

2019 及 2020 年显著性检测论文总结

May 8, 2020

Contents

1 概述与分类	1
1.1 内容概述	1
1.2 模型分类	2
2 模型结构	2
2.1 RGB 图像显著性检测	2
2.2 多模态图像显著性检测	3
2.3 视频显著性检测	4
3 模型与往年对比	5
3.1 边缘检测	5
3.2 注意力机制	5
4 训练方法与损失函数	6
4.1 训练方法	6
4.2 损失函数	6
5 公开数据集汇总	7
6 其他	7

1 概述与分类

1.1 内容概述

本次总结论文共有 50 篇，其中 19 年 35 篇，分别来自 CVPR (15 篇), ICCV (13 篇), NIPS (2 篇), AAAI (3 篇) 及 IJCAI (2 篇)。20 年目前已搜集并总结了来自 CVPR (7 篇) 及 AAAI (8 篇) 的论文。由于现阶段 20 年已公开显著性检测论文数量较少，因此总结主要针对 19 年相关论文。此外，由于主流显著性检测方法均基于深度学习，总结中没有将基于深度学习的模型与基于手工特征的方法进行对比。

1.2 模型分类

显著性检测从卷积神经网络，尤其是全卷积神经网络广泛应用以来得到了长足的发展，相关模型可以从多个角度进行分类。

从任务角度，显著性检测可分以下三类：

- 基于单张 RGB 图像的显著性检测，不属于后两类的模型均可归为此类。其中 [2, 8, 21, 36] 为图像 co-saliency 检测，近期该领域活跃程度上升明显；
- 基于多模态图像的显著性检测，包括 RGBD 图像 [5, 14, 17, 34, 39, 41]，光场图像 [15, 16, 22, 38] 以及多光谱图像 [19]；
- 视频显著性检测 [3, 6, 9, 12, 26, 40]

从网络学习方法上，可分为监督学习，弱监督学习与无监督学习。

- 监督学习应用最广泛，且 deep supervision 机制应用普遍；
- 弱监督学习 [30, 33, 35, 37] 使用了多种不同的弱监督信息，例如图像类别、摘要、涂鸦标注来指导显著性检测或辅助提升显著性检测效果。需注意的是 [37] 中图像摘要并非输入标注，而仅是模型内部的一个模块；
- 无监督学习 [13] 无需标注数据，模型使用多种人工特征对图像进行显著性检测得出粗糙检测结果，经优化后作为特征训练显著性检测网络。

后两种训练方法主要针对显著性数据集标注成本高的问题，试图通过低成本标注或无标注数据训练模型。具体学习方法将会在 4 中进行总结。

从模型主要所使用算法上，除典型的全卷积神经网络外，胶囊网络 [11]，变分自编码器 (VAE) [7]，条件变分自编码器 (CVAE) [34]，条件随机场 (CRF) [29, 32] 也得到了应用。而在视频显著性检测及光场显著性检测中，为利用图像帧、focal stack 间的相关性或深度信息，ConvLSTM [3, 14, 16, 22, 26, 38, 40] 与 convGRU [45] 也有广泛应用。此外 20 年模型中，[16, 45] 采用了 teacher-student 网络结构来进行知识蒸馏。

下面对以上模型分类进行详细总结。在 2 中，对显著性检测模型的常见结构进行了总结与分类；在 3 中，列举并对比了 19 及 20 年模型与以往显著性检测模型相比的显著特点；在 4 中，对模型训练方法及多种损失函数进行了总结；最后在 5 中，汇总了显著性检测各个领域常用及新发布的公开数据集。

2 模型结构

由于不同任务所对应模型结构差别较大，本章节将根据模型任务类型进行分类总结。

2.1 RGB 图像显著性检测

基于 RGB 图像的显著性检测模型研究热度最高，目前已经有部分公认的模型结构与设计思路，整体可分为特征提取与显著性检测两个部分。

特征提取

特征提取部分通常使用在 ImageNet 等大型数据集上预训练的图像分类网络去除末尾全连接层后作为 backbone 网络，如 VGG16, VGG19, ResNet34。backbone 网络逐层提取出不同尺度与深度的卷积特征图，通常为五层；随后，特征图会经过诸如多尺度特征融合，特征筛选的处理后，最终输出二值化的像素级显著性推断结果。

