深入了解协同显著性物体检测

范登平 [1,2,*] 林铮 [1,*] 季葛鹏 [3] 张鼎文 [4] 付华柱 [2] 程明明 [1] ✉

[1] 南开大学 [2] 阿联酋起源人工智能研究院 [3] 武汉大学 [4] 西安电子科技大学

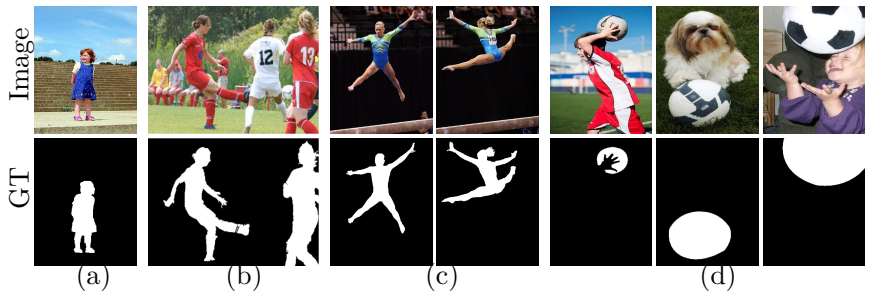

图 1: 不同的显著物体检测（SOD）任务。(a) 传统的显著性物体检测 [76]。(b) 图像内的协同显著性物体检测（CoSOD）[90]，其从单张图像中检测到常见的显著物体。(c) 现有的协同显著性物体检测，根据一对 [52] 或一组 [82] 外观相似的图像检测到显著物体。(d) 在自然环境下的 CoSOD 检测，需要大量的语义上下文，这使其比现存的 CoSOD 更具挑战性。基准工具包和结果见本文项目主页：http://dpfan.net/CoSOD3k/。

摘要

协同显著性物体检测（*CoSOD*）[1]是显著物体检测（*SOD*）的一个新兴且增长迅速的分支，目的是检测多个图像中同时出现的显著物体。然而，现有的 *CoSOD* 数据集往往存在严重的数据偏差，即假设每一组图像中都包含视觉外观相似的显著物体。这种偏差会导致过于理想的设置和影响模型的有效性，由于相似度通常是语义或概念上的，在现有数据集上训练的模型可能会在实际情况中表现不佳。为了解决这个问题，我们首先收集一个名为 ***CoSOD3k*** 的高质量新数据集，它包含了 *3,316* 张图像，共 *160* 组，提供了多个级别的标注信息，即类别，边界框，对象和实例。*CoSOD3k* 在多样性、难度和可扩展性方面取得了质的飞跃，使相关的视觉任务受益。此外，我们全面总结了 *34* 种最先进的算法，在四个现有数据集（*MSRC*, *iCoSeg*, *ImagePair* 和 *CoSal2015*）及 *CoSOD3k*（共计约 *61K* 最大规模图像）中对其中的 *19* 个进行了评测，并报告和分析了组级的性能。最后我们讨论了 *CoSOD* 的未来工作和挑战。我们的研究将大大促进 *CoSOD* 社区的发展。

1. 引言

在过去的几十年里，RGB 显著性物体检测 (SOD) [6, 18, 47, 91], RGB-D 显著性物体检测 [23, 26, 99, 104], 和视频显著性物体检测 [24] 已经成为计算机视觉领域的新兴研究方向 [30, 50, 72, 102]。SOD 模仿了人类的视觉系统，可以从单个图像中检测出最吸引人的物体，如 Fig. 1 (a) 所示。由于在面向集合的图像裁剪 [35]，协同分割 [78]，弱监督学习 [101]，图像检索 [11]，图像质量评估 [79] 和视频前景检测 [25] 等方面有众多应用，作为 SOD 的一个分支，通过一组图像协同地检测出显著性物体（CoSOD）的任务开始备受瞩目（见 Tab. 2）。

CoSOD 的目标是提取图像中共现的显著物体，例如在 Fig. 1中 (b) 穿红衣服的足球运动员或 (c) 蓝衣服的体操运动员。为了解决这个问题，现有的模型往往侧重于对象之间的外观相似性。然而，这可能会导致模型对数据选择出现偏差，这并不合适，因为，在实际应用中，一组图像中的显著对象，即使他们属于同一类，通常在纹理、颜色、场景、背景方面都有所不同（请参阅 Fig. 1（d）中的 CoSOD3k 数据集）。

[1] * 同等贡献。本文为 CVPR20 [22] 的中文翻译版

图 2: 来自本文的 CoSOD3k 数据集的示例图像。它提供了丰富标注信息，如，图像级类别（顶部），边界框，对象级标注，实例级标注。本文的 CoSOD3k 数据集将为 CoSOD 任务打下坚实的基础，并使更多相近领域受益，例如，协同分割，基于弱监督的定位。有关详细信息，请参阅补充材料。缩放文本可获得最佳阅读体验。

为了深入研究 CoSOD，本文做了三个贡献：

- 首先，我们构建了一个极具挑战性的 CoSOD3k 数据集，它的数据更真实。我们的 CoSOD3k 数据集是迄今为止最大规模的 CoSOD 数据集。如 Fig. 2所示，它有两方面特征：1）包含了 13 个超类，共计 160 组，3,316 张图像，其中每个超类都经过精心选择以覆盖各种场景；2）每张图像都伴随着**类别**，**边界框**，**对象级**，和**实例级**的标注，使各种视觉任务受益。
- 其次，我们呈现了第一个大规模的协同显著性物体检测调研，回顾了 34 个最新模型（SOTA），并在四个现有 CoSOD 数据集 [4,52,82,94] 以及本文的 *CoSOD3k* 上评估了其中的 19 个模型。我们提供了一个基准工具包来集成各种公开可用的 CoSOD 数据集和多个 CoSOD 度量标准，以实现便捷的性能评估。
- 最后，基于我们的综合评估结果，我们观察到了一些有趣的发现并讨论了未来研究中的一些重要问题。本文的研究可以作为推动大规模模型开发和对比的动力。

2. 相关工作

数据集： 目前为止，只有少数 CoSOD 数据集被提出 [4,11,52,82,90,93]，如 Tab. 1所示。MSRC [82] 和 Image Pair [52] 是最早的两个数据集。MSRC 设计的初衷是图像中的对象识别任务，并且在过去几年激发了许多有趣的想法。该数据集包括 8 个图像组，总共 240 张图片，并带有手动标注的像素级真值（Ground-truth）。Image Pair（由 Li 等人引入）[52]，是专为图像对设计，包含 105 组总共 210 张图像。*iCoSeg* [4] 数据集于 2010 年公开。这是一个相对较大的数据集，包含了 38 个类别，共 643 张图像。数据集中每个组包含 4 到 42 张图像，而不像 Image Pair 数据集中，每组只有的 2 张图像。*THUR15K* [11] 和 *CoSal2015* [94] 是两个大规模的公开可用数据集，并且 CoSal2015 被广泛用于评估 CoSOD 算法。与上述提及的数据集不同，WICOS [90] 数据集旨在检测单张图像中的协同显著物体，其中每个图像可视为一组。

Dataset	Year	#Gp	#Img	#Avg	IL	Ceg	BBx	HQ	Input
MSRC [82]	2005	8	240	30					Group images
iCoSeg [4]	2010	38	643	17				✓	Group images
Image Pair [52]	2011	105	210	2					Two images
THUR15K [11]	2014	5	15k	3k					Group images
CoSal2015 [94]	2015	50	2,015	40				✓	Group images
WICOS [90]	2018	364	364	1				✓	Single image
CoSOD3k(Ours)	2020	160	3,316	21	✓	✓	✓	✓	Group images

