

基于深度学习的图像显著性目标检测研究综述

李元贞¹, 赵俊松²

(1.河北工业大学人工智能与数据科学学院, 天津 300401;

2.中国民航大学信息网络中心, 天津 300300)

✉2081059521@qq.com; xmbdhyq@163.com



摘要: 随着近年来深度学习技术的发展, 图像显著性目标检测的研究重点偏向于利用深度学习方法解决问题。为了全面且深入地探究图像显著性目标检测领域, 基于深度学习框架, 回顾近五年出现的20余种深度学习方法, 归纳出四类深度学习策略, 并且对比了它们在4个显著性数据集上的评价结果。结果显示, 各类策略在不同数据集上的 F 度量值为0.800—0.950, 综合利用多种策略的方法可以取得更优的预测指标, 但仍然存在复杂场景干扰下检测有误的问题。针对现有问题, 提出加强深度学习方法在复杂数据集上的训练, 进而优化显著目标预测结果的定位准确性及边缘完整性。

关键词: 图像显著性目标检测; 深度学习框架; 深度学习策略

中图分类号: TP391 **文献标识码:** A

Overview of Image Salient Object Detection Research based on Deep Learning

LI Yuanzhen¹, ZHAO Junsong²

(1.School of Artificial Intelligence, Hebei University of Technology, Tianjin 300401, China;

2.Information Network Center, Civil Aviation University of China, Tianjin 300300, China)

✉2081059521@qq.com; xmbdhyq@163.com

Abstract: With the development of deep learning in recent years, research focus of image salient object detection tends to use deep learning methods to solve problems. In order to comprehensively and deeply explore the field of image salient object detection, this paper proposes to review more than 20 deep learning methods over the past five years based on deep learning framework, summarize four types of deep learning strategies, and compare their evaluation results on four datasets of salient object detection. The results show that F -measure of various strategies on different datasets are from 0.800 to 0.950. The comprehensive use of multiple strategies can achieve better prediction indicators, but there are still the problems of incorrect detection under the interference of complex scenes. In view of the existing problems, it is proposed to strengthen the training of deep learning methods on complex datasets, so as to optimize the positioning accuracy and edge integrity of salient object.

Keywords: image salient object detection; deep learning framework; deep learning strategies

1 引言(Introduction)

图像显著性目标检测旨在模拟人类视觉机制, 识别并分割出场景中人类感兴趣的部分, 常被作为其他视觉领域的预处理阶段, 应用于风格迁移^[1]等。早期, 基于深度学习的显著性目标检测为多层感知机的形式, 造成卷积后的特征输出时再次经过全连接层, 不适用于像素级别的显著性任务。为此, 大量研究者开始利用全卷积网络去除全连接层, 从而获得更加精确的结果。然而, 全卷积网络仍然存在显著目标检测边缘缺失、定位有误等问题。因此, 本文总结了不同问题的针对性解决方法及各自的作用原理, 从深度学习的网络框

架入手, 探究应用于上述网络各类深度学习策略的优点和缺点, 分析目前深度学习方法在图像显著性目标检测领域的评价结果, 并总结提出基于深度学习的图像显著性目标检测任务亟待解决的问题。

2 深度学习框架(Framework of deep learning)

单流单层输出结构在网络最深层输出预测结果, 该种架构简单, 但是对网络最后一层的特征样式依赖性较强, 随着网络层级的传递, 较浅层的一些细节信息会缺失, 预测的结果将变成一个模糊的显著区域, 而不是精准的显著目标。为此, 演变出一种单流多层融合输出的结构, 利用跳层连接的

方式^[2],打破特征在网络层级间的串联传递,融合各层级的特征和性能,可以使最终的预测结果兼顾细节与整体,很多基于多层级特征融合的策略,都源于此。多流网络则是利用多个分支,不同分支关注图像中不同类型的特征,特征间互为补充,以此兼顾多种特征性能,如边缘检测^[3]或RGB-D深度图^[4]。与多流网络不同,“U”形结构仅有一个输入,但是分为上采样阶段和下采样阶段,该网络架构的可扩展性强,是目前应用最广的架构,如特征金字塔网络(Feature Pyramid Networks, FPN)^[5],结合网络深浅层特征的同时,适应图像中目标的尺度变化。细化网络^[6]则是“U”形网络的进一步演变,通过堆叠多个上采样与下采样,逐渐优化网络学习过程中的特征表达,该方法也适用于优化显著目标的边缘细节,但是需要多次尝试,选出最优的迭代次数,并且网络的规模也相对较大。对应网络架构所适用的深度学习策略,详见表1。

