Vis. Pattern Recog., 2016, pp. 4753–4762.

[298] T. Miyato, S.-i. Maeda, M. Koyama, K. Nakae, and S. Ishii, "Distributional smoothing with virtual adversarial training," in Int. Conf. Learn. Represent., 2016.

[299] S. Wager, W. Fithian, S. Wang, and P. Liang, "Altitude training: Strong bounds for single-layer dropout," in Adv. Neural Inform. Process. Syst., 2014, pp. 100–108.

[300] J. C. Peterson, R. M. Battleday, T. L. Griffiths, and O. Russakovsky, "Human uncertainty makes classification more robust," in Int. Conf. Comput. Vis., 2019, pp. 9617–9626.

[301] Y. Li, G. Hu, Y. Wang, T. M. Hospedales, N. M. obertson, and Y. Yang, "Differentiable automatic data augmentation," in Eur. Conf. Comput. Vis., 2020, pp. 580–595.

[302] E. D. Cubuk, B. Zoph, D. Mane, V. Vasudevan, and Q. V. Le, "Autoaugment: Learning augmentation strategies from data," in IEEE Conf. Comput. Vis. Pattern Recog., 2019, pp. 113–123.

[303] 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 IEEE Conf. Comput. Vis. Pattern Recog., 2017, pp. 1568–1576.

[304] Z. Zhong, L. Zheng, G. Kang, S. Li, and Y. Yang, "Random erasing data augmentation," in AAAI Conf. Art. Intell., vol. 34, no. 07, 2020, pp. 13 001–13 008.

[305] Y.-T. Chang, Q. Wang, W.-C. Hung, R. Piramuthu, Y.-H. Tsai, and M.-H. Yang, "Mixup-cam: Weakly-supervised semantic segmentation via uncertainty regularization," in Brit. Mach. Vis. Conf., 2020.

[306] H. Guo, Y. Mao, and R. Zhang, "Mixup as locally linear out-of-manifold regularization," in AAAI Conf. Art. Intell., vol. 33, 2019, pp. 3714–3722.

[307] S. Gidaris, P. Singh, and N. Komodakis, "Unsupervised representation learning by predicting image rotations," in Int. Conf. Learn. Represent., 2018.

[308] X. Zhai, A. Oliver, A. Kolesnikov, and L. Beyer, "S4l: Self-supervised semi-supervised learning," in Int. Conf. Comput. Vis., 2019, pp. 1476–1485.

[309] X. Chen, S. Xie, and K. He, "An empirical study of training self-supervised visual transformers," arXiv preprint arXiv:2104.02057, 2021.

[310] X. Wang, K. He, and A. Gupta, "Transitive invariance for self-supervised visual representation learning," in Int. Conf. Comput. Vis., 2017, pp. 1329–1338.

[311] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick, "Momentum

contrast for unsupervised visual representation learning," in IEEE Conf. Comput. Vis. Pattern Recog., 2020, pp. 9729–9738.

[312] H. Caesar, J. Uijlings, and V. Ferrari, "Coco-stuff: Thing and stuff classes in context," in IEEE Conf. Comput. Vis. Pattern Recog., 2018, pp. 1209–1218.

[313] S. Lazebnik, C. Schmid, and J. Ponce, "A sparse texture representation using local affine regions," IEEE Trans. Pattern Anal. Mach. Intell., vol. 27, no. 8, pp. 1265–1278, 2005.

[314] T. Judd, F. Durand, and A. Torralba, "A benchmark of computational models of saliency to predict human fixations," in MIT Technical Report, 2012.

[315] F. Perazzi, J. Pont-Tuset, B. McWilliams, L. Van Gool, M. Gross, and A. Sorkine-Hornung, "A benchmark dataset and evaluation methodology for video object segmentation," in IEEE Conf. Comput. Vis. Pattern Recog., 2016, pp. 724–732.

[316] Q. Hou, M.-M. Cheng, X. Hu, A. Borji, Z. Tu, and P. Torr, "Deeply supervised salient object detection with short connections," IEEE Trans. Pattern Anal. Mach. Intell., vol. 41, no. 4, pp. 815–828, 2019.

[317] L. Yuan, F. E. Tay, G. Li, T. Wang, and J. Feng, "Revisiting knowledge distillation via label smoothing regularization," in IEEE Conf. Comput. Vis. Pattern Recog., 2020, pp. 3903–3911.

[318] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," in Adv. Neural Inform. Process. Syst. Worksh., 2015.

[319] C.-B. Zhang, P.-T. Jiang, Q. Hou, Y. Wei, Q. Han, Z. Li, and M.-M. Cheng, "Delving deep into label smoothing," arXiv preprint arXiv:2011.12562, 2020.

[320] C. Godard, O. Mac Aodha, and G. J. Brostow, "Unsupervised monocular depth estimation with left-right consistency," in IEEE Conf. Comput. Vis. Pattern Recog., 2017, pp. 270–279.

[321] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," IEEE Trans. Image Process., vol. 13, no. 4, pp. 600–612, 2004.

[322] 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 Trans. Pattern Anal. Mach. Intell., 2021.

[323] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," IEEE Trans. Pattern Anal. Mach. Intell., vol. 34, no. 10, pp. 1915–1926, 2011.