在特征提取环节，多数模型与 backbone 网络耦合度低，可更换 backbone 网络而无需调整模型结构。对于提取到的不同特尺度征图，论文中普遍有以下认识：

- 高层级特征抽象程度高，是显著性检测的主要特征依据，且可以指导对低层级特征的筛选；但由于其经过多层池化后分辨率低，推断结果往往缺乏足够的边缘分类精度。
- 低层级特征抽象层次低而细节丰富，可以有效提升边界分类准确率；但低层级特征往往会包含复杂背景及较多噪声，直接利用可能会降低模型整体表现；此外，低层级特征分辨率高，处理时间与空间开销明显 [27]。

针对以上特点及问题，部分模型提出了解决方案，如 [44] 中对高底层特征分别使用了不同的分支进行处理；[42] 等模型中丢弃了最浅层的特征图；而 [27] 中则更激进地直接抛弃了前两级低层级特征，仅利用后三层特征，显著提升了推断速度，在 352*352 像素输入、设备为 1080ti 的情况下可达 110fps。

显著性检测

在显著性检测环节，模型结构差别明显，但仍有以下共同之处：宏观上广泛地使用了 Encoder-Decoder 结构及其变体来进行多尺度特征融合，且多数以 U-Net 为基础进行实现，例如 [10] 中明确提及了其结构基于 U-Net，并通过改进的池化操作进行了优化；[18] 将模型分为了预测与优化两个连续的子模块，均采用了类似 U-Net 的结构。此外 [23] 中还将 U-Net 的 Top-Down 与 Bottom-Up 模式进行了进一步推广，提出了一种迭代优化方式。除常规的 Encoder-Decoder 结构外，[27] 中模型将高层特征提取部分分为了两个结构相同的分支，并在最后连接了 Partial Decoder 以代替传统的单 Decoder 结构。[28] 中模型同样采用了双分支结构，但两个分支非对称且分别进行边缘检测与显著性检测。

Encoder-Decoder 结构也存在明显的缺点：逐层池化后特征图分辨率降低且部分丢失了特征的位置信息，明显降低了预测结果边缘的分类准确率，针对此问题，论文中主要通过添加边缘检测模块及改进损失函数来解决，该部分将在 3 及 4 中详细总结。此外，注意力机制也在模型中得到了广泛的应用，该部分也会在 3 中详细讨论。

在细节上，不同模型间存在较大差异，以下为部分典型结构。特征图融合策略中，常用方法为拼接与对应元素相加，此外 [1] 中还使用了更加激进的哈达玛积操作；残差结构可以在增强网络学习能力的同时保持梯度，且使模块可以更灵活地加入现有网络结构，在模型中有众多显式或隐式的体现；空间金字塔结构可以有效提取特征图中的多尺度信息，典型代表有：[10] 使用不同 kernel size 的池化层来提取多尺度信息，[44] 使用多种膨胀率的卷积核进行空洞卷积。

2.2 多模态图像显著性检测

RGBD 图像显著性检测

RGBD 图像显著性检测基于 RGB 图像显著性检测在结构上相似程度很高，但相比 RGB 图像显著性检测有以下新挑战：一是缺乏在类似 ImageNet 等大规模数据集上预训练的 backbone 网络；二

是现有的跨模态融合方式效率较低；三是深度信息采集的准确率与精度较低。显著性检测领域研究主要针对后两者。

对于第二个问题，目前主流的融合方法除已被证明低效的直接拼接外，按模态融合发生的位置可分为以下三类：early fusion, middle fusion 及 late fusion。不过并非所有模型都可以明确分为以上类别，如 [17, 41] 在每层 rgb 特征提取块后均进行了跨模态融合，[5] 则在观察到 RGB 与深度图像训练及推断过程中特征图的相似性后，通过 naive gray color mapping 将深度图转化为 RGB 图像，与真实 RGB 图像拼接后作为模型输入，通过孪生网络结构来同时进行特征提取。该部分模型整体上差异较大，暂没有较统一的结构。