表1 显著性目标检测经典架构

Tab.1 Classic architectures of salient object detection

网络框架类型	适用策略
单流单层输出	多层感知机
单流多层融合输出	多层级、多尺度融合
多流网络	边缘检测分支、RGB-D
“U”形结构	FPN、多层级、多尺度融合
细化网络	“U”形结构演变

3 深度学习策略(Deep learning strategies)

图像显著性目标检测任务期待获取定位准确、边缘细节完整及内部均匀、高亮的目标。因此,通常利用边缘检测分支、细节优化网络或条件随机场(Conditional Random Field, CRF)对目标进行边缘特征增强;利用多尺度或多层级的特征融合策略,适应不同感受野范围的特征图像;利用通道注意力、空间注意力等注意力机制,加强显著目标的定位能力。除此之外,考虑设计更适用于显著性目标检测任务的损失函数,需要加强对网络训练过程中的监督,优化预测结果。

3.1 边缘增强

基于深度学习的图像显著性目标检测方法,随着卷积神经网络的层级加深,特征图尺寸会不断缩小,造成显著目标所包含的边缘细节缺失。为此,SU等^[7]基于选择性和不变性分析,将边缘检测和显著性定位问题拆分解,用边缘分支弥补显著分支的边缘细节。SONG等^[8]提出显著性检测的分层边缘细化网络,利用改进的“U”形结构粗略检测显著目标的区域,然后利用空间金字塔池化检测显著目标的边缘,一对一分层监督整个网络的不同输出。WANG等^[9]提出利用金字塔增强模块和显著性边缘检测模块获取边缘增强的显著性结果,通过进一步结合条件随机场,有效提升预测精度。虽然利用边缘检测分支、细化网络、条件随机场的方法能够缓解深度学习中网络细节信息缺失的问题,但是上述方法最大的弊端在于网络模型的运行规模过大,并且在复杂纹理场景下容易造成过度分割,使显著目标内部的均匀性受到影响。

3.2 特征融合

深度学习网络中不同层级的特征具有不同的特性,较深层的特征图分辨率较低,可以用来定位显著目标,较浅层的特征图分辨率较高,可以保留一些局部细节。因此,融合多个层级的特征有利于优化显著性检测结果。DAKHIA等^[10]提出通过结合不同层级局部信息和全局金字塔池实现多个特征的集成,有效聚合全局上下文信息,提高预测图质量。除了利用不同层之间的特征融合,还进一步延伸出单一层级的多尺度特征融合,以此适应不同感受野范围的显著目标。陈琴等^[11]提出利用FPN网络获取不同层级的特征,与此同时利用扩张卷积建立中心邻域金字塔,捕获不同级别的对比度信息。王正文等^[12]提出基于语义引导特征聚合的显著性目标检测网络,设计了包含不同空洞率的空洞卷积模块,以此得到不同尺寸大小的特征,融合它们增强语义信息;同时,设计多层级聚合模块,以级联的方式不断提取特征中的显著部分,细化显著物体的边缘细节。不同尺度和不同层级的特征融合,在某种程度上增强了特征的表达。然而,融合的特征是否需要,融合特征之间的差异性是否较大的问题仍然存在,不经过信息筛选的特征堆叠,反而会引入噪声。

3.3 注意力机制

视觉注意力机制在图像领域有着广泛的应用,很多显著性检测方法也通过加入注意力模块抑制背景噪声,突出显著目标。陈维婧等^[13]提出将改进的通道注意力机制与空间注意力机制并行结合,输出的两个注意力特征加权融合再反馈至通道-空间联合注意力机制中,从而得到细粒度更高的显著图。LIANG等^[14]提出双重注意中间表示模块和高级通道注意模块,前者使用两个分支自适应地整合中间层特征的空间和语义信息,后者通过两种不同的通道操作获得高层的语义特征,从而针对不同层级提取不同的注意力。王凯诚等^[15]提出了注意力精炼模块,对通道维度整合的注意力进行非局部的空间操作,并且用显著真值监督空间操作后的特征,增强预测位置的准确性。作为显著性目标检测任务的重要策略,注意力机制可以有效定位显著目标,但是针对低对比度、前景背景杂乱等复杂场景,则容易受到干扰。此外,注意力机制往往包含大量参数,会造成网络运行规模较大。

