

# 基于深度学习的显著性目标检测方法综述

罗会兰<sup>1</sup>,袁 璞<sup>1</sup>,童 康<sup>2</sup>

(1. 江西理工大学信息工程学院,江西赣州 341000;2. 南京航空航天大学电子信息工程学院,江苏南京 211106)

**摘要:** 显著性目标检测旨在对图像中最显著的对象进行检测和分割,是计算机视觉任务中重要的预处理步骤之一,且在信息检索、公共安全等领域均有广泛的应用.本文对近期基于深度学习的显著性目标检测模型进行了系统综述,从检测粒度的角度出发,综述了将深度学习引入显著性目标检测领域之后的研究成果.首先,从三个方面对显著性目标检测方法进行了论述:稀疏检测方法,密集检测方法以及弱监督学习下的显著性目标检测方法.然后,简要介绍了用于显著性目标检测研究的主流数据集和常用性能评价指标,并对各类主流模型在三个使用最广泛的数据集上进行了性能比较分析.最后,本文分析了显著性目标检测领域目前存在的问题,并对今后可能的研究趋势进行了展望.

**关键词:** 显著性目标检测;深度学习;卷积神经网络;视觉显著性;弱监督学习;计算机视觉任务

**中图分类号:** TP389.1

**文献标识码:** A

**文章编号:** 0372-2112(2021)07-1417-11

**电子学报 URL:** <http://www.ejournal.org.cn>

**DOI:** 10.12263/DZXB.20200651

## Review of the Methods for Salient Object Detection Based on Deep Learning

LUO Hui-lan<sup>1</sup>, YUAN Pu<sup>1</sup>, TONG Kang<sup>2</sup>

(1. College of Information Engineering, Jiangxi University of Science and Technology, Ganzhou, Jiangxi 341000, China;

2. College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing, Jiangsu 100084, China)

**Abstract:** Salient object detection aims to detect and segment the most salient objects in the image. It is one of the important preprocessing steps in computer vision tasks, and it has a wide range of applications in information retrieval, public safety and other fields. This paper systematically reviews the recent research on the salient object detection models based on deep learning. From the perspective of detection granularity, the research results of applying deep learning into the field of salient object detection are reviewed. First, the salient object detection methods are discussed from three aspects: sparse detection methods, dense detection methods and weakly-supervised learning methods. Then, the mainstream data sets and common performance evaluation indicators used for salient object detection research are briefly introduced, and the performance of various mainstream models on the three most widely used data sets are compared and analyzed. Finally, this paper analyzes the current problems in the field of salient object detection and prospects for possible future research trends.

**Key words:** salient object detection; deep learning; convolutional neural network; visual saliency; weakly supervised learning; computer vision task

### 1 引言

1998年Itti等人<sup>[1]</sup>基于对灵长类动物视觉系统的模拟提出了一种视觉注意力机制,由此开启了计算机视觉领域对于显著性检测的研究.早期的显著性检测大多将研究重心放在人眼注意点的检测上,即检测图像中最吸引人类眼光的地方,检测结果为一组光亮点组成的感兴趣区域.随着计算机视觉的进一步发展,单一

的视觉注意点检测逐渐难以满足其需求.因此,研究者提出将显著性检测看作是一个图像分割问题,将图像的显著性目标区域从背景中分割出来,并将该任务称为显著性目标检测(Salient Object Detection, SOD)<sup>[2]</sup>.

作为重要的图像预处理步骤,显著性目标检测广泛应用于各类计算机视觉任务中:利用显著性内容作

为图像描述的指导<sup>[3,4]</sup>;采用显著性检测指导对高分辨率的卫星图像进行无监督特征学习下的场景分类<sup>[5]</sup>;将无监督学习转化为多实例学习,实现定位和分类的任务<sup>[6]</sup>;提供对象级线索从而帮助无监督视频对象分割<sup>[7]</sup>;为目标检测任务提供图像的显著性轮廓信息<sup>[8]</sup>;用于构建测试视觉问答模型性能的数据集<sup>[9]</sup>等等。

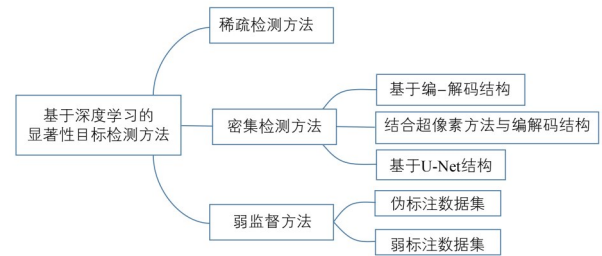
显著性目标检测领域的发展十分迅速,早期的综述论文只对一些传统方法进行了归纳<sup>[10,11]</sup>。近年来出现了一些论述前沿方法的文献,如文献<sup>[12]</sup>综述了阈值化分割在显著性目标检测中的应用,并讨论了一些虽还没被采用但很有潜力的阈值方法。文献<sup>[13]</sup>把显著性目标检测和类别目标检测并在一起进行了综述,介绍了目标检测中常用的模型和方法。文献<sup>[14]</sup>以时间顺序描述了显著性目标检测模型的演变,并简要介绍了显著性目标检测的起源和关键性技术。文献<sup>[15]</sup>根据特征提取策略的不同,综述了传统学习和深度学习下的显著性目标检测模型。文献<sup>[16]</sup>围绕网络架构、监督方式、学习范式等角度概述了基于深度学习的显著性目标检测模型,并通过研究噪声扰动和对抗攻击对算法和模型的影响,为检测模型的实际应用提供了可用建议。

不同于以上文献综述,本文从检测粒度的角度出发,综述了将深度学习引入显著性目标检测领域之后的研究成果,并将这些显著性目标检测方法分为三大类:稀疏检测方法,密集检测方法及弱监督学习方法。本文的稀疏与密集检测是以网络模型对图像的处理单元不同而定义的,前者的处理单元通常是超像素或候选对象等非独立像素的单元,后者的处理单元则是单个的像素<sup>[17]</sup>。这两种检测方法都需要大量的人工标注,弱监督方法的提出减轻了模型对人工标注数据集的依赖,本文也对这类方法做了详尽综述,并根据训练样本中标注来源的不同,将其分为基于伪标注数据集和基于弱标注数据集的方法。此外,本文还综述了用于显著性目标检测研究的七个主流数据集和三个评价指标,并选取了其中三个使用最广泛的数据集对各个主流模型进行性能分析。本文旨在为国内研究者提供一个包含最新方法的综述,增加对显著性目标检测研究的理解,以得到一些启发。

## 2 显著性目标检测方法

早期基于深度学习的显著性目标检测方法通常使用卷积神经网络在超像素层面提取特征,进行目标显著性的判断,得到稀疏的显著性检测结果图。随着全卷积神经网络<sup>[18]</sup>在像素级语义分割中取得的巨大成功,基于像素级别的密集检测方法随之产生,并逐渐成为显著性目标检测的主流方法。方法综述的整体结构如