表 1: 现有 CoSOD 以及本文的 CoSOD3k 数据集的统计信息表明，CoSOD3k 提供了更高质量和更丰富的标注。**#Gp**: 图像组数。**#Img**: 图片数量。**#Avg**: 每组平均图像数。**HQ**: 高质量的标注。**IL**: 是否提供了实例级别标注。**Ceg**: 是否为每个组提供类别标签。**BBx**: 是否为每个图像提供了包围盒标签。

尽管上述数据集已将 CoSOD 进行了不同程度的改进，但它们的种类仍然非常受限，只有几十个组。在如此小规模的数据集上，是无法完全评估模型的扩展性的。而且，这些数据集仅提供对象级标签。它们都没有提供丰富的标注，比如类别，边界

#	Model	Pub.	Year	#Training	Training Set	Main Component	SL.	Sp.	Po.	Ed.	Post.
1	WPL [35]	UIST	2010			Morphological, Translational Alignment	U				
2	PCSD [10]	ICIP	2010	120,000	8*8 image patch	sparse feature [31], Filter Bank	W				
3	IPCS [52]	TIP	2011			Ncut, co-multilayer Graph	U	✓			
4	CBCS [25]	TIP	2013			Contrast/Spatial/Corresponding Cue	U				
5	MI [51]	TMM	2013			Feature/Images Pyramid, Multi-scale Voting	U	✓			GCut
6	CSHS [60]	SPL	2013			Hierarchical Segmentation, Contour 3 Map [3]	U			✓	
7	ESMG [55]	SPL	2014			Efficient Manifold Ranking [85], OTSU [65]	U				
8	BR [7]	MM	2014			Common/Center Cue, Global Correspondence	U	✓			
9	SACS [8]	TIP	2014			Self-adaptive Weight, Low Rank Matrix	U	✓			
10	DIM‡ [93]	TNNLS	2015	1,000 + 9,963	ASD [1] + PV	SDAE model [93], Contrast/Object Prior	S	✓			
11	CODW‡ [95]	IJCV	2016		ImageNet [16] pre-train	SermaNet [68], RBM [5], IMC, IGS, IGC	W	✓	✓		
12	SP-MIL‡ [97]	TPAMI	2017	(240+643)*10%	MSRC-V1 [82] + iCoseg [4]	SPL [98], SVM, GIST [70], CNNs [9]	W	✓			
13	GD‡ [80]	IJCAI	2017	9,213	MSCOCO [56]	VGGNet16 [69], Group-wise Feature	S				
14	MVSRCC‡ [88]	TIP	2017			LBP, SIFT [62], CH, Bipartite Graph		✓	✓		
15	UMLF [28]	TCSVT	2017	(240 + 2015)*50%	MSRC-V1 [82] + CoSal2015 [95]	SVM, GMR [87], metric learning	S	✓			
16	DML‡ [54]	BMVC	2018	10,000 + 6,232 + 5,168	M10K [?] + THUR-15K [11] + DO	CAE, HSR, Multistage	S				
17	DWSI [90]	AAAI	2018			EdgeBox [107], Low-rank Matrix, CH	S		✓		
18	GONet‡ [34]	ECCV	2018		ImageNet [16] pre-train	ResNet-50 [29], Graphical Optimization	W	✓			CRF
19	COC‡ [32]	IJCAI	2018		ImageNet [16] pre-train	ResNet-50 [29], Co-attention Loss	W		✓		CRF
20	FASS‡ [106]	MM	2018		ImageNet [16] pre-train	DHS [57]/VGGNet, Graph optimization	W	✓			
21	PJO [74]	TIP	2018			Energy Minimization, BoWs	U	✓			
22	SPIG‡ [36]	TIP	2018	10,000+210 +2015+240	M10K [?]+IPCS [52] + CoSal2015 [95] + MSRC-V1 [82]	DeepLab, Graph Representation	S	✓			
23	QGF [37]	TMM	2018		ImageNet [16] pre-train	Dense Correspondence, Quality Measure	S	✓			THR
24	EHL‡ [71]	NC	2019	643	iCoseg [4]	GoogLeNet [73], FSM	S	✓			
25	IML‡ [66]	NC	2019	3624	CoSal2015 [95] + PV + CR	VGGNet16 [69]	S	✓			
26	DGFC‡ [81]	TIP	2019	>200,000	MSCOCO [56]	VGGNet16 [69], Group-wise Feature	S	✓			
27	RCANet‡ [45]	IJCAI	2019	>200,000	MSCOCO [56] + COS + iCoseg [4] + CoSal2015 [95] + MSRC [82]	VGGNet16 [69], Recurrent Units	S				THR
28	GS‡ [75]	AAAI	2019	200,000	COCO-SEG [75]	VGGNet19 [69], Co-category Classification	S				
29	MGCNet‡ [38]	ICME	2019			Graph Convolutional Networks [43]	S	✓			
30	MGLCN‡ [39]	MM	2019	N/A	N/A	VGGNet16, PiCANet [58], Inter-/Intra-graph	S	✓			
31	HC‡ [46]	MM	2019	N/A	N/A	VAE-Net [42], Hierarchical Consistency	S	✓	✓		CRF
32	CSMG‡ [100]	CVPR	2019	25,00	MB [59]	VGGNet16 [69], Shared Superpixel Feature	S	✓			
33	DeepCO^{3}‡ [33]	CVPR	2019	10,000	M10K [?]	SVFSal [96] / VGGNet [69], Co-peak Search	W		✓		
34	GWD‡ [44]	ICCV	2019	>200,000	MSCOCO [56]	VGGNet19 [69], RNN, Group-wise Loss	S				THR

表 2: 34 种经典和前沿的 CoSOD 方法的总结。**训练集:** PV = PASCAL VOC07 [17]. CR = Coseg-Rep [15]. DO = DUT-OMRON [87]. COS = COCO-subset. **主要成分:** IMC = 图像内对比度。IGS: 组内可分离性。IGC: 组内一致性。SPL: 自主学习。CH: 颜色直方图。GMR: 基于图的流形排序。CAE: 卷积自动编码器。HSR: 高空间分辨率。FSM: CBCS [25], RC [?], DCL [50], RFCN [77], DWSI [90] 五个显著性模型。**SL.** = 监督级别。W = 弱监督。S = 有监督。U = 无监督。**Sp.:** 是否使用超像素技术。**Po.:** 是否使用了弹出框算法。**Ed.:** 是否显式地使用边缘特征。**Post.:** 是否引入了后处理方法，例如 CRF，图割（GCut）或自适应/固定阈值（THR）。有关这些模型的更多详细信息，请参见两份调研报告 [14, 92]。

框，实例等，这些标注对于完成许多视觉任务和多任务建模是至关重要的。

传统方法: 早期的研究 [7, 28, 52, 74] 发现图像间的对应关系可以通过将输入图像分割成许多计算单位（例如，超像素区域 [103] 或像素簇 [25]）来建模。最近的综述也支持类似的观察结果 [14, 92]。这些方法从图像中提取启发式特征（例如轮廓 [60]，颜色，亮度），并捕获高级特征用不同方式，例如通过度量学习 [28] 或自适应加权 [8] 来表达语义属性。一些研究还调查了如何通过各种计算机制来捕获图像间的约束，例如平移对齐 [35]，高效流形排名 [55] 和全局对应 [7]。一些方法（例如 PCSD [10] 仅使用滤波器组技术）甚至不需要执行两个输入图像之间的对应匹配就能够在集中注意力发生之前实现协同物体检测。

