JOURNAL OF IMAGE AND GRAPHICS

主办：中国科学院空天信息创新研究院
中国图象图形学学会
北京应用物理与计算数学研究所

中国图象图形学报

2021
05
VOL.26

ISSN1006-8961
CN11-3758/TB

参与介质渲染 *P961*

中国图象图形学报

刊名题字：宋健 | 月刊（1996年创刊）

第26卷第5期（总第301期）
2021年5月16日

中国精品科技期刊
中国国际影响力优秀学术期刊
中国科技核心期刊
中文核心期刊

主管单位 中国科学院
主办单位 中国科学院空天信息创新研究院
中国图象图形学学会
北京应用物理与计算数学研究所

主　　编 吴一戎
编辑出版 《中国图象图形学报》编辑部
通信地址 北京市海淀区北四环西路19号
邮　　编 100190
电子信箱 jig@radi.ac.cn
电　　话 010-58887035
网　　址 www.cjig.cn

广告发布登记号 京朝工商广登字20170218号
总 发 行 北京报刊发行局
订　　购 全国各地邮局
海外发行 中国国际图书贸易集团有限公司
（邮政信箱：北京399信箱　邮编：100048）
印刷装订 北京科信印刷有限公司

Journal of Image and Graphics

Title inscription: Song Jian | Monthly, Started in 1996

Superintended by Chinese Academy of Sciences
Sponsored by Aerospace Information Research Institute, CAS
China Society of Image and Graphics
Institute of Applied Physics and Computational Mathematics

Editor-in-Chief Wu Yirong
Editor, Publisher Editorial Office of Journal of Image and Graphics
Address No. 19, North 4th Ring Road West, Haidian District, Beijing, P. R. China
Zip code 100190
E-mail jig@radi.ac.cn
Telephone 010-58887035
Website www.cjig.cn

Distributed by Beijing Bureau for Distribution of Newspapers and Journals
Domestic All Local Post Offices in China
Overseas China International Book Trading Corporation
(P.O.Box 399, Beijing 100048,P.R.China))
Printed by Beijing Kexin Printing Co., Ltd.

CN 11-3758/TB
ISSN 1006-8961
CODEN ZTTXFZ

国外发行代号 M1406
国内邮发代号 82-831
国内定价 60.00元

2021年5月 第26卷 第5期（总第301期）
中国图象图形学报
Zhongguo Tuxiang Tuxing Xuebao

目次

均匀参与介质的渲染方法研究(第0961页)

循环生成对抗网络的线稿图像自动提取(第1117页)

伪点云修正增强激光雷达数据(第1157页)

学者观点

均匀参与介质的渲染方法研究
王贝贝，樊家辉 …… 0961
布料仿真建模研究进展
靳雁霞，张晋瑞，贾瑶，马博 …… 0970

综述

中国图像工程:2020
章毓晋 …… 0978

图像处理和编码

双阶段信息蒸馏的轻量级图像超分辨率网络
李明鸿，常侃，李恒鑫，谭宇飞，覃团发 …… 0991
具有纹理感知能力的超像素分割方法
吴江，刘春晓 …… 1006

图像分析和识别

深度聚类注意力机制下的显著对象检测
陈庆文，谢宏文，查浩，奚瑜，张雪 …… 1017
特征增强策略驱动的车标识别
贺敏雪，余烨，程茹秋 …… 1030

医学图像处理

融合上下文和注意力的视盘视杯分割
刘洪普，赵一浩，侯向丹，郭鸿湧，丁梦园 …… 1041

遥感图像处理

复杂背景下SAR近岸舰船检测
阮晨，郭浩，安居白 …… 1058
对抗型长短期记忆网络的雷达回波外推算法
方巍， 庞林， 张飞鸿， 盛胜利 …… 1067
混沌系统和DNA编码的并行遥感图像加密算法
周辉，谢红薇，张昊，张慧婷 …… 1081

激光雷达大会2020

图匹配算法激光扫描点云树干分割
徐姗姗 …… 1095
多特征融合与残差优化的点云语义分割方法
杜静，蔡国榕 …… 1105

NCIG 2020

循环生成对抗网络的线稿图像自动提取
王素琴，张加其，石敏，赵银君 …… 1117
面向COVID-19疫情预测的图卷积神经网络时空数据学习
杨成意，刘峰，齐佳音，段妍，吕润倩，肖子龙 …… 1128
结合汉明码和图像矫正的彩色图像盲水印
刘得成，苏庆堂，袁子涵，张雪婷 …… 1138
密文域高嵌入率图像全位面可逆数据隐藏
周旭，吴福虎，陈志立，任帅 …… 1147
伪点云修正增强激光雷达数据
宋绪杰，戴孙浩，林春雨，詹书涛，赵耀 …… 1157