对于第三个挑战，就深度图而言，受限于采集设备性能，深度图像上会有随机分布的深度信息错误或者缺失，且错误信息主要发生在物体边缘区域，这些问题会导致不准确且模糊的显著性预测结果。针对以上问题，[39] 通过融合经 VGG 直接提取出的初始特征图和深度图二值化切片得到优化的 RGB 与深度特征图，并在观察到优化后的高层级 RGB 特征图比深度特征图包含更多边界信息后，从中提取出了边界信息，一定程度上解决了该问题。

光场图像显著性检测

光场图像通过光场相机拍摄，经过后处理可以得到全焦图像与对应的 focal stacks，也就是不同景深下的同一视角图像的集合。2D 的 RGB 图像与 3D 的 RGBD 图像难以捕获人眼注意力在不同深度上的转移，并且在复杂场景下检测难度较高，而利用 focal stacks 信息可以有效解决以上问题。

[22] 分别通过两个分支来处理全焦图像和 focal stacks，为利用 focal stacks 内部不同景深切片间的相关性，模型在 focal stacks 支路使用了 ConvLSTM 对切片序列进行处理，最终通过将两个支路对应元素相加来融合特征，属于 late fusion 策略。

[38] 首先通过 Channel Attention 机制对提取出的全焦图像和 focal stacks 特征进行优化，随后经过两层不同尺寸的 ConvLSTM 处理，逐渐优化检测结果，并将高层特征储存作为记忆，指导底层特征的选择。

[15] 着重考虑了光场图像所包含的几何信息，并将问题分解为了两个子问题：光场渲染和光场驱动的显著性检测。光场渲染/合成采用了基于深度学习的方法，通过中心图像和对应的深度信息生成光场图像；光场显著性检测网络包含多视角注意力模块和多视角检测模块，输入生成的光场图像后最终得到输出检测结果。（这篇文章很多地方没看懂，特别是 warp 操作）

[16] 使用了知识蒸馏结构，teacher 和 student 网络分别输入 focal stacks 和对应单张 RGB 图像训练，通过 teacher 网络指导 student 网络学习，最终得到的 student 网络仅需 RGB 图像输入，即可达到接近 teacher 网络的表现。

2.3 视频显著性检测

根据 [9] 中的分类，视频显著性检测可分为：利用循环神经网络（特别是 ConvLSTM）[3, 26, 40] 与直接利用卷积神经网络 [9, 12]。

前者往往通过 ConvLSTM 来捕获帧间的长期与短期相关性，而后者只需要短期内容来检测显著性物体。具体实践中，不同模型采用了多种特征，例如 [9] 中采用了光流影像，着重提取画面运动信息；[3] 关注并利用了 saliency-shift 现象，即显著性在帧间会存在平滑的转移；[12] 认为基于 ConvLSTM 的算法难以同时处理时间与空间信息，并提出了辅助池化模块降低运算量。除常规视频外，[40] 采用了立体视频输入，通过左右镜头采集图像的视差进行深度估计，辅助显著性检测。

在评估指标上，相比其他任务，视频显著性检测更加重视算法的实时性。因此模型往往不会采用与 RGB 图像显著性检测相似的 backbone 网络，而是使用 S3D、MobileNetV3 等轻量级网络进行特征提取。

3 模型与往年对比

19 年及 20 年显著性检测模型与往年相比，有两个明显的特点：一是更加注重物体边缘分类效果；二是广泛地使用了注意力机制。

3.1 边缘检测

19 年起，相当多的论文中都提及了显著性检测物体边缘模糊的问题，其主要原因为特征提取时的逐层池化使得空间分辨率降低，难以获得清晰的物体边缘，特别是在复杂场景下边界分类问题明显；条件随机场作为后处理手段虽然可以优化显著性预测结果，但考虑到模型运算量已基本不再使用。