深度学习方法: 深度 CoSOD 模型通常通过学习协同显著物体的表达来获得良好的性能。更具体地说，Zhang 等人 [93] 引入域适应模型来迁移 CoSOD 的先验知识。Wei 等人 [80] 在协作学习框架中，使用一组输入和输出来挖掘分组和单图像特征表示之间的协作和交互关系。沿着另一思路，MVSRCC [88] 模型采用了经典的特征，例如 SIFT，LBP 和颜色直方图，作为多视角特征。此外，其他几种基于更强大 CNN 模型的方法 [32, 33, 36, 71, 75, 81, 100]（例如 ResNet [29]，Res2Net [27]，GoogLeNet [73]，VGGNet [69]），获得了最佳性能。这些深度模型通常通过以下方式获得更好的性能，要么是弱监督的（例如 CODW [95]，SP-MIL [97]，GONet [34]，FASS [106]）或全监督学习的（例如，DIM [93]，GD [80]，DML [54]）。Tab. 2中列出了基于传统和深度学习的模型概要。

3. 提出的 *CoSOD3k* 数据集

3.1. 图像采集

我们建立了高质量的 CoSOD3k 数据集，其图像来源于大规模物体识别 ILSVRC [67] 数据集。使

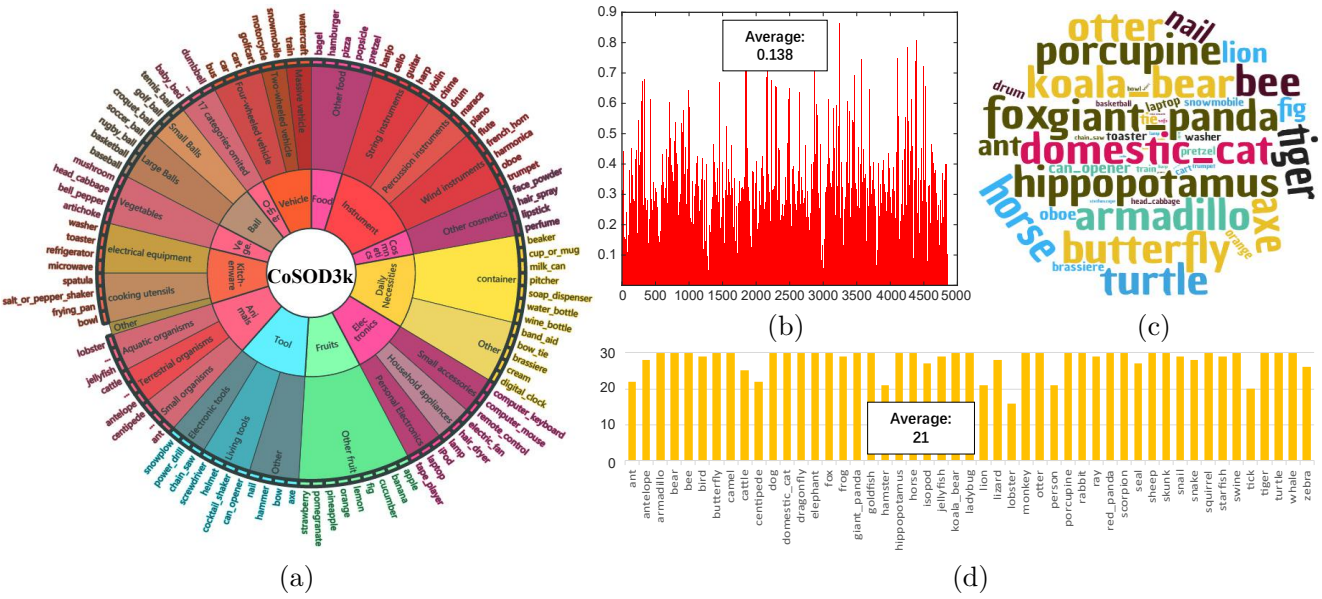

图 3: 提出的 CoSOD3k 数据集的统计数据。(a) 我们数据集的分类结构。(b) 实例大小的分布。(c) CoSOD3k 数据集的词云。(d) 49 个动物类别的图像数量。**最好在屏幕上查看，并放大以获取详细信息**。

用 ILSVRC 来构建我们的数据集有以下几点好处。ILSVRC 从 *Flickr* 中使用场景标签搜索得到的，因此它包括各种物体类别，各种真实的场景和不同的物体外观，并且涵盖了 CoSOD 中大多数具有挑战的情况，这为我们构建一个代表性的 CoSOD 基准数据集提供了坚实的基础。更重要的是，每个目标物体类别附带的轴对齐边框让我们能明确的标注实例级标签。

3.2. 数据标注

与 [21, 64] 相似，数据标注以分层（从粗到细）的方式进行（见 Fig. 2）。

分类标记. 我们为 *CoSOD3k* dataset 数据集建立了一个分级（三级）分类系统。选择 160 个常见类别以生成子类别（例如，蚂蚁，无花果，小提琴，火车等），它们与 ILSVRC 中的原始类别是一样的。然后，为每个子类分配一个上层类（中间层）。最后，我们将上层类集成到 13 个超类中。在 Fig. 3 (a) 中，给出了我们的 CoSOD3k 的分类结构。

边界标记. 二级标注是边界框，它广泛用于对象检测和定位。尽管 ILSVRC 数据集提供了边界框标注，但标记的对象不一定显著。借鉴许多著名的 SOD 数据集的做法 [1, 2, 12, 40, 48, 49, 59, 63, 76, 84, 86]，我们邀请了三位受试者在每幅图像中的那些吸引他们注意力的物体周围重新框出边界框。然后，我们合并由三个观察者标记的边界框，并让 CoSOD 领域中的另外两名高级研究人员再次检查标注。之后，和 [41] 的做法一致，我们丢弃了那些含有六个以上物体以及仅包含背景的图像。最后，我们在 160 个类别中收集了 3,316 张图像。

对象/实例级标注. 高质量像素级标注对于 Co-SOD 数据集是很有必要的。为此我们雇佣了 20 名专业标注者，并用 100 张图像示例训练他们。然后，他们被指示根据已经标注好的边框进一步用物体和实例级标签对图像进行标注。对于物体级别和实例级别的标签，每张图像的平均标注时间分别约为 8 分钟和 15 分钟。此外，我们还有三名志愿者对整个过程进行了三次以上的交叉检查，以确保获得高质量的标注。通过这种方式，我们获得了准确而具有挑战性的数据集，其中包含总共 3,316 个对象级和 4,915 个实例级显著物体标注。注意，我们最终边界框标签会根据像素级标注来重新修正从而紧紧贴近目标。

3.3. 数据集特征与统计

为了更深入地了解我们的 CoSOD3k，我们在下面展示其几个重要特征。

Metric	PCSD [10]	CODR [89]	ESMG [55]	CBCS [25]	IPCS [52]	SACS [8]	UMLF [28]	CSHS [60]	HCNco [61]	DIM [93]‡	EGNet [105]‡	CPD [83]‡	CSMG [100]‡
$S_\alpha \uparrow$	.401	.656	.664	.685	.747	.775	.810	.810	.838	.729	.842	.879	.902
$F_\beta \uparrow$	.378	.652	.651	.800	.786	.837	.870	.856	.867	.867	.835	.880	.925
$E_\xi \uparrow$	.598	.762	.767	.856	.848	.887	.898	.899	.896	.905	.887	.917	.952
$M \downarrow$	.242	.226	.198	.152	.168	.169	.163	.148	.073	.256	.076	.054	.067

表 3: Image Pair [52] 数据集上 13 种 CoSOD 方法的基准测试结果。简便起见，我们使用 ↑ 和 ↓ 表示越大越好和越小越好。前三名以红色，绿色和蓝色来突出显示。