前沿进展

中国遥感软件研制进展与发展方向——以像素专家PIE为例
刘东升，廖通逵，孙焕英，任芳 …… 1169

Volume 26, Number 5
Published May 16, 2021

CONTENTS
JOURNAL OF IMAGE AND GRAPHICS

Survey of rendering methods for homogeneous participating media(P0961)

Scholar View
Survey of rendering methods for homogeneous participating media
Wang Beibei, Fan Jiahui ····· 0961
Progress of cloth simulation modeling
Jin Yanxia, Zhang Jinrui, Jia Yao, Ma Bo ····· 0970

Review
Image engineering in China: 2020
Zhang Yujin ····· 0978

Image Processing and Coding
Lightweight image super-resolution network via two-stage information distillation
Li Minghong, Chang Kan, Li Hengxin, Tan Yufei, Qin Tuanfa ····· 0991
Superpixel segmentation with texture awareness
Wu Jiang, Liu Chunxiao ····· 1006

Image Analysis and Recognition
Salient object detection based on deep clustering attention mechanism
Chen Qingwen, Xie Hongwen, Zha Hao, Xi Yu, Zhang Xue ····· 1017
Vehicle logo recognition method based on feature enhancement
He Minxue, Yu Ye, Cheng Ruqiu ····· 1030

Image extraction of cartoon line art based on cycle-consistent adversarial networks(P1117)

Medical Image Processing
Optic disc and cup segmentation by combining context and attention
Liu Hongpu, Zhao Yihao, Hou Xiangdan, Guo Hongyong, Ding Mengyuan ····· 1041

Remote Sensing Image Processing
SAR inshore ship detection algorithm in complex background
Ruan Chen, Guo Hao, An Jubai ····· 1058
Radar echo extrapolation algorithm based on adversarial long short-term memory network
Fang Wei, Pang Lin, Zhang Feihong, Sheng Victor S ····· 1067
Parallel remote sensing image encryption algorithm based on chaotic map and DNA encoding
Zhou Hui, Xie Hongwei, Zhang Hao, Zhang Huiting ····· 1081

China LiDAR Conference 2020
Tree stem segmentation of laser scanning point clouds based on graph matching algorithm
Xu Shanshan ····· 1095
Point cloud semantic segmentation method based on multi-feature fusion and residual optimization
Du Jing, Cai Guorong ····· 1105

NCIG 2020
Image extraction of cartoon line art based on cycle-consistent adversarial networks
Wang Suqin, Zhang Jiaqi, Shi Min, Zhao Yinjun ····· 1117
Spatiotemporal data learning of graph convolutional neural network for epidemic prediction of COVID-19
Yang Chengyi, Liu Feng, Qi Jiayin, Duan Yan, Lyu Runqian, Xiao Zilong ····· 1128
Blind color image watermarking method combining Hamming code with image correction
Liu Decheng, Su Qingtang, Yuan Zihan, Zhang Xueting ····· 1138
All bit planes reversible data hiding for images with high-embedding-rate in ciphertext field
Zhou Xu, Wu Fuhu, Chen Zhili, Ren Shuai ····· 1147
LiDAR data enhancement via pseudo-LiDAR point cloud correction
Song Xujie, Dai Sunhao, Lin Chunyu, Zhan Shutao, Zhao Yao ····· 1157

LiDAR data enhancement via pseudo-LiDAR point cloud correction(P1157)

Frontier
Research progress and development direction of Chinese remote sensing software: taking PIE as an example
Liu Dongsheng, Liao Tongkui, Sun Huanying, Ren Fang ····· 1169

中图法分类号:TN957.52　　**文献标识码**:A　　**文章编号**:1006-8961(2021)05-1017-13

论文引用格式:Chen Q W, Xie H W, Zha H, Xi Y and Zhang X. 2021. Salient object detection based on deep clustering attention mechanism. Journal of Image and Graphics,26(05):1017-1029(陈庆文,谢宏文,查浩,奚瑜,张雪. 2021. 深度聚类注意力机制下的显著对象检测. 中国图象图形学报,26(05):1017-1029)[DOI:10.11834/jig.200081]

深度聚类注意力机制下的显著对象检测

陈庆文[1],谢宏文[2],查浩[2],奚瑜[1],张雪[3,4*]

1. 中国电建集团西北勘测设计研究院有限公司, 西安　710065; 2. 水电水利规划设计总院, 北京　100120;
3. 天津大学智能与计算学部, 天津　300350; 4. 天津市机器学习重点实验室, 天津　300350