图1所示。

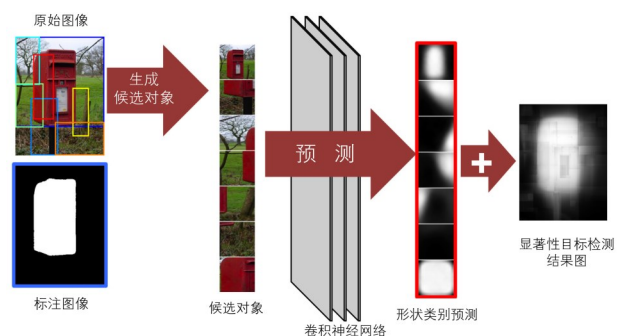


### 2.1 稀疏检测方法

以超像素作为处理单元的代表工作有 SuperCNN<sup>[19]</sup>, MCDL<sup>[20]</sup> 和 ELD<sup>[21]</sup> 等, SuperCNN<sup>[19]</sup> 首先使用 SLIC<sup>[22]</sup> 算法在不同尺度下分割出多幅超像素大小不一的图像,且为每一个超像素生成两组特征序列,然后将这些序列送入卷积神经网络中训练得到图像的层级特征,并采用 softmax 激活函数将特征向量转化为每个超像素的显著性得分,即每个区域具有显著性的概率,再使用 argmax 分类出具有显著性的超像素区域,得到显著性映射图,最后将不同尺度下生成的结果图进行加权融合,得到最终的显著性目标检测图。

不同于 SuperCNN<sup>[19]</sup> 中对多尺度特征融合的探索, MCDL<sup>[20]</sup> 致力于提取图像中的上下文信息,分别以每个超像素为中心,采用了大小不一的两个窗口提取各超像素的局部和全局上下文信息进行特征学习。Lee 等人提出的 ELD<sup>[21]</sup> 则在 MCDL<sup>[20]</sup> 的基础上引入了传统方法中手工提取的显著性特征,并利用该特征为深度卷积神经网络中提取的特征提供互补信息,最终推断出每个超像素的显著性。

以候选对象作为处理单元的工作有 SCSD<sup>[23]</sup>, LEGS<sup>[24]</sup>, MDF<sup>[25]</sup> 和 MAP<sup>[26]</sup> 等。SCSD<sup>[23]</sup> 的网络结构如图2所示,该模型首先通过 SSOR<sup>[27]</sup> 算法生成大量的候选对象,然后将这些候选对象送入卷积神经网络中,预测出每个候选对象中显著性对象所占区域的形状类别,并生成相应形状的掩码,即每个候选对象的显著性图,最后将这些候选对象的显著性映射相加即为整幅



图像的显著性目标检测结果. Wang 等人提出的 LEGS<sup>[24]</sup>使用 GOP 算法<sup>[28]</sup>生成了一系列候选对象,同时训练一个神经网络来学习图像的局部和全局对比度、纹理及形状等信息,通过将候选对象中的高级语义特征与神经网络中各层级的显著性特征相结合,为每个像素分配显著性值,该方法可以有效地减少检测结果中的高频背景噪声. MDF<sup>[25]</sup>首先对输入图像进行超像素分割,然后为每个超像素选取了三个相嵌套且逐步增大的矩形区域作为候选对象,并分别在卷积神经网络中获得对应的特征向量,再将它们一并送入带有多个全连接层的神经网络中进行回归训练,得到最终的显著性区域分类. 这些方法大多倾向于检测单个显著性对象,对此, Zhang 等人<sup>[26]</sup>基于最大后验原理和卷积神经网络提出了一种能生成大量带得分标注框的模型 MAP,通过一个子集优化公式筛选出一组用于描述显著性对象的标注框,该模型在具有多个显著对象的场景中可以取得很好的效果.

基于超像素或候选区域的方法往往以一个小区域为计算单元,对它们进行独立的特征提取和显著值判断,因此其生成的显著性目标检测图通常带有模糊的边界.

## 2.2 密集检测方法

### 2.2.1 基于编-解码结构的显著性目标检测方法

UCF<sup>[29]</sup>是最具代表性的基于编-解码结构的显著性目标检测模型,其网络结构如图 3 所示,输入图像首先输入编码器中提取出图像的特征表达,在此基础上进行图像信息的解码重构,然后通过 softmax 分类器得到每个像素的显著性预测.

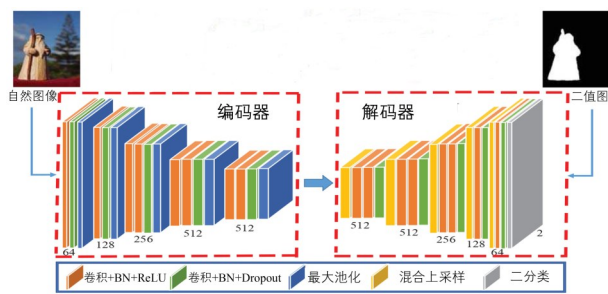


图3 UCF网络结构图<sup>[29]</sup>

然而,编码器在筛除无关特征的同时也会遗失一部分有效的显著性特征,而解码器在上采样的过程中也会不可避免地引入一部分的噪声信息. 对此, Wang 等人<sup>[30]</sup>将递归神经网络的循环计算结构与编-解码网络相结合,在UCF<sup>[29]</sup>的基础上添加了一个时间序列,使得基础网络在不同时刻反复训练,上一时刻网络输出的显著性映射图被当作是下一时刻网络训练的先验知识,与原始图像一并送入编码器中,取得了不错的效

果. Liu 等人<sup>[31]</sup>在文献[30]的基础上增加了编码器内的反馈路径,每个模块的输出以反馈的方式连接到下一时刻相同模块的输入,通过每一层多次共享权值,该网络架构可以在不显著增加参数数量的同时,学习到更多的上下文和空间结构信息.

为进一步学习图像中上下文信息之间的关系, Kuen<sup>[32]</sup>等人在图3中编码器和解码器的中间位置上添加了两个递归网络层,利用递归网络学习的上下文感知特征,模型在迭代过程中逐步将原编-解码网络中生成的粗糙映射图精细化. Xu<sup>[33]</sup>等人通过添加空间和信道维度上的注意力机制,将基于不同子区域的上下文信息聚合到全局上下文中,加强了全局语境,减少了深度特征分辨率的牺牲.

### 2.2.2 结合超像素方法与编解码结构的显著性目标检测方法

此类方法通常利用卷积神经网络获得具有全局信息的显著性估计,然后利用超像素分割所得到的局部上下文线索进行信息补充. 如图4中 Multi-Task DNN<sup>[34]</sup>的网络结构图所示,该模型首先使用全卷积网络捕获原始图像中不同层次上显著性目标的语义信息,从而得到初步的显著性预测,再利用一个基于超像素分割的非线性回归模型对初始预测图进行边界的细化调整,从而得到最终的细粒度显著性映射.

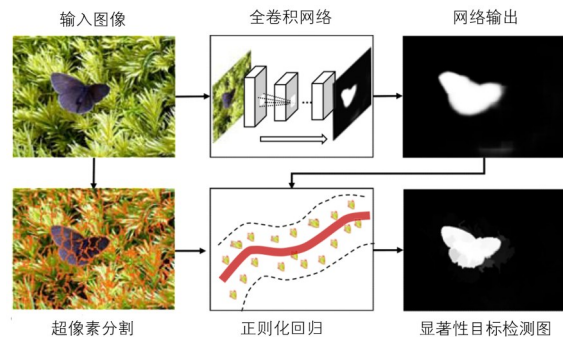


图4 Multi-Task DNN网络结构图<sup>[34]</sup>

虽然 Multi-Task DNN<sup>[34]</sup>通过结合一种拉普拉斯正则化回归方法得到了较为精细的显著性分割图,但该模型较难分辨物体边界邻近具有相似颜色的像素,从而导致边界模糊的问题. 对此, Yuan 等人<sup>[17]</sup>将图4中超像素分割的结果添加进卷积网络的输入,并将稀疏检测中提取的整体目标轮廓和低阶图像特征共同作为指导信息,提高了轮廓定位的精准性. Hu 等人<sup>[35]</sup>则引入了广泛应用在图像二值分割任务中的水平集方法(Level Set Method),水平集在模型的训练过程中被迭代更新从而使得网络的能量函数值最小化,以便将显著性目标从背景中准确地分割出来. 实验表明,水平集方法的加入能显著地提高检测结果的边界精度.