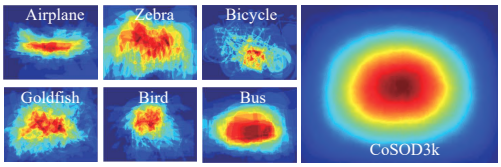

图 4: *CoSOD3k* 混合特定类别和整体类别掩模的重叠掩模可视化。

混合特定类别标注. Fig. 4 显示了单个类别和整个类别的平均真值标注，可以观察到一些类别具有独特形状（例如飞机，斑马和自行车），这呈现出当前的形状偏差图，而具有非刚性或凸形状的类别（例如，金鱼，鸟和公共汽车）可能没有明确的形状偏差。整个类别标注（Fig. 4的左边）趋向于呈现出中心偏差而没有形状偏差，这和显著物体的特点相符。众所周知，在拍摄照片时，人们通常倾向于更加关注场景的中心。因此，在算法中采用高斯函数时，SOD 模型很容易获得高分。由于篇幅所限，我们会在补充材料上展示了所有 160 种特定混合物类别的掩模。

足够的物体多样性. 由 Tab. 6的第二行和 Fig. 3 (c) 所示，我们的 CoSOD3k 涵盖了许多超级类，包括蔬菜，食品，水果，工具，必需品，交通，化妆品，球，仪器，厨具，动物（Fig. 3 d），以及其它的物品，从而能够更全面的了解现实世界场景。

实例大小. 实例大小定义为前景实例像素与总图像像素之比。表 4 总结了我们的 CoSOD3k 中的实例大小。实例大小的分布（Fig. 3 b）为 0.02%∼86.5%（平均：13.8%），呈现出很广的分布。

实例数量. 将物体分为实例对人类理解、分类和与外界交互至关重要。为了使基于学习的模型能够获得实例级别的信息，那么实例级别的标注就不可或缺。考虑到这一点，与现有的 CoSOD 数据集相比，我们的 CoSOD3k 包含具有实例级别标注的多实例场景。如 Tab. 4中所示，实例数 (1, 2, ≥3) 的比例为 7：2：1。

CoSOD3k	Instance Size.			# Instances		
	large (>30%)	middle	small (<5%)	1	2	≥ 3
# Images	439	3173	1303	2371	644	334

表 4: 在提出的 CoSOD 的数据集中统计实例大小和数量。

4. 基准实验

4.1. 实验设置

评估指标. 为了提供全面的评估，两个广泛使用的指标：最大 F 度量 (F_β) [1]，MAE (M) [13], 和两个最近提出的指标：S 度量 (S_α) [19]，最大 E 度量 (E_ξ) [20] 被用来评估多幅图像中的 CoSOD 性能。令 $D = \{G_1, \ldots, G_i, \ldots, G_q\}$ 表示具有 q 个图像组的完整数据集，而 I_k^i 是图像组 $G_i = \{I_1^i, \ldots, I_k^i, \ldots, I_{N_i}^i\}$ 中的第 k 张图像。N_i 是 G_i 中的图像数。N_D 是整个数据集 D 中的图像总数。对于每个指标 $\vartheta \in \{S_\alpha, E_\xi, F_\beta, M\}$，我们在整个数据集上计算其平均得分 (Tab. 5和 Tab. 3)。数据集 D 的平均指标定义为 $Q_\vartheta(D) = \frac{1}{N_D} \sum_{i=1}^{q} \sum_{k=1}^{N_i} \vartheta(I_k^i)$。为了深入了解在小组层面的算法性能，我们提供了小组平均分，即 $T_\vartheta(G_i) = \frac{1}{N_i} \sum_{k=1}^{N_i} \vartheta(I_k^i)$。

评估方法. 在这项研究中，我们评估/比较了 19 种 SOTA CoSOD 模型，其中包括 10 种传统方法 [8, 10, 25, 28, 52, 53, 55, 60, 61, 89] 和 9 种深度学习模型 [34, 83, 94, 97, 98, 100, 105]。这些方法是根据两个标准选择的：（1）有代表性，和（2）代码开源。

评测协议. 我们在四个现有的 CoSOD 数据集，即 Image Pair [52]，MSRC [82]，iCoSeg [4]，CoSal [94] 和我们的 *CoSOD3k* 上进行评估。总共有 363 个组，包含约 61K 张图像，这是最大、最全面的基准评测。公平起见，我们直接使用默认设置来运行代码（例如 PCSD [10], IPCS [52], CSHS [60], CBCS [25], RFPR [53], ESMG [55], SACS [8], CODR [89], HCNco [61], UMLF [28], CPD [83], EGNet [105]）或使用作者提

	Metric	CBCS [25]	ESMG [55]	RFPR [53]	CSHS [60]	SACS [8]	CODR [89]	UMLF [28]	DIM [93]‡	CODW [95]‡	MIL [98]‡	IML [66]‡	GONet [34]‡	SP-MIL [97]‡	CSMG [100]‡	CPD [83]‡	EGNet [105]‡
MSRC	$S_\alpha \uparrow$	.480	.532	.644	.666	.707	.754	.797	.657	.713	.720	.781	.795	.769	.722	.714	.702
	$F_\beta \uparrow$	.630	.606	.696	.727	.782	.776	.849	.705	.784	.768	.840	.846	.824	.847	.762	.752
	$E_\xi \uparrow$	.676	.675	.746	.784	.810	.822	.880	.725	.820	.800	.856	.863	.855	.859	.795	.794
	$M \downarrow$	.314	.303	.302	.289	.224	.198	.184	.309	.264	.216	.174	.179	.218	.190	.173	.186
CoSal2015	$S_\alpha \uparrow$	.544	.552	N/A	.592	.694	.689	.662	.592	.648	.673	-	.751	N/A	.774	.814	.818
	$F_\beta \uparrow$	.532	.476	N/A	.564	.650	.634	.690	.580	.667	.620	-	.740	N/A	.784	.782	.786
	$E_\xi \uparrow$	.656	.640	N/A	.685	.749	.749	.769	.695	.752	.720	-	.805	N/A	.842	.841	.843
	$M \downarrow$	.233	.247	N/A	.313	.194	.204	.271	.312	.274	.210	-	.160	N/A	.130	.098	.099
iCoSeg	$S_\alpha \uparrow$	.658	.728	.744	.750	.752	.815	.703	.758	.750	.727	.832	.820	.771	.821	.861	.875
	$F_\beta \uparrow$	.705	.685	.771	.765	.770	.823	.761	.797	.782	.741	.846	.832	.794	.850	.855	.875
	$E_\xi \uparrow$	.797	.784	.841	.841	.817	.889	.827	.864	.832	.799	.895	.864	.843	.889	.900	.911
	$M \downarrow$	.172	.157	.170	.179	.154	.114	.226	.179	.184	.186	.104	.122	.174	.106	.057	.060

表 5: 16 种领先的 CoSOD 方法在现有三个经典 [4,82,94] 数据集上的基准测试结果。“N / A” 表示代码或结果无法获得。“–” 表示整个数据集都被用作训练集。请注意，UMLF 方法采用来自 MSRC 和 CoSal2015 的一半图像来训练它们的模型。“score” 表示特定的模型在当前评估的数据集上已训练过（例如 SP-MIL，UMLF）。参考 Tab. 2了解更多训练详细信息（某些模型用了更多训练数据）。

供的 CoSOD 结果图（例如 IML [66]，CODW [95]，GONet [34]，SP-MIL [97]，CSMG [100]）。*iCoSeg* [4]，*CoSal2015* [94])。

4.2. 定量比较

Image Pair 的性能. Image Pair 是第一个 CoSOD 数据集 [52]，如 Tab. 3所示。Image Pair [52] 数据集在每个组中只有一对图像，并且大多数协同显著对象具有相似的外观。因此，与其他协同显著物体检测数据集相比，它相对容易，并且 top-1 模型（即 CSMG [100]）获得了很高的性能（S_α >0.9）。