摘　要:**目的**　为了得到精确的显著对象分割结果,基于深度学习的方法大多引入注意力机制进行特征加权,以抑制噪声和冗余信息,但是对注意力机制的建模过程粗糙,并将所有特征均等处理,无法显式学习不同通道以及不同空间区域的全局重要性。为此,本文提出一种基于深度聚类注意力机制(deep cluster attention,DCA)的显著对象检测算法 DCANet(DCA network),以更好地建模特征级别的像素上下文关联。**方法**　DCA 显式地将特征图分别在通道和空间上进行区域划分,即将特征聚类分为前景敏感区和背景敏感区。然后在类内执行一般性的逐像素注意力加权,并在类间进一步执行语义级注意力加权。DCA 的思想清晰易懂,参数量少,可以便捷地部署到任意显著性检测网络中。**结果**　在 6 个数据集上与 19 种方法的对比实验验证了 DCA 对得到精细显著对象分割掩码的有效性。在各项评价指标上,部署 DCA 之后的模型效果都得到了提升。在 ECSSD(extended cornplex scene saliency dataset)数据集上,DCANet 的性能比第 2 名在 F 值上提升了 0.9%;在 DUT-OMRON(Dalian University of Technology and OMRON Corporation)数据集中,DCANet 的性能比第 2 名在 F 值上提升了 0.5%,平均绝对误差(mean absolute error,MAE)降低了 3.2%;在 HKU-IS 数据集上,DCANet 的性能比第 2 名在 F 值上提升了 0.3%, MAE 降低了 2.8%;在 PASCAL(pattern analysis, statistical modeling and computational learning)-S(subset)数据集上,DCANet 的性能则比第 2 名在 F 值上提升了 0.8%,MAE 降低了 4.2%。**结论**　本文提出的深度聚类注意力机制通过细粒度的通道划分和空间区域划分,有效地增强了前景敏感类的全局显著得分。与现有的注意力机制相比,DCA 思想清晰、效果明显、部署简单,同时也为一般性的注意力机制研究提供了新的可行的研究方向。

关键词:显著性检测;注意力机制;深度聚类;空间—通道维度解耦;全卷积网络(FCN)

Salient object detection based on deep clustering attention mechanism

Chen Qingwen[1], Xie Hongwen[2], Zha Hao[2], Xi Yu[1], Zhang Xue[3,4*]

1. China Power Construction Corporation Northwest Survey Design and Research Institute Co., Ltd., Xi'an　710065, China; 2. China Renewable Energy Engineering Institute, Beijing　100120, China; 3. College of Intelligence and Computing, Tianjin University, Tianjin　300350, China; 4. Tianjin Key Laboratory of Machine Learning, Tianjin　300350, China

Abstract: **Objective** Salient object detection is a basic task in the field of computer vision, which simulates the human visual attention mechanism and quickly detects attractive objects in the scene that are most likely to represent user query varia-

收稿日期:2020-03-13;**修回日期**:2020-07-15;**预印本日期**:2020-07-22

*通信作者:张雪　zhx95@tju.edu.cn

基金项目:国家自然科学基金项目(61772360,61876125)

Supported by: National Natural Science Foundation of China (61772360,61876125)