进一步, Tang 等人<sup>[36]</sup>设计了一种带有像素级和超像素级预测分支的双分支网络结构, 然后将两个子网络分别获得的显著性映射通过卷积融合为最终的显著性检测结果. Li 等人<sup>[37]</sup>在此基础上添加了一个全连接的条件随机场<sup>[38]</sup>作为后处理阶段, 增强了对语义信息的挖掘, 同时提高了检测结果中的空间一致性, 减少了背景像素的误判.

联合训练方法综合了稀疏标注方法和全卷积网络方法, 旨在提高对图像语义关系的学习能力, 但模型较复杂, 容易导致信息冗余.

### 2.2.3 基于 U-Net 结构的显著性目标检测方法

基于简单编解码结构的网络模型主要利用卷积层的高层语义特征来预测显著性图, 缺乏低层次的空间细节信息, 这使得显著性检测结果无法保持良好的目标边界. U-Net<sup>[39]</sup>为解决该问题提供了很好的思路, 其网络结构如图 5 所示, 本小节针对基于 U-Net 结构的方法进行综述.

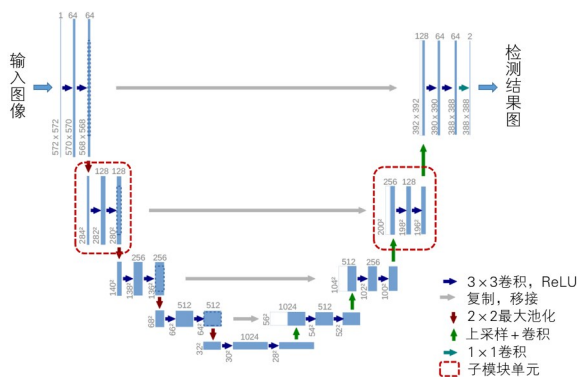


图 5 U-Net 网络结构图<sup>[39]</sup>

#### (1) 改进编-解码器之间跳跃连接的方法

当显著性对象处于杂乱的背景中时, 神经网络的检测精度和速度都会受到影响. 针对该问题, Luo 等人<sup>[40]</sup>在图 5 所示结构中每一条跳跃连接的路径上添加卷积块和平均池化块, 在获取局部和全局上下文信息的同时提升了空间信息的连贯性. Zhang 等人<sup>[41]</sup>在文献<sup>[40]</sup>的基础上, 为两两相邻的跳跃路径之间添加带门函数的连接, 这使得模型能在双向传递信息的路径上进行无效信息的筛选和传递速率的控制, 进而有效地集成多层特征和提高检测速率. 然而, 这些通过设计复杂结构来连接编-解码器间不同层次特征的过程往往比较烦琐, 且多层次特征间的间接整合可能由于长期的依赖问题而存在缺陷.

#### (2) 引入递归结构的改进方法

为了更好地整合不同层级中卷积层提取的特征, Liu 等人<sup>[42]</sup>提出了一种显著性区域检测方法, 将解码部分的卷积层替换成递归卷积层, 分层渐进地提取显著

性映射, 在获得低层空间细节信息的同时增强了模型集成上下文信息的能力. 由于解码器的低级特征中通常保留了大量的边缘信息, Zhang 等人<sup>[43]</sup>在文献<sup>[42]</sup>的基础上, 为图 5 所示的 U-Net 结构中编码部分的第一个子模块添加尺寸为  $1 \times 1$  的卷积层和 ReLU 层, 然后将生成的边界预测添加到解码部分里每一个子模块的原始预测中, 从而对这些子模块的输出进行边界细化处理, 有效地保留了显著性对象的边界.

#### (3) 融合上下文语义信息的方法

Liu 等人<sup>[44]</sup>在解码部分的子模块之间加入全局和局部上下文解码块, 为每一个像素选择性地构造上下文信息, 即对每个像素均生成一个注意力图, 且根据上下文相关性分配相应的注意力权重, 最终得到精准、均匀的检测结果. 在文献<sup>[44]</sup>的基础上, Zhang 等人<sup>[45]</sup>将编码部分中最末端子模块的输出以反馈的形式逐级连接到编码器中各个子模块的输入中, 这使得高层级的全局语义信息转移到了较浅层的卷积层中, 提高了网络的特征学习能力. Chen 等人<sup>[46]</sup>改进了文献<sup>[45]</sup>中的反馈方式, 将编码部分中最后一层的输出通过一个特征聚合模块连接到解码器中各个子模块的输出上, 将低层的细节信息、高层的语义信息和全局上下文信息融合在一起, 提高了显著性检测图的完备性.

Zhao 等人<sup>[47]</sup>将编码器的前半部分和后半部分分别定义为富含空间信息的低层和富含语义信息的高层, 并设立对应的注意力机制将这两部分的输出分开处理, 网络模型在这样的训练过程中可以根据不同层次的特点从而选择更有效的特征进行学习. 进一步, Noori 等人<sup>[48]</sup>对编码层的分类更细致, 他们根据尺度的不同为编码部分的每一层配置相应的注意力引导模块, 这使得模型在有效提取多尺度特征的同时, 增加对具有更多显著性特征映射图的关注, 并衰减显著性鉴别度较弱的特征图. 类似地, Liu 等人<sup>[49]</sup>通过探索池化层的作用, 在 U-Net 的基础上设计了位于解码器相邻子模块之间的全局信息引导模块和特征融合模块, 帮助高阶信息在解码过程中更有效的融合. Wang 等人<sup>[50]</sup>为各卷积层融合更深层次的特征, 并以周期性的方式进行更新, 高级语义信息能直接集成到所有低层特征中, 这避免了合并相邻层特征导致的长期依赖问题, 网络模型也在周期性的过程中逐步优化.

#### (4) 强化显著性目标边界检测的方法

为了提升显著性目标的边界检测精度, Feng 等人<sup>[51]</sup>使模型的编码和解码部分的每一个子模块间保持着双向反馈的连接, 同时模型学习从自身输出的显著性预测中提取边界, 增强对边界的处理能力. Zhao 等人<sup>[52]</sup>直接为模型添加了一个边界检测分支, 并加入显著性对象的轮廓信息对它进行训练, 获得了边界检测

精度的提升. Han 等人<sup>[53]</sup>将预测图与标注图之间的差通过公式转换为一个边缘约束项,并把它加入到损失函数中,以学习到更准确的边界预测. Qin 等人<sup>[54]</sup>则将传统的交叉熵损失函数替换成一个由二进制交叉熵、结构相似度和交并比损失函数融合而成的混合损失函数,并在通用的编-解码结构后增加了一个残差细化模块,利用显著性目标的边缘信息和与其邻近区域信息之间的互补性,生成具有清晰边界的显著性目标检测结果图.