[324] X. Zhu, C. Tang, P. Wang, H. Xu, M. Wang, J. Chen, and J. Tian, "Saliency detection via affinity graph learning and weighted manifold ranking," Neurocomputing, vol. 312, pp. 239–250, 2018.

[325] C. Tang, P. Wang, C. Zhang, and W. Li, "Salient object detection via weighted low rank matrix recovery," IEEE Signal Process. Let., vol. 24, no. 4, pp. 490–494, 2016.

[326] X. Huang and Y. Zhang, "Water flow driven salient object detection at 180 fps," Pattern Recogn., vol. 76, pp. 95–107, 2018.

[327] J. Zhang, D.-P. Fan, Y. Dai, S. Anwar, F. Saleh, S. Aliakbarian, and N. Barnes, "Uncertainty inspired rgb-d saliency detection," IEEE Trans. Pattern Anal. Mach. Intell., 2021.

[328] 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 IEEE Conf. Comput. Vis. Pattern Recog., 2020, pp. 8582–8591.

[329] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, "On calibration of modern neural networks," in Int. Conf. Mach. Learn., 2017, pp. 1321–1330.

[330] D.-P. Fan, G.-P. Ji, M.-M. Cheng, and L. Shao, "Concealed object detection," IEEE Trans. Pattern Anal. Mach. Intell., 2021.

[331] K. R. Cave, "Finding meaning in eye movements," Nat. Hum. Behav., vol. 1, no. 10, pp. 709–710, 2017.

[332] J. M. Henderson, T. R. Hayes, C. E. Peacock, and G. Rehrig, "Meaning and attentional guidance in scenes: A review of the meaning map approach," Vision, vol. 3, no. 2, p. 19, 2019.

[333] J. M. Henderson and T. R. Hayes, "Meaning-based guidance of attention in scenes as revealed by meaning maps," Nat. Hum. Behav., vol. 1, no. 10, pp. 743–747, 2017.

[334] X. Tian, K. Xu, X. Yang, B. Yin, and R. W. Lau, "Weakly-supervised salient instance detection," in Brit. Mach. Vis. Conf., 2020.

[335] Y.-H. Wu, Y. Liu, L. Zhang, W. Gao, and M.-M. Cheng, "Regularized densely-connected pyramid network for salient instance segmentation," IEEE Trans. Image Process., vol. 30, pp. 3897–3907, 2021.

[336] Y. Liu, D. Zhang, Q. Zhang, and J. Han, "Part-object relational visual saliency," IEEE Trans. Pattern Anal. Mach. Intell., 2021.

[337] A. R. Zamir, A. Sax, W. Shen, L. J. Guibas, J. Malik, and S. Savarese, "Taskonomy: Disentangling task transfer learning," in IEEE Conf. Comput. Vis. Pattern Recog., 2018, pp. 3712–3722.

[338] A. Li, J. Zhang, Y. Lv, B. Liu, T. Zhang, and Y. Dai, "Uncertainty-aware joint salient object and camouflaged object detection," in IEEE Conf. Comput. Vis. Pattern Recog., 2021.

[339] M. Zhuge, D. Gao, D.-P. Fan, L. Jin, B. Chen, H. Zhou, M. Qiu, and L. Shao, "Kaleido-bert: Vision-language pre-training on fashion domain," in IEEE Conf. Comput. Vis. Pattern Recog., 2021.

[340] G. Wang, C. Chen, D.-P. Fan, A. Hao, and H. Qin, "From semantic categories to fixations: A novel weakly-supervised visual-auditory saliency detection approach," in IEEE Conf. Comput. Vis. Pattern Recog., 2021.

[341] D. V. Ruiz, B. A. Krinski, and E. Todt, "Ida: Improved data augmentation applied to salient object detection," in 2020 33rd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), 2020, pp. 210–217.

[342] Y. Liu, X.-Y. Zhang, J.-W. Bian, L. Zhang, and M.-M. Cheng, "Samnet: Stereoscopically attentive multi-scale network for lightweight salient object detection," IEEE Trans. Image Process., vol. 30, pp. 3804–3814, 2021.

[343] D.-P. Fan, G.-P. Ji, X. Qin, and M.-M. Cheng, "Cognitive vision inspired object segmentation metric and loss function," SCIENTIA SINICA Informationis, 2020.

[344] Y. Mao, J. Zhang, Z. Wan, Y. Dai, A. Li, Y. Lv, X. Tian, D.-P. Fan, and N. Barnes, "Transformer transforms salient object detection and camouflaged object detection," arXiv preprint arXiv:2104.10127, 2021.

[345] N. Liu, N. Zhang, K. Wan, J. Han, and L. Shao, "Visual saliency transformer," arXiv preprint arXiv:2104.12099, 2021.

[346] H. Zhang, Y. Zeng, H. Lu, L. Zhang, J. Li, and J. Qi, "Learning to detect salient object with multi-source weak supervision," IEEE Trans. Pattern Anal. Mach. Intell., 2021.

[347] X. Zhao, Y. Pang, L. Zhang, H. Lu, and X. Ruan, "Self-supervised representation learning for rgb-d salient object detection," arXiv preprint arXiv:2101.12482, 2021.

[348] R. Deora, R. Sharma, and D. S. S. Raj, "Salient image matting," arXiv preprint arXiv:2103.12337, 2021.