bles and contain the most information. As a preprocessing step of other vision tasks, such as image resizing, visual tracking, person re-identification, and image segmentation, salient object detection plays a very important role. The traditional salient object detection method mainly uses the method of manually extracting features of the image to detect. However, this process is time-consuming and labor-intensive, and the results cannot meet the requirements. With the rise of deep learning, a large number of feature extraction algorithms based on convolutional neural networks have emerged. Compared with traditional feature extraction methods, using deep neural networks to extract features has better quality and more accurate prediction. In order to obtain accurate salient object segmentation results, deep learning-based methods mostly introduce attention mechanisms for feature weighting to suppress noise and redundant information. However, the modeling process of the existing attention mechanism is quite rough, which treats each position in the feature tensor equally and directly solves the attention score. This strategy cannot explicitly learn the global importance of different channels and different spatial regions, which may lead to missed detection or misdetection. To this end, in this study, we propose a deep clustering attention (DCA) mechanism to better model the feature-level pixel-by-pixel relationship. **Method** In this study, the proposed DCA explicitly divides the feature tensors into several categories channel-wise and spatial-wise; that is, it clusters the features into foreground and background sensitive regions. Then, general per-pixel attention weighting is performed within each class, and semantical attention weighting is further performed inter-classes. The idea of DCA is easy to understand, whose parameter quantity is also small and can be deployed in any salient detection network. This method can efficiently separate the foreground and background regions. In addition, through supervised learning on the edges of salient objects, the prediction can get clearer edges, and the results are more accurate. **Result** Comparison of 19 state-of-the-art methods on six large public datasets demonstrates the effectiveness of DCA in modeling pixel-wise attention, which is very helpful for obtaining finely salient object segmentation mask. On various evaluation indicators, the effects of the model after the deployment of DCA have improved. On the extended cornplex scene saliency (ECSSD) dataset, the performance of DCA-Net increased by 0.9% over the second place (F-measure value). On the Dalian University of Technology and OMRON Corporation (DUT-OMRON) dataset, the performance of DCANet increased by 0.5% over the second place (F-measure value), and the MAE decreased by 3.2%. On the HKU-IS dataset, the performance of DCANet is 0.3% higher than the second place (F-measure value), and the MAE is reduced by 2.8%. On the pattern analysis, statistical modeling and computational learning (PASCAL)-subset (S) dataset, the performance of DCANet is 0.8% higher than the second place (F-measure value), and the MAE is reduced by 4.2%. **Conclusion** The DCA proposed in this study effectively enhances the globally salient scores of foreground sensitive classes through more fine-grained channel partitioning and spatial region partitioning. This paper analyzes the deficiencies of the existing salient object detection algorithm based on attention mechanism and proposes a method for explicitly dividing feature channels and spatial regions. The attention modeling mechanism helps the model training process perceive and adapt tasks quickly. Compared with the existing attention mechanism, the idea of DCA is clear, the effect is significant, and it is simple to deploy. Meanwhile, DCA provides a viable new research direction for the study of more general attention mechanisms.

Key words: saliency detection; attention mechanism; deep clustering; spatial-channel decoupling; full convolutional network (FCN)

0 引 言

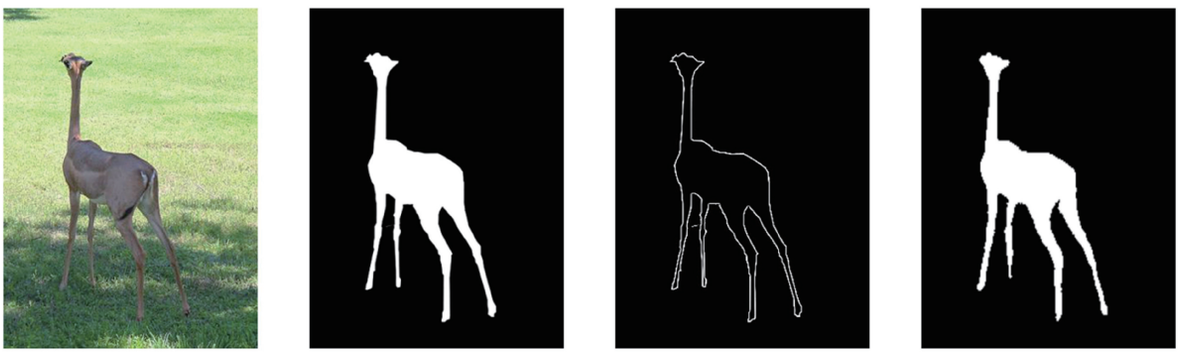
生物学和神经科学等研究表明,人类在观看静态图像或动态视频时,会选择性地过滤掉无用或不感兴趣的信息,而仅对感兴趣的局部区域进行重点处理,这种注意力分配机制称为人类视觉注意力机制,能够显著提升信息处理和获取速度。影响注意力分配的因素主要来自低级对比度特征和更高级别的语义特征。当眼部停留时间很短时,颜色、轮廓和亮度等低级特征是影响注意力分配的决定因素。随着在场景中停留时间的增加,注意力分配方式逐渐转变为基于知识指导的形式,语义特征或运动特征成为主要影响因素。为了模拟这种视觉注视机制,显著性检测应运而生。作为计算机视觉类任务的关键预处理步骤,显著性检测得到越来越多的关注,并已经广泛用于对象跟踪(Borji 等,2012)、人物重定位(Zhao 等,2013)、图像分割/编辑(Qin 等,2018)