MRSC 的性能. MSRC 数据集 [82] 在每个组中都有更多图像。从 Tab. 5实验表明 UMLF [28]，GONet [34]，IML [66] 和 SP-MIL [97] 是该数据集中前 4 名的模型。有趣的是，我们发现所有这些模型都采用超像素方法在多张图像上推理共现区域。这些工作在包含大量外观相似的显著物体的 MSRC 数据集上取得了良好的性能。然而，由于超像素技术专注于颜色相似性，因此它们在 iCoSeg 和我们的 CoSOD3k 上的性能急剧下降（例如，GONet: No. 2 → No.5），因此，这些方法在应对基于语义相似性的数据集不够鲁棒。

iCoSeg 的性能. iCoSeg 数据集 [4] 最初是为图像协同分割而设计的，但是现在已广泛用于 CoSOD 任务了。如 Tab. 5所示，两个 SOD 模型（EGNet [105] 和 CPD [83]）都达到了最先进的性能。一个可能的原因是 iCoSeg 数据集包含大量当个物体的图像，这些图像很容易被 SOD 模型检测到。这部分地说明 iCoSeg 数据集可能不适合协同显著性物体检测模型的评估。

CoSal2015 的性能. Tab. 5 展示了 CoSal2015 数据集上的评测结果。我们发现了一个有趣的现象，排名前 2 的模型仍是 EGNet [105] 和 CPD [83]，与 iCoSeg 数据集上的模型排名一致。这意味着某些性能最高的显著物体检测框架可能适合扩展到 CoSOD 任务。

CoSOD3k 的性能. Tab. 6中呈现了我们 CoSOD3k 的结果。为了提供对每个组的更深入的了解，我们在 13 个超类上报告了模型的性能。我们可以在其他分类上观察到较低的平均值，其他类（如，婴儿床，铅笔盒），乐器（如，钢琴，吉他，大提琴等），必需品（如，水罐），工具（如斧头，钉子，链锯），以及球类（如足球，网球），他们在现实情景中包含复杂的结构。每行的第一名性能（S_α =0.76）清楚地表明，提出的 CoSOD3k 数据集极具挑战性，并为进一步研究留有广阔的空间。注意，几乎所有基于深度学习的模型（例如，EGNet [105]，CPD [83]，IML [66]，CSMG [100] 等）都比传统方法（CODR [89]，CSHS [60]，CBCS [25] 和 ESMG [55]）表现更好，说明了利用深度学习技术解决 CoSOD 问题具有潜在优势。另一个有趣的发现是，边缘特征可以帮助结果提供良好的边界。例如，传统（CSHS [60]）和深度学习模型（如 EGNet [105]）的最佳方法都引入了边缘信息以帮助检测。

	Vege.	Food	Fruit	Tool	Nece.	Traf.	Cosm.	Ball	Inst.	Kitch.	Elec.	Anim.	Oth.	All
#Sub-class	4	5	9	11	12	10	4	7	14	9	9	49	17	160
CBCS(TIP'13) [25]	.512	.496	.602	.523	.506	.512	.505	.554	.516	.505	.511	.547	.498	.528
CSHS(SPL'13) [60]	.521	.549	.635	.556	.530	.574	.569	.525	.535	.554	.573	.592	.516	.563
ESMG(SPL'14) [55]	.488	.553	.649	.517	.458	.527	.484	.478	.545	.492	.516	.568	.486	.532
CODR(SPL'15) [89]	.632	.646	.696	.595	.586	.649	.602	.574	.576	.612	.616	.682	.573	.630
DIM‡(TNNLS'15) [93]	.593	.626	.663	.538	.534	.569	.530	.515	.540	.528	.545	.577	.517	.559
UMLF(TCSVT'17) [28]	.711	.689	.697	.534	.648	.669	.615	.567	.559	.671	.634	.667	.559	.632
IML‡(NC'19) [66]	.767	.693	.763	.671	.680	.762	.691	.664	.655	.727	.688	.791	.623	.720
CSMG‡(CVPR'19) [100]	.645	.774	.756	.612	.666	.770	.632	.714	.612	.751	.725	.780	.617	.711
Average	.643	.650	.704	.596	.608	.667	.608	.595	.577	.644	.630	.690	.570	.639

表 6: 我们的 CoSOD3k 数据集的每个超类平均表现 (S_α)。Vege. = 蔬菜, Nece. = 必需品, Traf. = 交通, Cosm.= 化妆品, Inst. = 仪器, Kitch. = 厨具, Elec. = 电子产品, Anim. = 动物, Oth. = 其他. "All" 是指整个数据集的分数。我们仅评估 10 个最新且公布代码的模型。注意 CPD 和 EGNet 是基准测试中排名前 2 的 SOD 模型 (http://dpfan.net/socbenchmark)。

4.3. 定性比较

Fig. 5展示了 CoSOD3k 上 10 个最新视觉算法的两组结果。可以看出，SOD 模型（例如 EGNet [105] 和 CPD [83]）检测到所有显著物体，但是忽略了共同的信息。例如，（检测）香蕉的结果包含其他几个不相关的物体，如橘子，菠萝和苹果。类似的情况也发生在马类别的图像中，其中围栏（第二张图像）和骑手（第一张和第四张图片）与马一起被检测到了。另一方面，CoSOD 方法（如 CSMG [100]）可以确定共同的显著物体，但不能得到准确的预测结果，尤其是在物体边界上。基于以上观察，我们推断出 CoSOD 问题尚未解决，后续模型仍有很大的空间。

5. 讨论

从评测中可以观察到在大多数情况下，目前的 SOD 方法（例如 EGNet [105] 和 CPD [83]）比 CoSOD 方法（例如 CSMG [100] 和 SP-MIL [97]）获得了非常有竞争力或更好的表现。然而，这并不意味着当前的数据集不够复杂，直接使用 SOD 方法可以获得良好性能的，SOD 方法在 CoSOD 数据集上的表现事实上是低于在 SOD 数据集上的，例如 HKU-IS [49]（EGNet 的 $F_\beta = 0.937$ ）和 ECSSD [86]（EGNet [105] 的 $F_\beta = 0.943$）。相反，这是因为 CoSOD 中的许多问题仍未得到充分研究，这使现有 CoSOD 模型的有效性降低。在本节中，我们讨论四个重要的问题，现有的显著性物体检测方法尚未完全解决这些问题，可以在今后进行研究。

可拓展性. 可拓展性问题是设计 CoSOD 算法时需要考虑的最重要问题之一。具体来说，它表示 CoSOD 模型处理大规模图像场景的能力。众所周知，CoSOD 的一个关键特性是需要考虑每组中的多个图像。然而，现实中，图像组可能包含多个相关图像。在这种情况下，不考虑方法的可拓展性问题将具有巨大的计算成本且需要运行很长时间，这在实践中是不可接受的。因此，如何解决可拓展性问题成为该领域的关键，尤其是在将 CoSOD 方法应用于真实场景。

稳定性. 另一个重要问题是稳定性问题。在处理包含多个图像的图像组时，某些现有方法（例如 HCNco [61]，PCSD [10]，IPCS [52]）将图像组分为图像对或图像子组（例如 GD [80]）。另一类方法采用基于 RNN 的模型（例如，GWD [44]），该模型需要分配输入图像的顺序。所有这些策略都会使整个过程变得不稳定，因为没有一个原则性策略来指导如何划分图像组或分配相关图像的输入顺序。这也将影响 CoSOD 方法的的应用。

兼容性. 将 SOD 引入 CoSOD 是建立 CoSOD 框架直接而有效的策略。然而，大多数现有的工作仅介绍了 SOD 模型的结果或特征作为有用的信息线索。利用 SOD 技术的另一步骤是将基于 CNN 的 SOD 网络与 CoSOD 模型结合起来为 CoSOD 构建统一的、端到端的可训练框架。为了实现这一目标，需要考虑 CoSOD 框架的兼容性，从而方便地集成现有的 SOD 技术。