### 2.3 弱监督方法

目前基于深度学习的显著性目标检测方法极大地提升了显著性目标检测的性能,但这些方法的成功在很大程度上依赖于带有像素级标注的大规模数据集. 为了降低训练成本,有些研究者尝试使用没有任何人工标注的数据集或仅使用带有弱标注的数据集来完成对图像的显著性目标检测学习,本文将这些方法统称为弱监督方法. 根据训练图像中标注信息的不同,本小节将弱监督方法分为基于伪标注数据集的方法和基于弱标注数据集的方法.

#### 2.3.1 基于伪标注数据集的方法

自从引入深度学习技术以来,Zhang<sup>[55]</sup>等人首次尝试在训练过程中不使用任何带人工标注的图像数据集,他们选用了一些传统方法下的显著性目标检测模型针对每个训练图像生成相应的一系列映射图,并将这些映射通过权重分配融合成单张样本标注图,即伪标签,然后将其与训练图像一并送入一个基于 DHSNet<sup>[42]</sup>的模型中进行网络训练. 实验结果表明,该模型可以达到相同网络下全监督模型所取得精度的 95% 至 98%.

Zhang 等人<sup>[56]</sup>在文献<sup>[55]</sup>的基础上对伪标签噪声进行显式建模,显著性检测模型和噪声拟合模型在协同工作中联合优化,使模型生成低噪声输出. 该模型的精度甚至比同年的全监督网络 RFCN<sup>[30]</sup>略微高出了 0.8%,证明了联合优化网络性能的优越性. 为了对伪标签进一步精细化,Nguyen 等人<sup>[57]</sup>在文献<sup>[55]</sup>的基础上将每类传统方法生成的伪标签送入一个带自我监督的迭代细化网络,再将细化后的伪标签融合为最终用于网络训练的伪标注数据集. 细化网络的加入极大地降低了伪标签中的噪声,在减轻各类伪标签缺陷的同时保留了不同方法所产生伪标签的多样性,并提高了网络的泛化能力.

为综合不同传统模型的优点,Quan 等人<sup>[58]</sup>设计了一种可以将已有的无监督模型结合起来的推理机制. 该方法首先采用一些现有的无监督显著性检测模型对输入图像生成弱显著性检测图,然后为每个模型定义一个标注精度参数,同时为输入图像的各区域定义一

个标注难度参数,以便同时考虑到各检测模型间的性能差异和各类图像区域上的特征差异,进而在设定的融合原则下将各模型生成的弱显著性检测图整合成强显著性检测图,最后利用局部空间一致性作为约束条件,对整合结果进行细化,得到最终的显著性目标检测结果.

采用伪标签作为训练样本的方法利用了已有的无监督显著性目标检测模型来提供所需的伪监督,在很大程度上降低了人工标注的代价,但这些方法的最终性能主要取决于所选用模型生成的伪标签质量,检测精度的提升受到了限制.

#### 2.3.2 基于弱标注数据集的方法

弱标注数据指的是相对于像素级标注来说,成本更低的标注,如给出如图 6(b)所示的图像级标注,或图 6(c)的图像描述文本,或图 6(d)所示的涂鸦标注. 弱标注数据集可以避免依赖传统方法生成伪标签所带来的限制,同时为显著性目标检测训练提供更可靠的依据.

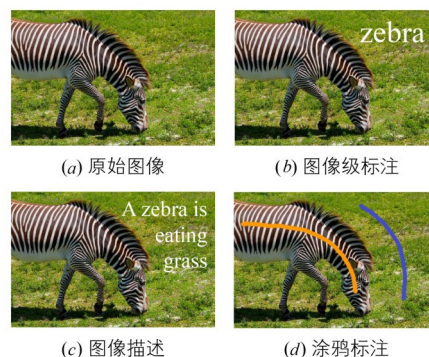


图 6 各类弱标注图像

Wang<sup>[59]</sup>等人使用如图 6(b)所示的图像级标签作为训练集的标注,并利用全卷积网络为图像标签中的对象类别分配相应的区域,然后通过由卷积层和二分图层组成的前景推理网络来捕获所有潜在的显著性对象区域. 然而,这些类别标签的信息量是有限的,它们难以帮助网络去详细地预测整个范围内的显著性映射. 相比之下,如图 6(c)所示的图像描述没有固定在图像的类别上,而是作为一个更加高级的全局概念,为网络提供丰富的显著性线索. 对此,Zhang 等人<sup>[60]</sup>使用 MSCOCO<sup>[61]</sup>和 SALICON<sup>[62]</sup>构建了一个以图像描述为标注的数据集 COCO-CapSal. 在该数据集上,CapSal<sup>[60]</sup>和 LAWS<sup>[63]</sup>利用长短时记忆网络(LSTM)网络来提取标题中潜在的上下文信息,并以此作为衡量语义内容的指标,再使用全卷积网络对输入图像进行显著性目标的推理和检测训练. 实验结果表明图像描述在复杂场景下能有效地保持检测性能,但由于图像描述通常也包含背景,这会导致检测结果产生偏差.

Zeng 等人<sup>[64]</sup>将图像级标签和图像描述共同作为标注样本进行训练,并针对这两种弱监督信息分别构建了分类网络和描述网络.分类网络用于显著性目标类别的预测,并标注出与分类结果相关的区域.描述网络用于生成对图像的描述并定位相应的区域.这两个网络通过一个转移损失函数使各自的监督信息可以在网络间进行传输,模型在这两个弱监督信息的协同训练中得到优化.

与上述探索自然语言与显著性对象之间联系的角度不同,Zhang 等人<sup>[65]</sup>使用如图 6(d)所示的涂鸦标注来学习图像中显著性目标的检测.在训练过程中,模型将图像的像素分为前景、背景和未知像素,首先使用一个基于 VGG16 的显著性预测网络中使用部分交叉熵损失函数对未知像素进行训练,获得初始的粗糙显著性映射,然后通过一个包含结构感知损失函数的边缘检测分支,补充图像中的边缘结构信息,从而得到优化后的检测结果.

### 3 数据集与评价指标

#### 3.1 数据集

MSRA<sup>[2]</sup>可以算作是第一个人工标注的显著性目标检测数据集,它包含了从各种图像论坛和图像搜索引擎收集的 20840 幅图像,图像的注释由边界框标记而成.MSRA10K<sup>[66]</sup>(也称为 THUS10K)数据集由 MSRA 中的 10000 张图像组成,并增加了像素级标注.由于其大规模和精确的注释,该数据集被广泛用于显著性目标检测模型的训练中.

SOD<sup>[67]</sup>由来自 BSD<sup>[68]</sup>的 300 张具有多个显著性对象的图像构成,这些对象与图像背景的颜色对比度大多都偏低,甚至还会与图像边界发生重叠,故该数据集具有非常高的挑战性,现被广泛用于显著性目标检测模型的评估中.

ECSSD<sup>[69]</sup>包含从 BSD<sup>[68]</sup>、VOC2012<sup>[70]</sup>数据集和互联网上选取的 1000 张具有复杂场景的图像,这些图像大多都只有一个显著性对象,其前景和背景信息都包含着多样化的图案.