3.4 损失监督

图像显著性目标检测任务中主要采用交叉熵损失函数,如公式(1)所示。利用显著性真值图像 $G(x, y)$ 与预测图像 $P(x, y)$ 计算得到,预测结果与真值越相近,损失值越小。

$$L_{bce} = \sum_{p \in P(x, y), g \in G(x, y)} [-g \log p + (1 - g) \log (1 - p)] \quad (1)$$

但是,传统的交叉熵损失函数对图像中的每一个像素点是同等看待的,而针对显著性问题,前景目标及边缘区域应当受到更多的关注。为此,FENG等^[16]提出带边缘权重的交叉熵损失函数,使处于目标边缘位置的像素特征更具判别性;WEI等^[17]在利用边缘权重损失函数的同时,提出区分前景目标与背景区域,从而关注显著目标,抑制背景噪声。尽管如此,交叉熵损失还是逐像素级别的损失计算,不符合人类从整体结构角度观察物体的习惯。为此,LIU等^[18]提出结合交并

比(Intersection over Union, IoU)损失关注显著目标的结构特性,如公式(2)所示。

$$L_{IoU} = 1 - \frac{\sum_{p \in P(x,y), g \in G(x,y)} p \cdot g}{\sum_{p \in P(x,y), g \in G(x,y)} p + g - p \cdot g} \quad (2)$$

目前,大多数图像显著性监督都是在交叉熵损失与IoU损失的基础上,设计各类权重,区分前景与背景,通过整合得到一个新的损失。除此之外,研究者对监督方式的选择也有所不同。CHEN等^[19]对网络不同层级的特征采取同时监督。WU等^[20]提出下采样阶段对浅层网络采取边缘真值的监督,对深层网络采取显著性真值的监督,此外利用上述两种真值对上采样阶段的特征采取一种损失交替的监督方式,突出前景轮廓检测且产生均匀高亮的显著图。与上述两种类型对多个特征进行监督不同,陈家祯等^[21]仅对网络输出层的结果进行监督,从总体上看,多个损失监督的方式相比仅对结果监督的方式,可以在一定程度上提升训练结果的性能。但是,某些特定属性的数据集真值监督,对于现实生活场景的复杂性与多样性的满足程度远远不够,为了推进图像显著性在实际生活中的应用,也有考虑用弱监督的方式作用于显著性任务,ZHANG等^[22]就提出利用辅助的边缘检测任务及门控结构的感知损失构造出伪标签,将标签交替迭代地用于网络训练中,即一边训练网络完成预测,一边用预测结果作为新的真值训练网络。

4 深度学习在图像显著性目标检测领域的成效 (Effect of deep learning in image salient object detection)

为了进一步了解基于深度学习的显著性目标检测研究现状,本文利用柱状图对前文提及的代表性研究方法加以示意,详见图1,纵坐标的数值表示 F 度量值的大小,取值范围为0.000—1.000,并且数值越大,表明预测效果越好,不同灰度柱状图的对应研究方法名称在图下方标注,其中BANet和ITBINet是利用边缘检测分支和细化网络的边缘增强策略,MAFNet和GCPANet是利用特征融合策略,PAGE是重点利用注意力机制,MLM为关注损失的策略,F3Net则同时兼顾边缘增强、特征融合与损失。本文展示了DUTS-TE、ECSSD、HKU-IS和PASCAL-S四个广泛使用的显著性数据集结果,首先单独观察每个子图的数据集,柱状条的高度较为平均且数值均在0.800—0.950,表明目前基于深度学习的显著性目标检测的研究水平趋于一致且均能达到一定的效果,有利于显著目标的准确定位与边缘完整性。其次通过横向对比不同数据集的 F 度量值发现,DUTS-TE和PASCAL-S数据集的整体指标值要略低于其他两个数据集,这是因为DUTS-TE和PASCAL-S数据集相较于其他数据集包含更加复杂的前景和背景,对显著目标的识别具有一定的挑战性。因此,基于深度学习的显著性目标检测目前主要面临的问题是复杂自然场景下的目标识别。此外,F3Net方法在多个数据集上都能取得优秀的表现,表明在网络特征提取的不同阶段,综合使用相应的深度学习策略,可以有效增强特征表达,提升预测结果的精准性。