指标. CoSOD 的当前评估指标是根据 SOD 设计的，即直接计算每个组的 SOD 评分的平均值。与 SOD 相比，CoSOD 涉及不同图像之间的协同物体的关系信息，这对于 CoSOD 评估更为重要，并且带来了更多挑战。例如，当前 CoSOD 指标假设目标物体在所有图像中具有相似的大小。作为不同图像

图 5: CoSOD3k 上现有的前十大模型的定性示例。补充材料中显示了更多示例。

中具有不同大小的物体，CoSOD 指标（第4部分中提到的 $S_\alpha, E_\xi, F_\beta, M$ ）倾向于较大物体。此外，目前 CoSOD 指标偏向于单一图像中目标检测的性能，而不是多幅图像中相应物体的识别。因此，如何为 CoSOD 设计合适的指标是一个开放性的问题。

6. 结论

在本文中，我们完整地调研了协同显著性物体检测（CoSOD）任务。通过识别严重的数据偏差，即假设每一组图像都包含视觉外观相似的显著物体，我们在现有的 CoSOD 数据集中构建了一个新的高质量数据集，名为 CoSOD3k，它包含了语义或概念层次上具有相似性的协同显著物体。值得注意的是，CoSOD3k 是最具挑战性的 CoSOD 数据集，到目前为止，它包含 160 个组和总共 3316 张图像，这些图像分别带有类别，边界框，对象级和实例级标注。它在多样性，难度和可扩展性方面取得了重大飞跃，有利于相关视觉任务，例如协同分割，弱监督定位和实例级检测，并对这些研究领域的未来发展有很大的帮助。

此外，本文还对 34 种前沿算法进行了全面的研究，并对其中 19 种算法在现有的 4 个数据集和提出的 *CoSOD3k* 数据集上进行了基准测试。根据评估结果，我们对 CoSOD 研究领域的核心问题进行了深入讨论。我们希望这项工作中提出的研究能够更好地促进 CoSOD 社区的发展。未来，我们计划增加数据集的规模来启发更多新的思路。

参考文献

[1] Radhakrishna Achanta, Sheila Hemami, Francisco Estrada, and Sabine Süsstrunk. Frequency-tuned salient region detection. In *IEEE CVPR*, pages 1597–1604, 2009.

[2] Sharon Alpert, Meirav Galun, Ronen Basri, and Achi Brandt. Image segmentation by probabilistic bottom-up aggregation and cue integration. In *IEEE CVPR*, 2007.

[3] Pablo Arbelaez, Michael Maire, Charless Fowlkes, and Jitendra Malik. Contour detection and hierarchical image segmentation. *IEEE TPAMI*, 33(5):898–916, 2010.

[4] Dhruv Batra, Adarsh Kowdle, Devi Parikh, Jiebo Luo, and Tsuhan Chen. icoseg: Interactive co-segmentation with intelligent scribble guidance. In *IEEE CVPR*, 2010.

[5] Yoshua Bengio et al. Learning deep architectures for ai. *FTML*, 2(1):1–127, 2009.

[6] Ali Borji, Ming-Ming Cheng, Qibin Hou, Huaizu Jiang, and Jia Li. Salient object detection: A survey. *Computational Visual Media*, 5(2):117–150, 2019.

[7] Xiaochun Cao, Yupeng Cheng, Zhiqiang Tao, and Huazhu Fu. Co-saliency detection via base reconstruction. In *ACM MM*, pages 997–1000, 2014.

[8] Xiaochun Cao, Zhiqiang Tao, Bao Zhang, Huazhu Fu, and Wei Feng. Self-adaptively weighted co-saliency detection via rank constraint. *IEEE TIP*, 23(9):4175–4186, 2014.

[9] Ken Chatfield, Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. Return of the devil in the details: Delving deep into convolutional nets. In *BMVC*, 2014.

[10] Hwann-Tzong Chen. Preattentive co-saliency detection. In *IEEE ICIP*, pages 1117–1120, 2010.

[11] Ming-Ming Cheng, Niloy J Mitra, Xiaolei Huang, and Shi-Min Hu. Salientshape: group saliency in image collections. *The Visual Computer*, 30(4):443–453, 2014.

[12] Ming-Ming Cheng, Niloy J. Mitra, Xiaolei Huang, Philip H. S. Torr, and Shi-Min Hu. Global contrast based salient region detection. *IEEE TPAMI*, 37(3):569–582, 2015.

[13] Ming-Ming Cheng, Jonathan Warrell, Wen-Yan Lin, Shuai Zheng, Vibhav Vineet, and Nigel Crook. Efficient salient region detection with soft image abstraction. In *IEEE ICCV*, pages 1529–1536, 2013.

[14] Runmin Cong, Jianjun Lei, Huazhu Fu, Ming-Ming Cheng, Weisi Lin, and Qingming Huang. Review of visual saliency detection with comprehensive information. *IEEE TCSVT*, 29(10):2941–2959, 2018.

[15] Jifeng Dai, Ying Nian Wu, Jie Zhou, and Song-Chun Zhu. Cosegmentation and cosketch by unsupervised learning. In *IEEE ICCV*, pages 1305–1312, 2013.

[16] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *IEEE CVPR*, pages 248–255, 2009.

[17] Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes (voc) challenge. *IJCV*, 2010.

[18] Deng-Ping Fan, Ming-Ming Cheng, Jiang-Jiang Liu, Shang-Hua Gao, Qibin Hou, and Ali Borji. Salient objects in clutter: Bringing salient object detection to the foreground. In *ECCV*, pages 186–202. Springer, 2018.

[19] Deng-Ping Fan, Ming-Ming Cheng, Yun Liu, Tao Li, and Ali Borji. Structure-measure: A New Way to Evaluate Foreground Maps. In *IEEE ICCV*, pages 4548–4557, 2017.

[20] Deng-Ping Fan, Cheng Gong, Yang Cao, Bo Ren, Ming-Ming Cheng, and Ali Borji. Enhanced-alignment Measure for Binary Foreground Map Evaluation. In *IJCAI*, 2018.

[21] Deng-Ping Fan, Ge-Peng Ji, Guolei Sun, Ming-Ming Cheng, Jianbing Shen, and Ling Shao. Camouflaged object detection. In *IEEE CVPR*, 2020.

[22] Deng-Ping Fan, Zheng Lin, Ge-Peng Ji, Dingwen Zhang, Huazhu Fu, and Ming-Ming Cheng. Taking a deeper look at co-salient object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

[23] Deng-Ping Fan, Zheng Lin, Zhao Zhang, Menglong Zhu, and Ming-Ming Cheng. Rethinking RGB-D Salient Object Detection: Models, Datasets, and Large-Scale Benchmarks. *IEEE TNNLS*, 2020.

[24] Deng-Ping Fan, Wenguan Wang, Ming-Ming Cheng, and Jianbing Shen. Shifting more attention to video

salient object detection. In *IEEE CVPR*, pages 8554–8564, 2019.

[25] Huazhu Fu, Xiaochun Cao, and Zhuowen Tu. Cluster-based co-saliency detection. *IEEE TIP*, 22(10):3766–3778, 2013.

[26] Keren Fu, Deng-Ping Fan, Ge-Peng Ji, and Qijun Zhao. JL-DCF: Joint Learning and Densely-Cooperative Fusion Framework for RGB-D Salient Object Detection. In *IEEE CVPR*, 2020.

[27] Shang-Hua Gao, Ming-Ming Cheng, Kai Zhao, Xin-Yu Zhang, Ming-Hsuan Yang, and Philip Torr. Res2Net: A New Multi-scale Backbone Architecture. *IEEE TPAMI*, 2020.

[28] Junwei Han, Gong Cheng, Zhenpeng Li, and Dingwen Zhang. A unified metric learning-based framework for co-saliency detection. *IEEE TCSVT*, 28(10):2473–2483, 2017.

[29] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *IEEE CVPR*, pages 770–778, 2016.