为解决该问题，论文中主要可分为两种思路：一是显式地引入边缘检测模块或与边缘检测网络联合训练以提升边界分类准确率；二是通过构造新的损失函数，增加边缘像素的权重，使模型更关注边界分类效果。本章节主要关注前者，后者将在 4 中详细总结。

通过额外网络结构增强边缘像素分类结果有以下几种思路：

- 独立模块：[10] 在主模型外额外添加了可选的边缘检测支路，主模型可独立运行；
- 分支结构：[28] 中采用的双分支结构，分别对显著性物体和边缘提取特征，并逐层迭代优化；[20] 则使用了三个分支，分别检测边缘、内部及边缘于内部之间的空隙；
- 融合结构：[42] 在从高到低逐层融合特征后提取出边缘信息，并显式地添加了边缘检测 loss；
- 其他：[4] 方法较独特，在注意力反馈模块中通过对原始预测结果分别进行腐蚀、膨胀再叠加，从而得到聚焦于物体边缘的三元注意力图，迭代纠正先前在边界上的预测错误。

3.2 注意力机制

注意力机制早期主要在自然语言处理领域应用广泛。近年来，受人类视觉系统（Human visual system）启发，注意力机制也逐渐融入了计算机视觉领域。注意力机制主要可分为软注意力（Soft Attention）及硬注意力（Hard Attention）。软注意力可导，易于融合进现有网络结构并进行端到端训练，从 19 年起，在显著性检测中应用明显增加；硬注意力相对较少，在高分辨率图像的显著性检测 [31] 中有所应用。下面分别对两种注意力机制进行总结。

Soft Attention

软注意力机制，通过作用方式可分为 Channel Attention 与 Spatial Attention。Channel Attention 多数可视为受 SEnet（Squeeze-and-Excitation Networks）的启发，通过学习到的注意力对图像不同通道进行加权处理，实验表明可以较明显提高模型表现；由于其残差结构，Channel Attention 可以方便地添加进现有模型。而 Spatial Attention 则直接为特征图的每一个空间位置赋予权重，在空间上对特征进行筛选。多数模型同时使用了两种软注意力机制。例如 [44] 对高层特征使用 Channel Attention 进行通道加权，对低层特征使用 Spatial Attention 进行特征筛选；[24] 使用空间金字塔结构获得多尺度 Attention map，增加显著性物体区域权重。

Hard Attention

硬注意力机制应用较少，在高分辨率图像中，直接对显著性物体进行检测难度较大，因此 [31] 在全局信息指导下进行对图像的显著性区域进行裁剪，再进一步进行显著性检测，可以视为一种硬注意力机制的应用。

4 训练方法与损失函数

本章节将从模型训练方法、损失函数两个方面来对模型进行总结。

4.1 训练方法

显著性检测的训练方法可分为监督学习，半监督学习与无监督学习。监督学习使用最普遍，且 deep supervision 机制应用广泛，loss 会作用于模型的多个不同尺度特征图，或分别用不同的 loss 作用于不同的网络层级，如 [4] 中 Boundary-Enhanced Loss 仅作用于最浅的两层特征图，而 cross-entropy loss 作用于网络的全部五层特征图。

半监督学习利用像素级标注之外的其他标注数据训练显著性检测网络，主要意义在于降低对高成本显著性标注数据的依赖程度。[33] 利用了图像类别标注及图像摘要，在多种 loss 的辅助下通过类别和摘要信息训练检测网络，并提出了一种利用弱监督信息的通用框架；[30] 通过稀疏标注的视频帧生成连续的伪标注；[35] 通过涂鸦进行前景和背景标注来训练模型。

无监督学习 [13] 直接输入无标注图像，首先通过多种传统方法分别提取出显著性伪标签，其次，并不直接将每个分支生成的伪标签输入显著性检测网络进行推断，而是分别先利用额外的神经网络迭代式降噪，之后综合使用了移动平均和 CRF 等方法优化输出后，输入最终的显著性检测网络。