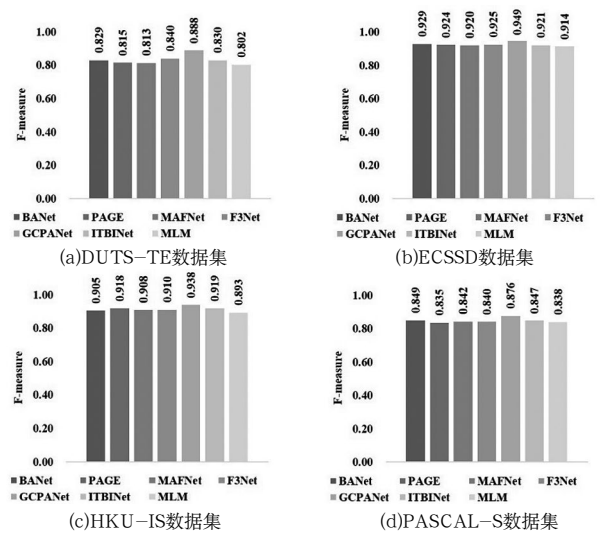


图1 4个数据集的 F 度量值柱状图

Fig.1 Histograms of F -measure on 4 datasets

5 结论(Conclusion)

本文围绕深度学习,对20余种显著性目标检测方法展开了研究论述。首先,归纳了5种深度学习框架;其次,探究依托于深度学习架构上的4种深度学习策略,分析各自的作用原理与优点、缺点。研究表明,兼顾整体与细节的方法更能满足定位准确、边缘完整的显著性目标检测要求。通过进一步分析在DUTS-TE、ECSSD、HKU-IS和PASCAL-S数据集上的 F 度量指标发现,目前的深度学习方法对纹理清晰、目标单一的显著性数据集已经取得了较好的检测效果,但是针对包含复杂前景背景的数据集,检测会受到一定的干扰。未来将重点解决复杂场景下的显著目标识别,构建自然场景下的复杂图像数据集,并且加强在复杂数据集上的深度学习训练。

参考文献(References)

- [1] 王杨,郁振鑫,卢嘉.融合视觉注意机制的图像显著性区域风格迁移方法[J].计算机工程与科学,2022,44(01):118-123.
- [2] HOU Q B, CHENG M M, HU X W, et al. Deeply supervised salient object detection with short connections[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 41(4):815-828.
- [3] 张冬明,新国庆,代锋,等.基于深度学习的显著性目标检测算法[J].计算机学报,2019,42(09):2076-2086.
- [4] XU Y Q, YU X, ZHANG J, et al. Weakly supervised rgb-d salient object detection with prediction consistency training and active scribble boosting[J]. IEEE Transactions on Image Processing, 2022, 31:2148-2161.
- [5] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]// Institute of Electrical and Electronics Engineers. Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017:936-944.
- [6] WANG W G, SHEN J B, CHENG M M, et al. An iterative and cooperative top-down and bottom-up inference network for salient object detection[C]// Institute of Electrical and Electronics Engineers. Proceedings of 2019 IEEE/CVF

- Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2019:5961–5970.
- [7] SU J M, LI J, ZHANG Y, et al. Selectivity or invariance: boundary-aware salient object detection[C]// Institute of Electrical and Electronics Engineers. Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2019:3798–3807.
- [8] SONG D W, DONG Y S, LI X L. Hierarchical edge refinement network for saliency detection[J]. IEEE Transactions on Image Processing, 2021, 30:7567–7577.
- [9] WANG W G, ZHAO S H, SHEN J B, et al. Salient object detection with pyramid attention and salient edges[C]// Institute of Electrical and Electronics Engineers. Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2019:1448–1457.
- [10] DAKHIA A, WANG T T, LU H C. Multi-scale pyramid pooling network for salient object detection[J]. Neurocomputing, 2019, 333:211–220.
- [11] 陈琴,朱磊,后云龙,等.基于深度中心邻域金字塔结构的显著目标检测[J].模式识别与人工智能,2020,33(06):496–506.
- [12] 王正文,宋慧慧,樊佳庆,等.基于语义引导特征聚合的显著性目标检测网络[J/OL].自动化学报:1–10[2021–12–22]. <https://doi.org/10.16383/j.aas.c210425>.
- [13] 陈维婧,周萍,杨海燕,等.通道-空间联合注意力机制的显著性检测模型[J].计算机工程与应用,2021,57(19):214–219.
- [14] LIANG Y H, QIN G H, SUN M H, et al. MAFNet: multi-style attention fusion network for salient object detection[J]. Neurocomputing, 2021, 422:22–33.
- [15] 王凯诚,鲁华祥,龚国良,等.基于注意力机制的显著性目标检测方法[J].智能系统学报,2020,15(5):956–963.
- [16] FENG G, BO H, SUN J, et al. CACNet: salient object detection via context aggregation and contrast embedding[J]. Neurocomputing, 2020, 403:33–44.
- [17] WEI J, WANG S, HUANG Q. F³Net: Fusion, feedback and focus for salient object detection[C]// Association for the Advance of Artificial Intelligence. Proceedings of 2020 AAAI Conference on Artificial Intelligence. Menlo Park: AAAI, 2020, 34(07):12321–12328.
- [18] LIU Y, HAN J G, ZHANG Q, et al. Deep salient object detection with contextual information guidance[J]. IEEE Transactions on Image Processing, 2020, 29:360–374.
- [19] CHEN Z, XU Q, CONG R, et al. Global context-aware progressive aggregation network for salient object detection[C]// Association for the Advance of Artificial Intelligence. Proceedings of 2020 AAAI Conference on Artificial Intelligence. Menlo Park: AAAI, 2020, 34(07):10599–10606.
- [20] WU R, FENG M, GUAN W, et al. A mutual learning method for salient object detection with intertwined multi-supervision[C]// Institute of Electrical and Electronics Engineers. Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2019:8142–8151.
- [21] 陈家祯,黄添强,林文忠,等.融合小型深度生成模型的显著性检测[J].电子学报,2021,49(4):768–774.
- [22] ZHANG J, YU X, LI A, et al. Weakly-supervised salient object detection via scribble annotations[C]// Institute of Electrical and Electronics Engineers. Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2020:12543–12552.

作者简介:

李元贞(1996–),女,硕士生.研究领域:图像处理,深度学习,显著性目标检测.

赵俊松(1988–),男,硕士,工程师.研究领域:图像处理,数据挖掘.

(上接第8页)

- [5] 王卫东,李净,张福存,等.基于BP神经网络的太阳辐射预测——以兰州市为例[J].干旱区资源与环境,2014,28(2):185–189.
- [6] 张立影,孟令甲,王泽忠.基于双层BP神经网络的光伏电站输出功率预测[J].电测与仪表,2015,52(11):31–35.
- [7] 冯姣姣,王维真,李净,等.基于BP神经网络的华东地区太阳辐射模拟及时空变化分析[J].遥感技术与应用,2018,33(05):881–889,955.
- [8] 黄海静,张京红,覃文娜,等.基于小波神经网络法的海南地区逐时太阳总辐射预测研究[J].气象与环境学报,2016,32(1):60–65.
- [9] 谷同辉,管晓丹,高照远,等.黄河流域蒸散发与气温和降水以及风速的相关性分析[J].气象与环境学报,2022,38(01):48–56.
- [10] 蒋美云.基于模拟退火算法优化的BP神经网络预测模型[J].软件工程,2018,229(07):36–38.
- [11] 陈力,殷时蓉,罗天洪,等.基于BP神经网络的智能车辆换道决策模型研究[J].汽车工程学报,2022,12(01):83–89.
- [12] 海仁古丽·阿不力提甫,姑丽加玛丽·麦麦提艾力.一种基于小波神经网络的图像边缘检测[J].电子设计工程,2021,451(05):190–193.
- [13] 曹双华,曹家枞,刘凤强.小波分析在太阳辐射神经网络预测中的应用研究[J].东华大学学报(自然科学版),2004,30(6):18–22.
- [14] 秦航远,刘金朝,王卫东,等.基于多分辨率分析的轨道刚度检测数据自适应预处理方法[J].铁道建筑,2020,60(01):99–102.

作者简介:

鲁玉军(1976–),男,博士,教授.研究领域:智能制造,数字化设计与制造.

周世豪(1995–),男,硕士.研究领域:智能制造.

胡小勇(1994–),男,硕士.研究领域:智能制造.