[30] Xiaodi Hou and Liqing Zhang. Saliency detection: A spectral residual approach. In *IEEE CVPR*, pages 1–8, 2007.

[31] Xiaodi Hou and Liqing Zhang. Dynamic visual attention: Searching for coding length increments. In *NIPS*, 2009.

[32] Kuang-Jui Hsu, Yen-Yu Lin, and Yung-Yu Chuang. Co-attention cnns for unsupervised object co-segmentation. In *IJCAI*, pages 748–756, 2018.

[33] Kuang-Jui Hsu, Yen-Yu Lin, and Yung-Yu Chuang. DeepCO3: Deep Instance Co-Segmentation by Co-Peak Search and Co-Saliency Detection. In *IEEE CVPR*, 2019.

[34] Kuang-Jui Hsu, Chung-Chi Tsai, Yen-Yu Lin, Xiaoning Qian, and Yung-Yu Chuang. Unsupervised CNN-based co-saliency detection with graphical optimization. In *ECCV*, pages 485–501. Springer, 2018.

[35] David E Jacobs, Dan B Goldman, and Eli Shechtman. Cosaliency: Where people look when comparing images. In *ACM UIST*, pages 219–228, 2010.

[36] Dong-ju Jeong, Insung Hwang, and Nam Ik Cho. Co-salient object detection based on deep saliency networks and seed propagation over an integrated graph. *IEEE TIP*, 27(12):5866–5879, 2018.

[37] Koteswar Rao Jerripothula, Jianfei Cai, and Junsong Yuan. Quality-guided fusion-based co-saliency estimation for image co-segmentation and colocalization. *IEEE TMM*, 20(9):2466–2477, 2018.

[38] Bo Jiang, Xingyue Jiang, Jin Tang, Bin Luo, and Shilei Huang. Multiple graph convolutional networks for co-saliency detection. In *IEEE ICME*, pages 332–337, 2019.

[39] Bo Jiang, Xingyue Jiang, Ajian Zhou, Jin Tang, and Bin Luo. A unified multiple graph learning and convolutional network model for co-saliency estimation. In *ACM MM*, pages 1375–1382, 2019.

[40] Huaizu Jiang, Ming-Ming Cheng, Shi-Jie Li, Ali Borji, and Jingdong Wang. Joint Salient Object Detection and Existence Prediction. *Front. Comput. Sci.*, 2017.

[41] Edna L Kaufman, Miles W Lord, Thomas Whelan Reese, and John Volkmann. The discrimination of visual number. *The American Journal of Psychology*, 62(4):498–525, 1949.

[42] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. In *ICLR*, 2014.

[43] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *ICLR*, 2017.

[44] Bo Li, Zhengxing Sun, Qian Li, Yunjie Wu, and Anqi Hu. Group-wise deep object co-segmentation with co-attention recurrent neural network. In *IEEE ICCV*, 2019.

[45] Bo Li, Zhengxing Sun, Lv Tang, Yunhan Sun, and Jinlong Shi. Detecting robust co-saliency with recurrent co-attention neural network. In *IJCAI*, pages 818–825, 2019.

[46] Bo Li, Zhengxing Sun, Quan Wang, and Qian Li. Co-saliency detection based on hierarchical consistency. In *ACM MM*, pages 1392–1400, 2019.

[47] Chongyi Li, Runmin Cong, Junhui Hou, Sanyi Zhang, Yue Qian, and Sam Kwong. Nested network with two-stream pyramid for salient object detection in optical remote sensing images. *TGRS*, 57(11):9156–9166, 2019.

[48] Guanbin Li, Yuan Xie, Liang Lin, and Yizhou Yu. Instance-level salient object segmentation. In *IEEE CVPR*, pages 247–256, 2017.

[49] Guanbin Li and Yizhou Yu. Visual saliency based on multiscale deep features. In *IEEE CVPR*, 2015.

[50] Guanbin Li and Yizhou Yu. Deep contrast learning for salient object detection. In *IEEE CVPR*, 2016.

[51] Hongliang Li, Fanman Meng, and King Ngi Ngan. Co-salient object detection from multiple images. *IEEE TMM*, 15(8):1896–1909, 2013.

[52] Hongliang Li and King Ngi Ngan. A co-saliency model of image pairs. *IEEE TIP*, 20(12):3365–3375, 2011.

[53] Lina Li, Zhi Liu, Wenbin Zou, Xiang Zhang, and Olivier Le Meur. Co-saliency detection based on region-level fusion and pixel-level refinement. In *IEEE ICME*, 2014.

[54] Min Li, Shizhong Dong, Kun Zhang, Zhifan Gao, Xi Wu, Heye Zhang, Guang Yang, and Shuo Li. Deep learning intra-image and inter-images features for co-saliency detection. In *BMVC*, page 291, 2018.

[55] Yijun Li, Keren Fu, Zhi Liu, and Jie Yang. Efficient saliency-model-guided visual co-saliency detection. *IEEE SPL*, 22(5):588–592, 2014.

[56] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *ECCV*, pages 740–755. Springer, 2014.

[57] Nian Liu and Junwei Han. Dhsnet: Deep hierarchical saliency network for salient object detection. In *IEEE CVPR*, pages 678–686, 2016.

[58] Nian Liu, Junwei Han, and Ming-Hsuan Yang. PiCANet: Learning pixel-wise contextual attention for saliency detection. In *IEEE CVPR*, pages 3089–3098, 2018.

[59] Tie Liu, Jian Sun, Nanning Zheng, Xiaoou Tang, and Heung-Yeung Shum. Learning to detect a salient object. In *IEEE CVPR*, pages 1–8, 2007.

[60] Zhi Liu, Wenbin Zou, Lina Li, Liquan Shen, and Olivier Le Meur. Co-saliency detection based on hierarchical segmentation. *IEEE SPL*, 21(1):88–92, 2013.

[61] Jing Lou, Fenglei Xu, Qingyuan Xia, Wankou Yang, and Mingwu Ren. Hierarchical co-salient object detection via color names. In *IEEE ACPR*, pages 718–724, 2017.

[62] David G Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004.

[63] David Martin, Charless Fowlkes, Doron Tal, and Jitendra Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *IEEE ICCV*, 2001.

[64] Kaichun Mo, Shilin Zhu, Angel X Chang, Li Yi, Subarna Tripathi, Leonidas J Guibas, and Hao Su. Partnet: A large-scale benchmark for fine-grained and hierarchical part-level 3d object understanding. In *CVPR*, pages 909–918, 2019.

[65] Nobuyuki Otsu. A threshold selection method from gray-level histograms. *IEEE TSMC*, 9(1):62–66, 1979.

[66] Jingru Ren, Zhi Liu, Xiaofei Zhou, Cong Bai, and Guangling Sun. Co-saliency detection via integration of multi-layer convolutional features and inter-image propagation. *Neurocomputing*, 371:137–146, 2020.

[67] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *IJCV*, 115(3):211–252, 2015.

[68] Pierre Sermanet, David Eigen, Xiang Zhang, Michaˇel Mathieu, Rob Fergus, and Yann LeCun. Overfeat: Integrated recognition, localization and detection using convolutional networks. In *ICLR*, 2014.

[69] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In *ICLR*, 2015.

[70] Parthipan Siva, Chris Russell, Tao Xiang, and Lourdes Agapito. Looking beyond the image: Unsupervised learning for object saliency and detection. In *IEEE CVPR*, pages 3238–3245, 2013.

[71] Shaoyue Song, Hongkai Yu, Zhenjiang Miao, Dazhou Guo, Wei Ke, Cong Ma, and Song Wang. An easy-to-hard learning strategy for within-image co-saliency detection. *Neurocomputing*, 358:166–176, 2019.

[72] Jinming Su, Jia Li, Yu Zhang, Changqun Xia, and Yonghong Tian. Selectivity or invariance: Boundary-aware salient object detection. In *IEEE ICCV*, 2019.