4.2 损失函数

显著性检测领域最常用的损失函数为二值交叉熵损失函数 (Binary Cross-Entropy Loss Function)，不过出于突出图像部分区域或弥补交叉熵损失缺陷等目的，部分模型提出了新的损失函数或者引入了现有的其他损失函数。以下为部分示例。

- 边缘检测：[44] 中运用拉普拉斯算子构造损失函数，在对图像求梯度后再次运用交叉熵损失函数得到对边缘敏感的损失函数，与交叉熵损失函数加权相加后可以提高模型对边界的分类效果；[18] 混合使用了三种损失函数：交叉熵，SSIM loss 与 IoU loss，分别对像素级、patch 级以及全局级的预测结果进行优化，间接增强了边缘分类结果；
- 改进交叉熵损失函数：[25] 中指出了交叉熵损失函数的三个问题：1. 不关注图像结构信息，2. 当背景占比很大时，前景的 loss 会被稀释，3. 只能平等的处理每个像素的 loss；为此作者提出了 PPA loss，每个像素的权重通过周围像素来定义，使难分类的像素获得更高的权重；
- 其他目的：针对 co-saliency 任务特点，[8] 基于 perceptual loss 提出了 co-perceptual loss，推动模型学习图像组内相关性。[43] 则改造了常用的评估指标 F-measure，使其可导并直接作为损失函数使用。

5 公开数据集汇总

本章节列举了所总结论文中使用及新发布的公开数据集，括号中为别名或注释。新发布的数据集均通过引用对应论文进行了标注。

RGB 显著性检测数据集

- 常规 RGB 显著性数据集: DUTS, ECSSD, THUR, HKU-IS, PASCAL-S, DUT-OMRON, SOD, HRSOD [31] (高分辨率显著性数据集)
- 带有图像摘要的 RGB 显著性数据集: COCO-CapSal [37]
- Co-Saliency 检测数据集: iCoseg, MSRC, Cosal2015, COCO-SEG [21], CoSOD3k [2]
- 眼动数据 RGB 显著性数据集: SALICON, OSIE-SR

RGBD 显著性检测数据集

- NJU2000, NLPR (RGBD1000), SSB (STEREO), RGBD135 (DES)

光场显著性检测数据集

- LFSD, HFUT-Lytro, DUTLF [22]

视频显著性检测数据集

- 常规视频数据集: VOS, MCL, UVSD, SegV2, DAVIS, FBMS, ViSal, DAVSOD [3];
- 常规视频眼动数据集: DHF1K, Hollywood2, UCFSports;
- 立体视频眼动数据集: SAVAM, SVS [40]。

6 其他

本次总结论文中，被引数在 20 以上的文章有: [3, 4, 10, 18, 23, 24, 27, 41, 42, 44]，绝大部分为 CVPR19 会议论文。