和机器人导航(Craye 等,2016)等领域。

根据任务和输入数据形式的不同,显著性检测分为显著对象分割(Borji 等,2015)和眼部注视点预测(Huang 等,2015)。本文主要针对显著对象分割任务展开讨论。早期的显著性检测主要基于手工特征提取图结构,提取图像的颜色、纹理和亮度等低级特征,并对这些特征分别构建单一模式显著图,采用基于图的随机游走策略(Yuan 等,2018)、马尔可夫随机场(Harel 等,2007)和流形排序(Zhang 等,2017c)等方式进行特征整合,得到全局性的显著预测结果。这种算法基于低级的对比特征,在处理具有复杂语义交互关系的场景时,由于缺乏高级语义信息的指导,检测效果并不理想。随着深度学习技术的飞速发展,卷积神经网络(convolutional neural networks,CNNs)开始在结构化数据处理方面崭露头角,得益于其自适应提取图像数据多尺度特征的巨大优势,大量基于 CNNs 的显著性研究问世(Sun 等,2019),并在诸多公开基准数据集上击败传统方法,成为主流方法。

基于 CNNs 的显著性检测算法,大多通过构建特征金字塔来保证检测结果的准确性(Hou 等,2017)。该类方法对前向求解过程获取的多尺度特征反向拼接,即将语义信息丰富的特征用于指导空间结构信息丰富的特征,从而保证信息的完整性,得到兼顾语义准确性和空间信息完整性的显著预测图。但是,对多尺度特征不加区分地进行融合会导致冗余和噪声信息前向传递和放大,影响检测结果的准确性。为此,将基于注意力机制的特征提取策略用于特征筛选(Wang 等,2018b;Feng 等,2019),通过在特征级别执行注意力求解,每一个位置的特征从原始的权重 1 重新约束为一个代表其重要性的概率值。注意力矩阵与原始特征的逐像素相乘有效保留和放大了正向积极特征,同时抑制了干扰和无用区域的特征。这种注意力矩阵类似于一种掩码,保留关键区域的同时屏蔽无关区域,不仅有助于得到准确的特征,也加快了后期的显著性解码过程。

基于上述特征加权机制,本文提出一种深度聚类注意力机制(deep cluster attention,DCA),以更为合理地求解注意力掩码。与以往注意力模型均等处理所有特征的方式不同,此处显式地将特征在通道和空间级别划分为前景敏感类和背景感敏感类。对背景敏感类不加区分地进行抑制,而对前景敏感类则进一步在类内执行精细化的注意力加权,从而得到准确的注意力掩码。本文的主要贡献为:1)系统分析了现有基于注意力机制的显著性检测算法的利弊,提出了基于深度聚类注意力机制的显著对象检测算法 DCANet(DCA network);2)提出了通道聚类注意力机制(channel deep cluster attention, CDCA)和空间聚类注意力机制(spatial deep cluster attention, SDCA),将特征分别在通道和空间上进行聚类划分,得到前背景分离明确的通道注意力得分,以及局部相关性和全局一致性的空间注意力得分,有助于更完备的空间显著区域选择。本文提出的 DCANet 的预测效果图和模型结构图如图 1 和图 2 所示。



(a) 输入图像　(b) 真值标签　(c) 边缘真值　(d) 本文预测图

图 1　本文 DCANet 的预测效果图

Fig. 1　The salient object detection result of our proposed DCANet

((a) input image; (b) ground truth; (c) boundary ground truth; (d) saliency prediction of ours)