PASCAL-S<sup>[71]</sup>由 850 幅来自 PASCAL VOC 2010<sup>[72]</sup>数据集上的图片组成,并在其原有的注释基础上增加了眼动注视点记录和显著性对象分割标记.

DUT-OMRON<sup>[73]</sup>包含 5168 张背景相对复杂且内容丰富的图片,每张图片都有一个或多个具有复杂背景的显著性对象,及与其相对应的像素级标注.

HKU-IS<sup>[25]</sup>包含 4447 张带有显著性物体像素级标注的图像,这些图像至少满足以下条件中的一个:存在多个不相连的显著性对象;至少有一个显著性对象触碰到所在图像的边界;颜色对比度小于 0.7.

DUTS<sup>[59]</sup>是最大的显著性目标检测数据集,分为含 10553 张图像的训练集和 5019 张图像的测试集.训练图像选自 ImageNet<sup>[74]</sup>训练或验证集,测试图像选自 ImageNet<sup>[74]</sup>测试集和 SUN<sup>[75]</sup>数据集,其精确的像素级标注由 50 名参与者手动标注而成.

#### 3.2 评估指标

##### 3.2.1 PR 曲线

精度(Precision,  $P$ )和召回率(Recall,  $R$ )也被称作查准率和查全率,是通过二值化下预测的显著性掩码和实际的显著性掩码计算得到的.

$$\begin{aligned} \text{Precision} &= \frac{TP}{TP + FP} \\ \text{Recall} &= \frac{TP}{TP + FN} \end{aligned} \quad (1)$$

式(1)中的 TP(True Positive)表示被模型预测为正确的正样本个数,FP(False Positive)表示被模型误判为正确的负样本个数,FN(False Negative)表示被模型误判为负的正样本个数,另外,TN(True Negative)表示的是被模型预测为负的正样本个数.

通过应用从 0 到 255 的阈值将显著性目标检测图二值化,每个阈值产生一组精度与召回值,从而描绘出呈现模型性能的 PR 曲线.通常而言,曲线越贴近右上方,模型性能就越好.

##### 3.2.2 F-measure

F-measure<sup>[76]</sup>综合考虑查准率和查全率,计算方法如式(2)所示:

$$F_{\beta} = \frac{(1 + \beta^2)\text{Precision} \times \text{Recall}}{\beta^2\text{Precision} + \text{Recall}} \quad (2)$$

式(2)中的  $\beta^2$  根据经验设置为 0.3,用以提升精度在评估中所占的比重.除了绘制 F-measure 曲线,人们往往会直接取其中的最大值作为模型的评价标准.

##### 3.2.3 平均绝对误差 MAE

检测图与真值图之间像素的平均绝对误差(Mean Absolute Error, MAE)计算方法如式(3)所示:

$$\text{MAE} = \frac{1}{W \times H} \sum_{x=1}^W \sum_{y=1}^H |S(x, y) - \text{GT}(x, y)| \quad (3)$$

式(3)中  $W$  和  $H$  分别表示图像的宽度和高度, $S(x, y)$  为模型检测得到的显著图, $\text{GT}(x, y)$  代表人工标注的真实显著图.一般地,MAE 的值越小,模型的性能也越好.

### 4 实验对比

为了从检测性能上直观地展示本文所论述的方法,表 1 给出了不同的显著性目标检测模型在 PASCAL-S<sup>[71]</sup>、ECSSD<sup>[69]</sup>和 HKU-IS<sup>[25]</sup>数据集上的性能对比,并使用 F-measure 和平均绝对误差作为评估指标,所有数据

均取自原始文献中的实验报告. 其中,“§”表示该项评估指标在对应的方法中没有具体的数据(通常以图的形式呈现);“-”表示该方法没有在对应的数据集上进行测试;“b”代表深度监督学习下的稀疏检测模型;“#”

代表弱监督模型.

观察表 1 中所有方法所使用的测试集可以发现,相较于其他数据集,MSRA10K 和 DUTS 在模型训练中被广泛使用,这在一定程度上体现了数据规模对于深度

表 1 不同方法在三个数据集上的性能对比

年份	模型名称	训练集	测试集					
			PASCAL-S		ECSSD		HKU-IS	
			F 最大值	MAE	F 最大值	MAE	F 最大值	MAE
2015	bLEGS <sup>[24]</sup>	MSRA-B +PASCAL-S	0.669	0.170	0.775	0.137	-	-
	bMCDL <sup>[20]</sup>	MSRA10K	0.793	§	0.732	§	-	-
2016	bLDHF <sup>[21]</sup>	MSRA10K	0.771	0.121	0.867	0.080	-	-
	SDCR <sup>[36]</sup>	MSRA10K	0.776	0.064	0.849	0.016	0.821	0.043
2017	DHSNet <sup>[42]</sup>	MSRA10K+DUT-O	0.799	0.810	0.893	0.060	-	-
	BR <sup>[34]</sup>	MSRA10K	0.805	0.182	0.793	0.160	-	-
2018	RACDNN <sup>[32]</sup>	DUT-OMRON	-	-	0.878	0.081	0.856	0.070
	DCL <sup>[37]</sup>	MSRA10K	0.822	0.108	-	-	0.904	0.049
2017	DLS <sup>[35]</sup>	MSRA10K	0.651	0.136	0.766	0.090	0.748	0.072
	UCF <sup>[29]</sup>	MSRA10K	0.741	0.117	0.852	0.069	0.823	0.062
	Amulet <sup>[43]</sup>	MSRA10K	0.763	0.098	0.868	0.059	0.854	0.052
	NLDF <sup>[40]</sup>	MSRA-B	0.804	0.116	0.886	0.075	0.874	0.060
	NLDF <sup>[40]</sup>	MSRA-B	0.831	0.099	0.905	0.063	0.902	0.048
	#WSS <sup>[59]</sup>	ImageNet	0.720	§	0.823	§	0.821	§
2018	#SbF <sup>[55]</sup>	MSRA10K	0.680	0.140	0.787	0.085	-	-
	RFCN <sup>[30]</sup>	MSRA10K	0.778	0.105	0.871	0.067	0.856	0.055
	PAGE <sup>[45]</sup>	DUTS	0.803	0.092	0.891	0.064	0.886	0.048
	RLN <sup>[31]</sup>	DUTS	-	-	0.903	0.045	0.882	0.037
	MCFEM <sup>[41]</sup>	DUTS	0.862	0.074	0.928	0.044	0.920	0.038
	PiCANetV <sup>[44]</sup>	DUTS	0.881	0.079	0.933	0.036	0.925	0.031
	PiCANetR <sup>[44]</sup>	DUTS	0.883	0.077	0.940	0.035	0.927	0.031
2019	#MNL <sup>[56]</sup>	MSRA-B	0.842	0.139	0.878	0.070	-	-
	BASNet <sup>[54]</sup>	DUTS	0.854	0.076	0.942	0.037	0.928	0.032
	AFNet <sup>[51]</sup>	DUTS	0.868	0.071	0.935	0.042	0.923	0.036
	PoolNet <sup>[49]</sup>	DUTS	0.880	0.065	0.945	0.038	0.935	0.030
	EGNet <sup>[52]</sup>	DUTS	0.869	0.074	0.943	0.041	0.937	0.031
	PFAN <sup>[47]</sup>	DUTS	0.892	0.068	0.931	0.033	0.926	0.032
	#FMN <sup>[63]</sup>	MSCOCO	0.741	0.119	0.831	0.088	0.821	0.067
	#CapSal <sup>[60]</sup>	COCO-CapSal <sup>[60]</sup>	0.823	0.075	-	-	0.836	0.059
	#DeepUSPS <sup>[57]</sup>	ImageNet	-	-	0.874	0.063	-	-
2020	#MSWS <sup>[64]</sup>	ImageNet+DUTS	0.790	0.134	0.878	0.096	-	-
	GCPANet <sup>[46]</sup>	DUTS	0.876	0.061	0.949	0.035	0.938	0.031
	DPNet <sup>[33]</sup>	DUTS	0.881	0.070	0.949	0.031	0.937	0.028
	PFPNetV <sup>[50]</sup>	DUTS	0.891	0.071	0.938	0.040	0.928	0.035
	PFPNetR <sup>[50]</sup>	DUTS	0.892	0.068	0.949	0.033	0.939	0.030
	DFNetV <sup>[48]</sup>	DUTS	0.866	0.075	0.933	0.040	0.921	0.033
	DFNetR <sup>[48]</sup>	DUTS	0.896	0.054	0.949	0.028	0.936	0.028
2020	#WSSA <sup>[65]</sup>	S-DUTS <sup>[65]</sup>	0.7884	0.140	0.865	0.061	0.858	0.047