References

- [1] Zuyao Chen, Qianqian Xu, Runmin Cong, and Qingming Huang. Global context-aware progressive aggregation network for salient object detection. *arXiv preprint arXiv:2003.00651*, 2020.
- [2] Deng-Ping Fan, Zheng Lin, Ge-Peng Ji, Dingwen Zhang, Huazhu Fu, and Ming-Ming Cheng. Taking a deeper look at the co-salient object detection. In *IEEE CVPR*, 2020.
- [3] Deng-Ping Fan, Wenguan Wang, Ming-Ming Cheng, and Jianbing Shen. Shifting more attention to video salient object detection. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.
- [4] Mengyang Feng, Huchuan Lu, and Errui Ding. Attentive feedback network for boundary-aware salient object detection. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.
- [5] 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. *arXiv preprint arXiv:2004.08515*, 2020.
- [6] Yuchao Gu, Lijuan Wang, Ziqin Wang, Yun Liu, Ming-Ming Cheng, and Shao-Ping Lu. Pyramid constrained self-attention network for fast video salient object detection.
- [7] Bo Li, Zhengxing Sun, and Yuqi Guo. Supervae: Superpixelwise variational autoencoder for salient object detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 8569–8576, 2019.
- [8] Bo Li, Zhengxing Sun, Lv Tang, Yunhan Sun, and Jinlong Shi. Detecting robust co-saliency with recurrent co-attention neural network. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, pages 818–825. AAAI Press, 2019.
- [9] Haofeng Li, Guanqi Chen, Guanbin Li, and Yizhou Yu. Motion guided attention for video salient object detection. In *The IEEE International Conference on Computer Vision (ICCV)*, October 2019.
- [10] Jiang-Jiang Liu, Qibin Hou, Ming-Ming Cheng, Jiashi Feng, and Jianmin Jiang. A simple pooling-based design for real-time salient object detection. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.
- [11] Yi Liu, Qiang Zhang, Dingwen Zhang, and Jungong Han. Employing deep part-object relationships for salient object detection. In *The IEEE International Conference on Computer Vision (ICCV)*, October 2019.
- [12] Kyle Min and Jason J. Corso. Tased-net: Temporally-aggregating spatial encoder-decoder network for video saliency detection. In *The IEEE International Conference on Computer Vision (ICCV)*, October 2019.
- [13] Tam Nguyen, Maximilian Dax, Chaithanya Kumar Mummadi, Nhung Ngo, Thi Hoai Phuong Nguyen, Zhongyu Lou, and Thomas Brox. Deepusps: Deep robust unsupervised saliency prediction via self-supervision. In *Advances in Neural Information Processing Systems*, pages 204–214, 2019.

- [14] Yongri Piao, Wei Ji, Jingjing Li, Miao Zhang, and Huchuan Lu. Depth-induced multi-scale recurrent attention network for saliency detection. In *The IEEE International Conference on Computer Vision (ICCV)*, October 2019.
- [15] Yongri Piao, Zhengkun Rong, Miao Zhang, Xiao Li, and Huchuan Lu. Deep light-field-driven saliency detection from a single view. In *International Joint Conference on Artificial Intelligence (IJCAI)*, 2019.
- [16] Yongri Piao, Zhengkun Rong, Miao Zhang, and Huchuan Lu. Exploit and replace: An asymmetrical two-stream architecture for versatile light field saliency detection.
- [17] Yongri Piao, Zhengkun Rong, Miao Zhang, Weisong Ren, and Huchuan Lu. A2dele: Adaptive and attentive depth distiller for efficient rgb-d salient object detection. In *CVPR*, 2020.
- [18] Xuebin Qin, Zichen Zhang, Chenyang Huang, Chao Gao, Masood Dehghan, and Martin Jagersand. Basnet: Boundary-aware salient object detection. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.
- [19] Shaoyue Song, Hongkai Yu, Zhenjiang Miao, Jianwu Fang, Kang Zheng, Cong Ma, and Song Wang. Multi-spectral salient object detection by adversarial domain adaptation.
- [20] Jinming Su, Jia Li, Yu Zhang, Changqun Xia, and Yonghong Tian. Selectivity or invariance: Boundary-aware salient object detection. In *The IEEE International Conference on Computer Vision (ICCV)*, October 2019.
- [21] Chong Wang, Zheng-Jun Zha, Dong Liu, and Hongtao Xie. Robust deep co-saliency detection with group semantic. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 8917–8924, 2019.
- [22] Tiantian Wang, Yongri Piao, Xiao Li, Lihe Zhang, and Huchuan Lu. Deep learning for light field saliency detection. In *The IEEE International Conference on Computer Vision (ICCV)*, October 2019.
- [23] Wenguan Wang, Jianbing Shen, Ming-Ming Cheng, and Ling Shao. An iterative and cooperative top-down and bottom-up inference network for salient object detection. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.
- [24] Wenguan Wang, Shuyang Zhao, Jianbing Shen, Steven C. H. Hoi, and Ali Borji. Salient object detection with pyramid attention and salient edges. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.
- [25] Jun Wei, Shuhui Wang, and Qingming Huang. F3net: Fusion, feedback and focus for salient object detection. *arXiv preprint arXiv:1911.11445*, 2019.
- [26] Xinyi Wu, Zhenyao Wu, Jinglin Zhang, Lili Ju, and Song Wang. Salsac: A video saliency prediction model with shuffled attentions and correlation-based convlstm.
- [27] Zhe Wu, Li Su, and Qingming Huang. Cascaded partial decoder for fast and accurate salient object detection. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