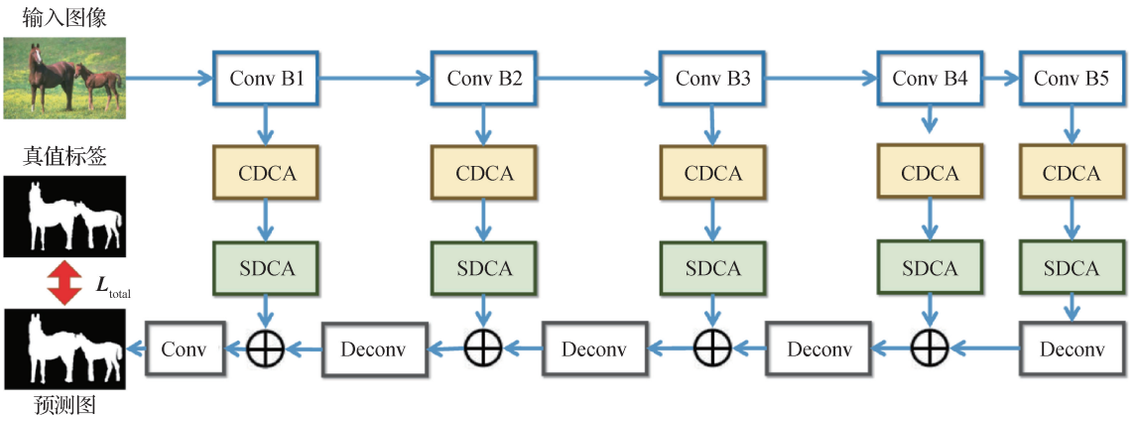


图 2　本文提出的 DCANet 的模型结构图

Fig. 2　The architecture of the proposed DCANet

1　相关工作

1.1　显著对象检测

早期的深度显著对象检测模型通常利用多层感知机(multi-layer perceptron, MLP)预测提取的深度特征的每个单元的显著性得分(Li 和 Yu, 2015; Han 等, 2015; Wang 等, 2015)。其中, Li 和 Yu(2015)采用多尺度特征聚合的方式, 赋予不同尺度的输出不同的监督信息, 以提高预测精度, 同时提出了一个大型的显著对象分割数据集 HKU-IS, 为后续研究提供了一个新的可参照基准数据集。Han 等人(2015)提出栈式去噪自编码器(stacked denoising auto-encoders, SDAEs), 并通过两个阶段的学习来求解每个位置的显著得分, 第 1 个学习阶段 SDAEs 以无监督的形式从原始图像数据中学习能够保留代表性特征并剔除高频噪声信息的特征, 第 2 个学习阶段则以有监督的形式学习最后一层分类器, 以求解最终的显著值。

随着全卷积神经网络(fully convolutional neural networks, FCNs)这种高效和对等的结构逐渐成为深度显著对象检测模型结构的主流(Li 等, 2016; Wang 等, 2018a), 为了得到边界精确的显著图, 研究者试图寻找最佳的多尺度信息融合策略, 以解决由下采样导致的空间尺度信息丢失问题。Hou 等人(2017)构建横向连接结构(deeply-supervised skip-layer structure, DSS)将深层特征连接到浅层特征, 以期利用高级语义特征指导低级结构特征准确地定位显著区域, 同时保留低级特征中蕴含的完整边界信息。在此基础上, Zhang 等人(2017a)同样采用级联结构(aggregating multi-level network, Amulet)将多级特征聚合为多分辨率输出。每个分辨率的输出都构成独立的数据链路, 以自顶向下的方式进行完善, 同时利用边界优化策略进行细节修复。Cong 等人(2019)和 Wang 等人(2019)对各个时期的多种综合性信息的视觉显著性检测工作进行了详细的总结说明。

尽管显著性检测方面的工作已经相对成熟, 然而在部分细节问题上仍然可以更进一步。比如在对前后背景分离时没有考虑边缘部分的细节处理, 以及由于边缘导致的问题如何优化等。而本文对边缘部分单独进行监督学习, 可以大大提高模型的边缘感知能力, 缓解边缘检测不准带来的低精度问题。

1.2　注意力机制

虽然特征金字塔结构有效融合了语义信息和边界信息, 提高了显著性预测结果的准确性, 但是无区分地将所有多尺度特征进行合并存在诸多弊端。例如, 许多噪声信息和冗余信息会不断放大和前馈, 不仅增加了特征解码过程的训练难度, 也导致预测结果出现一定偏差。为此, 在构建特征金字塔之前对原始特征进行筛选和加权, 构建注意力机制成为一种有效的解决方案。Wang 等人(2017)提出基于全局平滑池化的通道注意力机制(weakly supervised saliency detection, WSS), 对最顶层的语义特征执行通道加权, 求解每个通道的特征表征, 然后利用 Softmax 进行归一化处理, 得到全局显著比重。但 WSS 只考虑了通道注意力求解, 缺乏对空间维度的特征加权, 因此对信息的过滤和筛选并不充分。

Wang 等人(2018b)提出 DGRL(detect globally refine locally),使用 Inception 结构获取多尺度特征,并使用具有单个输出通道的 Softmax 激活卷积层计算注意力得分。这种注意力机制对空间位置进行加权,但是缺乏通道维度的注意力计算,并且所有位置的权重总和为 1,这意味着同时存在两个位置都具有很高的显著概率是不可能的。显然这种设计违背了图像的局部强相关性特点。对空间注意力矩阵的计算应允许多个位置同时具有较高的显著概率,这有助于保留更多的显著对象候选区。Feng 等人(2019)提出 AFNet(attentive feedback network),认为上述基于权值矩阵和原始特征的 Hadamard 乘积而完成的注意力求解过程是不完备的,因为这种操作严重依赖指导信息的准确性,一旦权值求解过程出现误差,不准确的指导信息将导致价值信息丢失,而错误信息却有可能过度使用,这将导致分割主要目标时的对象漂移。为了避免这种障碍,AFNet 使用三元注意力图作为特征指导信息,将原始特征区分为 3 部分,分别表示自信的前景区域、自信的背景区域和不确定区域。这种对模糊信息的容错性能够有效避免价值信息丢失和噪声信息前馈。