神经网络训练的重要性。随着显著性目标检测领域的发展,研究者逐渐约定俗成地使用MSRA10K和DUTS作为训练集,然后将训练好的模型在ECSSD、PASCAL-S和HKU-IS等常用数据集上进行测试。

如表1所示,稀疏检测方法在2017年以后基本完全被密集检测方法所替代,这是由于稀疏检测方法通常以一个区域为计算单元,需要候选区域提取和区域到像素的映射等步骤,导致检测精度的提升受到限制。而密集检测方法以整幅图像为输入,为每个像素进行显著性的判断,实现像素级预测的同时大幅度提升了显著性目标检测的精度。与同期监督学习下的深度神经网络模型相比,弱监督模型减少了对人工标注数据的依赖,但提取有效特征的能力也相应地下降了,因此其检测性能普遍较深度监督模型的更低。然而,有些弱监督方法可以追平甚至超过同时期部分深度监督方法的性能。如2017年提出的WSS<sup>[59]</sup>在各个数据集上的F最大值指标均超过了同年的深度监督模型DLS<sup>[35]</sup>,这是由于DLS<sup>[35]</sup>模型在训练过程中稀释了高级特征,同时还引入了超像素分割所携带的噪声。2018年的MN-LP<sup>[56]</sup>的性能指标也比同期的RFCN<sup>[30]</sup>略微高些。此外,从2019年至今的研究工作中,暂未出现性能超越监督学习方法的弱监督模型,但由于弱监督学习方法能在大幅度降低人工成本的同时仍保持较高的性能,依然值得对其开展更多的相关研究工作。总体上,显著性目标检测模型的性能随着时间的推移逐渐提高,这展示了显著性计算模型的发展。

## 5 总结与展望

本文首先对深度监督学习下的显著性目标检测模型进行了综述,并从检测粒度的角度分类综述了这些方法间的联系及各自的优缺点。然后根据训练图像中标注信息来源的不同,详细阐述了弱监督学习下的网络模型。另外,本文介绍了七个主流数据集和三个主流评估基准,并从中选取了三个最为广泛使用的测试集和两个便于数据直观对比的评估指标,通过从原始文献中采集实验数据来进行模型间的对比分析。

与传统方法相比,现今显著性目标检测模型的性能已取得显著的提升,但仍然存在许多的挑战。新的研究趋势包括:在特征提取方面,如何学习到更能表达显著性语义概念的特征;利用多任务学习,通过执行多个计算机视觉任务(如目标检测、语义分割等)产生丰富的信息,进而大幅度提高单独训练模型的性能。弱监督学习下的显著性目标检测模型如何融合多源信息进行训练也是非常值得研究的方向。随着社交媒体的普及和大数据技术的发展,研究者很容易获取多源信息,如很多社交媒体信息可以提供图片和文字形式的描

述,这可以为弱监督学习下的检测模型提供更多无需人工标注的训练样本。

## 参考文献

- [1] Itti L, Koch C, Niebur E. A model of saliency-based visual attention for rapid scene analysis[J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 1998, 20(11): 1254 – 1259.
- [2] Liu T, Zheng N-N, Sun J, et al. Learning to detect a salient object[A]. *IEEE Conference on Computer Vision and Pattern Recognition*[C]. Minneapolis, MN, USA: IEEE, 2007. 1 – 8.
- [3] Fang H G S, Iandola F, et al. From captions to visual concepts and back[A]. *IEEE Conference on Computer Vision and Pattern Recognition*[C]. Boston, Massachusetts, USA: IEEE, 2015. 1473 – 1482.
- [4] K X, J B, R K. Show, attend and tell: Neural image caption generation with visual attention[A]. *International Conference on Machine Learning*[C]. Lille, France: International Machine Learning Society, 2015. 2048 – 2057.
- [5] Zhang F, Du B, Zhang L. Saliency-guided unsupervised feature learning for scene classification[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2015, 53(4): 2175 – 2184.
- [6] Zhu J-Y, Wu J, Xu Y, et al. Unsupervised object class discovery via saliency-guided multiple class learning[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014, 37(4): 862 – 875.
- [7] Wang W, Shen J, Yang R, et al. Saliency-aware video object segmentation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 40(1): 20 – 33.
- [8] 毕威,黄伟国,张永萍,等. 基于图像显著轮廓的目标检测[J]. *电子学报*, 2017, 45(8): 1902 – 1910.  
BI Wei, HUANG Wei-guo, ZHANG Yong-ping, et al. Object detection based on salient contour of image[J]. *Acta Electronica Sinica*, 2017, 45(8): 1902 – 1910. (in Chinese)
- [9] Das A A H, Zitnick L, et al. Human attention in visual question answering: Do humans and deep networks look at the same regions?[J]. *Computer Vision and Image Understanding*, 2017, 163: 90 – 100.
- [10] Borji A, Cheng M-M, Jiang H, et al. Salient object detection: a benchmark[J]. *IEEE Transactions on Image Processing*, 2015, 24(12): 5706 – 5722.
- [11] Borji A, Itti L. State-of-the-art in visual attention modeling [J]. *IEEE Trans Pattern Anal Mach Intell*, 2013, 35(1): 185 – 207.
- [12] Kumar N. Thresholding in salient object detection: a sur-