- [28] Zhe Wu, Li Su, and Qingming Huang. Stacked cross refinement network for edge-aware salient object detection. In *The IEEE International Conference on Computer Vision (ICCV)*, October 2019.
- [29] Yingyue Xu, Dan Xu, Xiaopeng Hong, Wanli Ouyang, Rongrong Ji, Min Xu, and Guoying Zhao. Structured modeling of joint deep feature and prediction refinement for salient object detection. In *The IEEE International Conference on Computer Vision (ICCV)*, October 2019.
- [30] Pengxiang Yan, Guanbin Li, Yuan Xie, Zhen Li, Chuan Wang, Tianshui Chen, and Liang Lin. Semi-supervised video salient object detection using pseudo-labels. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 7284–7293, 2019.
- [31] Yi Zeng, Pingping Zhang, Jianming Zhang, Zhe Lin, and Huchuan Lu. Towards high-resolution salient object detection. In *The IEEE International Conference on Computer Vision (ICCV)*, October 2019.
- [32] Yu Zeng, Yunzhi Zhuge, Huchuan Lu, and Lihe Zhang. Joint learning of saliency detection and weakly supervised semantic segmentation. In *The IEEE International Conference on Computer Vision (ICCV)*, October 2019.
- [33] Yu Zeng, Yunzhi Zhuge, Huchuan Lu, Lihe Zhang, Mingyang Qian, and Yizhou Yu. Multi-source weak supervision for saliency detection. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.
- [34] 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. *arXiv preprint arXiv:2004.05763*, 2020.
- [35] Jing Zhang, Xin Yu, Aixuan Li, Peipei Song, Bowen Liu, and Yuchao Dai. Weakly-supervised salient object detection via scribble annotations. *arXiv preprint arXiv:2003.07685*, 2020.
- [36] Kaihua Zhang, Tengteng Li, Bo Liu, and Qingshan Liu. Co-saliency detection via mask-guided fully convolutional networks with multi-scale label smoothing. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.
- [37] Lu Zhang, Jianming Zhang, Zhe Lin, Huchuan Lu, and You He. Capsal: Leveraging captioning to boost semantics for salient object detection. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.
- [38] Miao Zhang, Jingjing Li, JI WEI, Yongri Piao, and Huchuan Lu. Memory-oriented decoder for light field salient object detection. In *Advances in Neural Information Processing Systems*, pages 896–906, 2019.
- [39] Miao Zhang, Weisong Ren, Yongri Piao, Zhengkun Rong, and Huchuan Lu. Select, supplement and focus for rgb-d saliency detection. In *CVPR*, 2020.
- [40] Qiudan Zhang, Xu Wang, Shiqi Wang, Shikai Li, Sam Kwong, and Jianmin Jiang. Learning to explore intrinsic saliency for stereoscopic video. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

- [41] 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 *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.
- [42] 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 *The IEEE International Conference on Computer Vision (ICCV)*, October 2019.
- [43] Kai Zhao, Shanghua Gao, Wenguan Wang, and Ming-Ming Cheng. Optimizing the f-measure for threshold-free salient object detection. In *The IEEE International Conference on Computer Vision (ICCV)*, October 2019.
- [44] Ting Zhao and Xiangqian Wu. Pyramid feature attention network for saliency detection. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.
- [45] Ziqi Zhou, Zheng Wang, Huchuan Lu, Song Wang, and Meijun Sun. Multi-type self-attention guided degraded saliency detection.