与上述所有基于注意力机制的显著对象检测算法不同,本文方法对原始特征同时在通道和空间维度进行显式注意力建模。考虑到每个通道对某种单一的特征模式敏感,对通道进行聚类,将其划分为前景敏感类和背景敏感类,这种聚类操作有效地在通道维度排除了背景噪声干扰。对空间进行聚类,分别计算每个像素点独立存在时的全局显著得分以及归结到某种固定大小的局部特征块时的全局显著得分,从而得到兼顾局部相似性和全局一致性的空间注意力掩码。实验结果证明,本文提出的聚类注意力机制能够有效地建模特征的上下文关联,得到精确的显著特征。DCA 模块独立于特征编码器存在,可以容易地插入到任何网络结构中。

2　深度聚类注意力显著性模型

本文提出的基于深度聚类注意力机制的显著性检测算法的模型结构如图 2 所示。为了获取充分的多尺度特征,将 VGGNet(Visual Geometry Group network)(Simonyan 和 Zisserman,2014)用于特征编码,同时为了保存充足的空间结构,丢弃了 VGGNet 的最后一个池化层以及后续的全连接层,只保留前面 5 个卷积块,且假设 5 个卷积块的特征分别为 $\{\boldsymbol{B}_1, \boldsymbol{B}_2, \boldsymbol{B}_3, \boldsymbol{B}_4, \boldsymbol{B}_5\}$,分别对不同尺度的特征执行通道注意力求解和空间注意力求解。

2.1　深度聚类注意力机制

2.1.1　通道聚类注意力 Channel-DCA

图 3 是通道聚类注意力机制示意图。对于特定尺度的原始特征 $\boldsymbol{B}_i^{w\times h\times c}$,通道注意力机制在通道维度上执行注意力加权并求解经过加权的优化特征 $\boldsymbol{B}_{i\mathrm{DCA}}^{w\times h\times c}$。

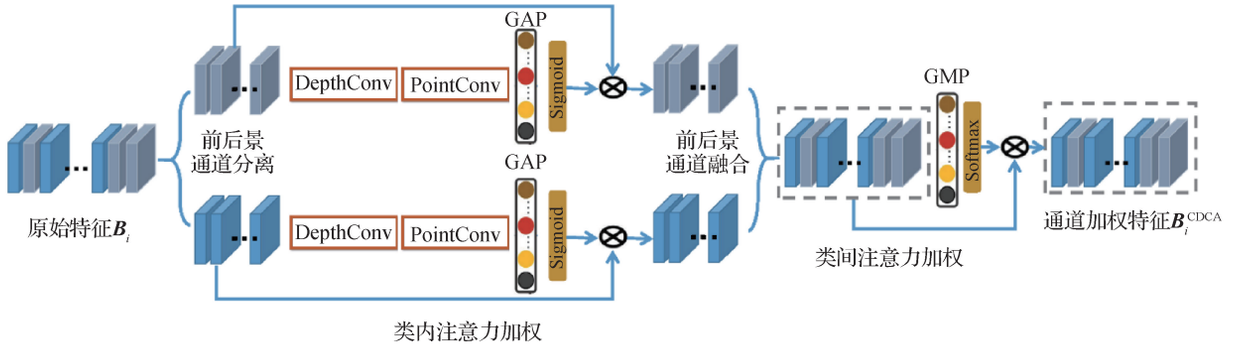


图 3　通道聚类注意力机制

Fig. 3　Channel deep clustered attention

为了减少张量运算,首先使用通道数为 128 的 1 × 1 点积(point convolution,PC)运算来减少特征通道数。具体为

$$\widehat{\boldsymbol{B}}_i = F_{1\times1}^{128}(\boldsymbol{B}_i,[\boldsymbol{\omega};b]) \tag{1}$$

式中,$\widehat{\boldsymbol{B}}_i$ 代表通道数减少后的特征,$F_{1\times1}^{128}(\cdot)$ 操作代表通道数为 128 的 1 × 1 点积运算,$\boldsymbol{B}_i$ 为原始特征,$\boldsymbol{\omega}$ 表示权重参数,b 为偏置项。