[73] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *IEEE CVPR*, pages 1–9, 2015.

[74] Chung-Chi Tsai, Weizhi Li, Kuang-Jui Hsu, Xiaoning Qian, and Yen-Yu Lin. Image co-saliency detection and co-segmentation via progressive joint optimization. *IEEE TIP*, 28(1):56–71, 2018.

[75] Chong Wang, Zheng-Jun Zha, Dong Liu, and Hongtao Xie. Robust deep co-saliency detection with group semantic. In *AAAI*, pages 8917–8924, 2019.

[76] Lijun Wang, Huchuan Lu, Yifan Wang, Mengyang Feng, Dong Wang, Baocai Yin, and Xiang Ruan. Learning to detect salient objects with image-level supervision. In *IEEE CVPR*, pages 136–145, 2017.

[77] Linzhao Wang, Lijun Wang, Huchuan Lu, Pingping Zhang, and Xiang Ruan. Saliency detection with recurrent fully convolutional networks. In *ECCV*, pages 825–841, 2016.

[78] Wenguan Wang and Jianbing Shen. Higher-order image co-segmentation. *IEEE TMM*, 18(6):1011–1021, 2016.

[79] Xiaochuan Wang, Xiaohui Liang, Bailin Yang, and Frederick WB Li. No-reference synthetic image quality assessment with convolutional neural network and local image saliency. *Computational Visual Media*, 2019.

[80] Lina Wei, Shanshan Zhao, Omar El Farouk Bourahla, Xi Li, and Fei Wu. Group-wise deep co-saliency detection. In *IJCAI*, 2017.

[81] Lina Wei, Shanshan Zhao, Omar El Farouk Bourahla, Xi Li, Fei Wu, and Yueting Zhuang. Deep group-wise fully convolutional network for co-saliency detection with graph propagation. *IEEE TIP*, 28(10):5052–5063, 2019.

[82] John Winn, Antonio Criminisi, and Tom Minka. Object categorization by learned universal visual dictionary. In *IEEE ICCV*, pages 1800–1807, 2005.

[83] Zhe Wu, Li Su, and Qingming Huang. Cascaded partial decoder for fast and accurate salient object detection. In *IEEE CVPR*, pages 3907–3916, 2019.

[84] Changqun Xia, Jia Li, Xiaowu Chen, Anlin Zheng, and Yu Zhang. What is and what is not a salient object? learning salient object detector by ensembling linear exemplar regressors. In *IEEE CVPR*, pages 4142–4150, 2017.

[85] Bin Xu, Jiajun Bu, Chun Chen, Deng Cai, Xiaofei He, Wei Liu, and Jiebo Luo. Efficient manifold ranking for image retrieval. In *ACM SIGIR*, pages 525–534, 2011.

[86] Qiong Yan, Li Xu, Jianping Shi, and Jiaya Jia. Hierarchical saliency detection. In *IEEE CVPR*, pages 1155–1162, 2013.

[87] Chuan Yang, Lihe Zhang, Huchuan Lu, Xiang Ruan, and Ming-Hsuan Yang. Saliency detection via graph-based manifold ranking. In *IEEE CVPR*, pages 3166–3173, 2013.

[88] Xiwen Yao, Junwei Han, Dingwen Zhang, and Feiping Nie. Revisiting co-saliency detection: A novel approach based on two-stage multi-view spectral rotation co-clustering. *IEEE TIP*, 26(7):3196–3209, 2017.

[89] Linwei Ye, Zhi Liu, Junhao Li, Wan-Lei Zhao, and Liquan Shen. Co-saliency detection via co-salient object discovery and recovery. *IEEE SPL*, 22(11):2073–2077, 2015.

[90] Hongkai Yu, Kang Zheng, Jianwu Fang, Hao Guo, Wei Feng, and Song Wang. Co-saliency detection within a single image. In *AAAI*, 2018.

[91] Yi Zeng, Pingping Zhang, Jianming Zhang, Zhe Lin, and Huchuan Lu. Towards high-resolution salient object detection. In *IEEE ICCV*, pages 7234–7243, 2019.

[92] Dingwen Zhang, Huazhu Fu, Junwei Han, Ali Borji, and Xuelong Li. A review of co-saliency detection algorithms: Fundamentals, applications, and challenges. *ACM TIST*, 9(4):1–31, 2018.

[93] Dingwen Zhang, Junwei Han, Jungong Han, and Ling Shao. Cosaliency detection based on intrasaliency prior transfer and deep intersaliency mining. *IEEE TNNLS*, 27(6):1163–1176, 2015.

[94] Dingwen Zhang, Junwei Han, Chao Li, and Jingdong Wang. Co-saliency detection via looking deep and wide. In *IEEE CVPR*, pages 2994–3002, 2015.

[95] Dingwen Zhang, Junwei Han, Chao Li, Jingdong Wang, and Xuelong Li. Detection of co-salient objects by looking deep and wide. *IJCV*, 120(2):215–232, 2016.

[96] Dingwen Zhang, Junwei Han, and Yu Zhang. Supervision by fusion: Towards unsupervised learning of deep salient object detector. In *IEEE ICCV*, pages 4048–4056, 2017.

[97] Dingwen Zhang, Deyu Meng, and Junwei Han. Co-saliency detection via a self-paced multiple-instance

learning framework. *IEEE TPAMI*, 39(5):865–878, 2016.

[98] Dingwen Zhang, Deyu Meng, Chao Li, Lu Jiang, Qian Zhao, and Junwei Han. A self-paced multiple-instance learning framework for co-saliency detection. In *IEEE ICCV*, pages 594–602, 2015.

[99] Jing Zhang, Deng-Ping Fan, Yuchao Dai, Saeed Anwar, Fatemeh Sadat Saleh, Tong Zhang, and Nick Barnes. UC-Net: Uncertainty Inspired RGB-D Saliency Detection via Conditional Variational Autoencoders. In *IEEE CVPR*, 2020.

[100] Kaihua Zhang, Tengpeng Li, Bo Liu, and Qingshan Liu. Co-saliency detection via mask-guided fully convolutional networks with multi-scale label smoothing. In *CVPR*, pages 3095–3104, 2019.

[101] Lu Zhang, Jianming Zhang, Zhe Lin, Huchuan Lu, and You He. Capsal: Leveraging captioning to boost semantics for salient object detection. In *IEEE CVPR*, 2019.

[102] Pingping Zhang, Dong Wang, Huchuan Lu, Hongyu Wang, and Xiang Ruan. Amulet: Aggregating multi-level convolutional features for salient object detection. In *IEEE ICCV*, pages 202–211, 2017.

[103] Jiaxing Zhao, Ren Bo, Qibin Hou, Ming-Ming Cheng, and Paul Rosin. Flic: Fast linear iterative clustering with active search. *Computational Visual Media*, 4(4):333–348, 2018.

[104] Jia-Xing Zhao, Yang Cao, Deng-Ping Fan, Ming-Ming Cheng, Xuan-Yi Li, and Le Zhang. Contrast prior and fluid pyramid integration for rgbd salient object detection. In *IEEE CVPR*, pages 3927–3936, 2019.

[105] Jia-Xing Zhao, Jiang-Jiang Liu, Deng-Ping Fan, Yang Cao, Jufeng Yang, and Ming-Ming Cheng. EGNet: Edge Guidance Network for Salient Object Detection. In *IEEE ICCV*, pages 8779–8788, 2019.

[106] Xiaoju Zheng, Zheng-Jun Zha, and Liansheng Zhuang. A feature-adaptive semi-supervised framework for co-saliency detection. In *ACM MM*, pages 959–966, 2018.

[107] C Lawrence Zitnick and Piotr Dollár. Edge boxes: Locating object proposals from edges. In *ECCV*, 2014.