- vey[J]. *Multimedia Tools and Applications*, 2018, 77(15): 19139 – 19170.
- [13] Han J, Zhang D, Cheng G, et al. Advanced deep-learning techniques for salient and category-specific object detection: a survey[J]. *IEEE Signal Processing Magazine*, 2018, 35(1): 84 – 100.
- [14] Borji A, Cheng M-M, Hou Q, et al. Salient object detection: a survey[J]. *Computational Visual Media*, 2019, 5(2): 117 – 150.
- [15] 吴加莹,杨赛,堵俊,等. 自底向上的显著性目标检测研究综述[J]. *计算机科学*, 2019, 46(03): 48 – 52.  
WU Jia-ying, YANG Sai, DU Jun, et al. Review of bottom-up salient object detection[J]. *Chinese Journal of Computers*, 2019, 46(03): 48 – 52. (in Chinese)
- [16] Wang W, Lai Q, Fu H, et al. Salient Object Detection in the Deep Learning Era: An In-depth Survey[EB/OL]. <http://arXiv.org/abs/1904.09146>, 2019.
- [17] Yuan Y, Li C, Kim J, et al. Dense and sparse labeling with multidimensional features for saliency detection[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2016, 28(5): 1130 – 1143.
- [18] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[A]. *IEEE Conference on Computer Vision and Pattern Recognition*[C]. Boston, Massachusetts, USA: IEEE, 2015. 3431 – 3440.
- [19] He S, Lau R W H, Liu W, et al. Supercnn: a superpixel-wise convolutional neural network for salient object detection[J]. *International Journal of Computer Vision*, 2015, 115(3): 330 – 344.
- [20] Zhao R, Ouyang W, Li H, et al. Saliency detection by multi-context deep learning[A]. *IEEE Conference on Computer Vision and Pattern Recognition*[C]. Boston, Massachusetts, USA: IEEE, 2015. 1265 – 1274.
- [21] Lee G T Y W, Kim J. Deep saliency with encoded low level distance map and high level features[A]. *IEEE Conference on Computer Vision and Pattern Recognition*[C]. Las Vegas, NV, USA: IEEE, 2016. 660 – 668.
- [22] Achanta R, Shaji A, Smith K, et al. SLIC superpixels compared to state-of-the-art superpixel methods[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, 34(11): 2274 – 2282.
- [23] Kim J, Pavlovic V. A shape preserving approach for salient object detection using convolutional neural networks [A]. *International Conference on Pattern Recognition*[C]. Cancun, Mexico: IEEE, 2016. 609 – 614.
- [24] Wang L, Lu H, Ruan X, et al. Deep networks for saliency detection via local estimation and global search[A]. *IEEE Conference on Computer Vision and Pattern Recognition* [C]. Boston, Massachusetts, USA: IEEE, 2015. 3183 – 3192.
- [25] Li G, Yu Y. Visual saliency based on multiscale deep features[A]. *IEEE Conference on Computer Vision and Pattern Recognition*[C]. Boston, Massachusetts, USA: IEEE, 2015. 5455 – 5463.
- [26] Zhang J, Sclaroff S, Lin Z, et al. Unconstrained salient object detection via proposal subset optimization[A]. *IEEE Conference on Computer Vision and Pattern Recognition* [C]. Las Vegas, NV, USA: IEEE, 2016. 5733 – 5742.
- [27] Uijlings J R, Van De Sande K E, Gevers T, et al. Selective search for object recognition[J]. *International Journal of Computer Vision*, 2013, 104(2): 154 – 171.
- [28] Krähenbühl P, Koltun V. Geodesic object proposals[A]. *European Conference on Computer Vision*[C]. Zurich, Switzerland: Springer, 2014. 725 – 739.
- [29] Zhang P, Wang D, Lu H, et al. Learning uncertain convolutional features for accurate saliency detection[A]. *IEEE International Conference on Computer Vision*[C]. Honolulu, Hawaii, USA: IEEE, 2017. 212 – 221.
- [30] Wang L, Wang L, Lu H, et al. Salient object detection with recurrent fully convolutional networks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, 41(7): 1734 – 1746.
- [31] Wang T, Zhang L, Wang S, et al. Detect globally, refine locally: a novel approach to saliency detection[A]. *IEEE Conference on Computer Vision and Pattern Recognition* [C]. Salt Lake City, Utah, USA: IEEE, 2018. 3127 – 3135.
- [32] Kuen J, Wang Z, Wang G. Recurrent attentional networks for saliency detection[A]. *IEEE Conference on Computer Vision and Pattern Recognition*[C]. Las Vegas, NV, USA: IEEE, 2016. 3668 – 3677.
- [33] Xu X, Chen J, Zhang H, et al. Dual pyramid network for salient object detection[J]. *Neurocomputing*, 2020, 375: 113 – 123.
- [34] Li X, Zhao L, Wei L, et al. Deepsaliency: multi-task deep neural network model for salient object detection[J]. *IEEE Trans Image Process*, 2016, 25(8): 3919 – 3930.
- [35] Hu P, Shuai B, Liu J, et al. Deep level sets for salient object detection[A]. *IEEE Conference on Computer Vision and Pattern Recognition*[C]. Honolulu, HI, USA: IEEE, 2017. 2300 – 2309.
- [36] Tang Y, Wu X. Saliency detection via combining region-level and pixel-level predictions with cnns[A]. *European Conference on Computer Vision*[C]. Amsterdam, The

- Netherlands: Springer, 2016. 809 – 825.
- [37] Li G Y Y. Deep contrast learning for salient object detection [A]. IEEE Conference on Computer Vision and Pattern Recognition[C]. Las Vegas, NV, USA: IEEE, 2016. 478 – 487.
- [38] Krähenbühl P, Koltun V. Efficient inference in fully connected crfs with gaussian edge potentials[A]. Advances in Neural Information Processing Systems[C]. Granada, Andalusia, Spain: Curran Associates, Inc, 2011. 109 – 117.
- [39] Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation[A]. International Conference on Medical Image Computing and Computer-assisted Intervention[C]. Munich, Germany: Springer, 2015. 234 – 241.
- [40] Zhiming Luo A M, Andrew Achkar, Justin Eichel, Shaozi Li, and Pierre-Marc Jodoin. Non-local deep features for salient object detection[A]. IEEE Conference on Computer Vision and Pattern Recognition[C]. Honolulu, HI, USA: IEEE, 2017. 6609 – 6617.
- [41] Zhang L, Dai J, Lu H, et al. A bi-directional message passing model for salient object detection[A]. IEEE Conference on Computer Vision and Pattern Recognition[C]. Salt Lake City, Utah, USA: IEEE, 2018. 1741 – 1750.
- [42] Liu N, Han J. Dhsnet: deep hierarchical saliency network for salient object detection[A]. IEEE Conference on Computer Vision and Pattern Recognition[C]. Las Vegas, NV, USA: IEEE, 2016. 678 – 686.
- [43] Zhang P, Wang D, Lu H, et al. Amulet: aggregating multi-level convolutional features for salient object detection[A]. IEEE International Conference on Computer Vision[C]. Honolulu, Hawaii, USA: IEEE, 2017. 202 – 211.
- [44] Liu N, Han J, Yang M-H. Picanet: learning pixel-wise contextual attention for saliency detection[A]. IEEE Conference on Computer Vision and Pattern Recognition[C]. Salt Lake City, Utah, USA: IEEE, 2018. 3089 – 3098.
- [45] Zhang X, Wang T, Qi J, et al. Progressive attention guided recurrent network for salient object detection[A]. IEEE Conference on Computer Vision and Pattern Recognition [C]. Salt Lake City, Utah, USA: IEEE, 2018. 714 – 722.
- [46] Chen Z, Xu Q, Cong R, et al. Global context-aware progressive aggregation network for salient object detection [A]. Association for the Advancement of Artificial Intelligence[C]. New York, USA: AAAI, 2020.
- [47] Zhao T, Wu X. Pyramid feature attention network for saliency detection[A]. IEEE Conference on Computer Vision and Pattern Recognition[C]. Long Beach, California, USA: IEEE, 2019. 3085 – 3094.
- [48] Noori M, Mohammadi S, Majelan S G, et al. DFNet: discriminative feature extraction and integration network for salient object detection[J]. Engineering Applications of Artificial Intelligence, 2020, 89: 103419.
- [49] Liu J-J, Hou Q, Cheng M-M, et al. A simple pooling-based design for real-time salient object detection[A]. IEEE Conference on Computer Vision and Pattern Recognition[C]. Long Beach, California, USA: IEEE, 2019.
- [50] Wang B, Chen Q, Zhou M, et al. Progressive feature polishing network for salient object detection.[A]. Association for the Advancement of Artificial Intelligence[C]. New York, USA: AAAI, 2020.
- [51] Feng M, Lu H, Ding E. Attentive feedback network for boundary-aware salient object detection[A]. IEEE Conference on Computer Vision and Pattern Recognition[C]. Long Beach, California, USA: IEEE, 2019. 1623 – 1632.
- [52] Zhao J-X, Liu J, Fan D-P, et al. EGNet: edge guidance network for salient object detection[A]. IEEE International Conference on Computer Vision[C]. Seoul, Korea: IEEE, 2019.
- [53] Han L, Li X, Dong Y. Convolutional edge constraint-based U-Net for salient object detection[J]. IEEE Access, 2019, 7: 48890 – 48900.
- [54] Qin X, Zhang Z, Huang C, et al. Basnet: boundary-aware salient object detection[A]. IEEE Conference on Computer Vision and Pattern Recognition[C]. Long Beach, California, USA: IEEE, 2019. 7479 – 7489.
- [55] Zhang D, Han J, Zhang Y. Supervision by fusion: towards unsupervised learning of deep salient object detector[A]. IEEE International Conference on Computer Vision[C]. Honolulu, Hawaii, USA: IEEE, 2017. 4048 – 4056.
- [56] Zhang J, Zhang T, Dai Y, et al. Deep unsupervised saliency detection: a multiple noisy labeling perspective[A]. IEEE Conference on Computer Vision and Pattern Recognition [C]. Salt Lake City, Utah, USA: IEEE, 2018. 9029 – 9038.
- [57] Nguyen T, Dax M, Mummadi C K, et al. Deepusps: deep robust unsupervised saliency prediction via self-supervision[A]. Advances in Neural Information Processing Systems[C]. Vancouver, Canada: Morgan Kaufmann Publishers Inc., 2019. 204 – 214.
- [58] Quan R, Han J, Zhang D, et al. Unsupervised salient object detection via inferring from imperfect saliency models[J]. IEEE Transactions on Multimedia, 2017, 20(5): 1101 – 1112.
- [59] Wang L, Lu H, Wang Y, et al. Learning to detect salient objects with image-level supervision[A]. IEEE Conference on Computer Vision and Pattern Recognition[C]. Honolulu, HI, USA: IEEE, 2017. 136 – 145.

- [60] Zhang L, Zhang J, Lin Z, et al. Capsal: leveraging captioning to boost semantics for salient object detection[A]. IEEE Conference on Computer Vision and Pattern Recognition[C]. Long Beach, California, USA: IEEE, 2019. 6024 – 6033.
- [61] Lin T-Y, Maire M, Belongie S, et al. Microsoft coco: common objects in context[A]. European Conference on Computer Vision[C]. Zurich, Switzerland: Springer, 2014. 740 – 755.
- [62] Jiang M, Huang S, Duan J, et al. Salicon: saliency in context[A]. IEEE Conference on Computer Vision and Pattern Recognition[C]. Boston, Massachusetts, USA: IEEE, 2015. 1072 – 1080.
- [63] Qian M, Qi J, Zhang L, et al. Language-aware weak supervision for salient object detection[J]. Pattern Recognition, 2019, 96: 106955.
- [64] Zeng Y, Zhuge Y, Lu H, et al. Multi-source weak supervision for saliency detection[A]. IEEE Conference on Computer Vision and Pattern Recognition[C]. Long Beach, California, USA: IEEE, 2019. 6074 – 6083.
- [65] Zhang J, Yu X, Li A, et al. Weakly-supervised salient object detection via scribble annotations[A]. IEEE Conference on Computer Vision and Pattern Recognition[C]. Seattle, USA: IEEE, 2020. 12546 – 12555.
- [66] Cheng M-M, Zhang G-X, Mitra N J, et al. Global contrast based salient region detection[A]. IEEE Conference on Computer Vision and Pattern Recognition[C]. NW Washington, DC, USA: IEEE, 2011. 409 – 416.
- [67] Movahedi V, Elder J H. Design and perceptual validation of performance measures for salient object segmentation [A]. IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops[C]. San Francisco, CA, USA: IEEE, 2010. 49 – 56.
- [68] Martin D, Fowlkes C, Tal D, et al. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics[A]. Proceedings of the 8th IEEE International Conference on Computer Vision[C]. Vancouver, BC, Canada: IEEE, 2001. 416 – 423.
- [69] Yan Q, Xu L, Shi J, et al. Hierarchical saliency detection [A]. IEEE Conference on Computer Vision and Pattern Recognition[C]. Portland, Oregon: IEEE, 2013. 1155 – 1162.
- [70] Everingham M, Winn J. The PASCAL Visual Object Classes Challenge 2012 (VOC2012) Development Kit[M]. UK: The VOC Workshop, 2012.
- [71] Li Y, Hou X, Koch C, et al. The secrets of salient object segmentation[A]. IEEE Conference on Computer Vision and Pattern Recognition[C]. Columbus, Ohio, USA: IEEE, 2014. 280 – 287.
- [72] Everingham M, Gool L V, Williams C K I, et al. The pascal visual object classes (voc) challenge[J]. International Journal of Computer Vision, 2010, 88(2): 303 – 338.
- [73] Yang C, Zhang L, Lu H, et al. Saliency detection via graph-based manifold ranking[A]. IEEE Conference on Computer Vision and Pattern Recognition[C]. Portland, Oregon: IEEE, 2013. 3166 – 3173.
- [74] Deng J, Dong W, Socher R, et al. Imagenet: a large-scale hierarchical image database[A]. IEEE Conference on Computer Vision and Pattern Recognition[C]. Miami, FL, USA: IEEE, 2009. 248 – 255.
- [75] Xiao J, Hays J, Ehinger K A, et al. Sun database: large-scale scene recognition from abbey to zoo[A]. IEEE Computer Society Conference on Computer Vision and Pattern Recognition[C]. San Francisco, CA, USA: IEEE, 2010. 3485 – 3492.
- [76] Achanta R, Hemami S, Estrada F, et al. Frequency-tuned salient region detection[A]. IEEE Conference on Computer Vision and Pattern Recognition[C]. Miami, FL, USA: IEEE, 2009. 1597 – 1604.

#### 作者简介



罗会兰 女, 1974年9月生于江西上高. 2008年获浙江大学工学博士学位. 现为江西理工大学图像处理实验室教授、硕士生导师. 主要从事机器学习、模式识别等方面的研究.

E-mail: luohuilan@sina.com



袁璞 女, 1997年5月生于江西吉安. 2018年进入江西理工大学. 在读硕士研究生, 研究方向为图像修复、显著性目标检测.

E-mail: emmanuel\_97@163.com



童康 男, 1992年生于江苏南京, 博士研究生, 主要研究方向为计算机视觉与模式识别.

E-mail: tkangev@nuaa.edu.cn