然后,将 128 个通道自适应地划分为前景敏感

类和背景敏感类两个类别。对前 64 个通道用真值标签进行监督学习,对后 64 个通道用反转后的真值标签(即原真值标签中值为 0 的像素改为 1,原真值标签中值为 1 的像素改为 0)进行监督学习,进而实现对前景背景的分离。通道聚类注意力机制中的前后景通道分离特征可视化示意图如图 4 所示,这种显式的类别划分方式能够促使前景和背景两个类别之间的语义差距越来越大,类内差距越来越小。即

$$\boldsymbol{C}_1 = \widehat{\boldsymbol{B}}_i^{w\times h\times 1\cdots 64},\ \boldsymbol{C}_2 = \widehat{\boldsymbol{B}}_i^{w\times h\times 65\cdots 128} \tag{2}$$

式中,$\boldsymbol{C}_1$ 代表前景特征,$\boldsymbol{C}_2$ 代表背景特征,$\widehat{\boldsymbol{B}}_i^{w\times h\times 1\cdots 64}$ 和 $\widehat{\boldsymbol{B}}_i^{w\times h\times 65\cdots 128}$ 分别是通道减少后的特征中的前 64 个和后 64 个特征通道组成的特征。

对于单独的每个类簇,通道注意力机制执行类内加权。因为类内特征模式更加一致,在类内求解注意力得分有助于得到聚焦于当前类别代表对象的特征通道。为了减少需要训练的参数量,首先使用深度可分离卷积(depthwise convolution, DC)运算单独求解每一个通道的特有特征模式,然后使用 *filter* × 1 × 1 的点积运算进行通道信息融合,*filter* 为卷积核的数量。具体为

$$\boldsymbol{C}'_j = F_{\mathrm{PC}}^{64}(F_{3\times 3}^{\mathrm{DC}}(\boldsymbol{C}_j, [\boldsymbol{\omega}_{<3,3>}; b_1]), [\boldsymbol{\omega}_{<1,1>}; b_2]) \tag{3}$$

式中,$\boldsymbol{C}'_j$ 表示通道信息融合后的特征,$F_{\mathrm{PC}}^{64}(\cdot)$ 表示对 64 个特征通道进行点积运算,$F_{3\times 3}^{\mathrm{DC}}(\cdot)$ 表示进行 3 × 3 的深度可分离卷积运算,$\boldsymbol{C}_j$ 为处理前特征,$\boldsymbol{\omega}_{<3,3>}$ 为卷积核大小为 3 的权重参数,$\boldsymbol{\omega}_{<1,1>}$ 为卷积核大小为 1 的权重参数,b_1 和 b_2 为偏置项。

从式(3)可以看出,先 DC 后 PC 的方式有效减少了参数量,是一般卷积运算参数量的 1/*filter*,这里 *filter* = 64。

接下来,利用全局平均池化(global average pooling, GAP)对 $\boldsymbol{C}'_j$ 求解每个通道上的代表值,并将每个通道上的特征均值作为当前通道的代表值,即

$$\boldsymbol{S}_{C_j}^c = \frac{1}{w\times h}\sum_{q=1;p=1}^{w,h}\boldsymbol{C}_i'^{q\times p\times c} \tag{4}$$

式中,$\boldsymbol{S}_{C_j}^c$ 为每个通道上的代表值,w 和 h 分别为特征图的宽和高,c 为通道数。

为了尽可能多地保留关注当前类别对象的多个通道,采用 Sigmoid 函数进行归一化操作,即对代表向量 $\boldsymbol{S}_i \in \{\boldsymbol{S}_{Ci}^1, \boldsymbol{S}_{Ci}^2, \cdots, \boldsymbol{S}_{Ci}^{64}\}$ 求解对应通道的全局显著概率。具体为

$$P_i^c = \frac{1}{1+\exp(-\boldsymbol{S}_i^c)} \tag{5}$$

式中,P_i^c 为对应通道的全局显著概率,$\boldsymbol{S}_i^c$ 为对应通道的代表向量。

依据求解的注意力向量,对通道权重进行重分配,得到类内加权特征 $\boldsymbol{C}_{\mathrm{aug}}$

$$\boldsymbol{C}_{\mathrm{aug}}^i = \boldsymbol{C}^i \cdot P^i \tag{6}$$

式中,$\boldsymbol{C}^i$ 为权重分配前对应通道的特征,P^i 为对应通道的权重。

经过类内加权初步得到了每个类别的代表性特征,然后将所有的类别在通道维度进行整合,并通过张量转置进行多类通道重排,以实现类间信息交互。具体为

$$\boldsymbol{B}_i^{\mathrm{T}} = Concat(\boldsymbol{C}_{\mathrm{aug}}^1, \boldsymbol{C}_{\mathrm{aug}}^2)^{\mathrm{T}} \tag{7}$$

式中,$Concat(\cdot)$ 为通道连接操作。

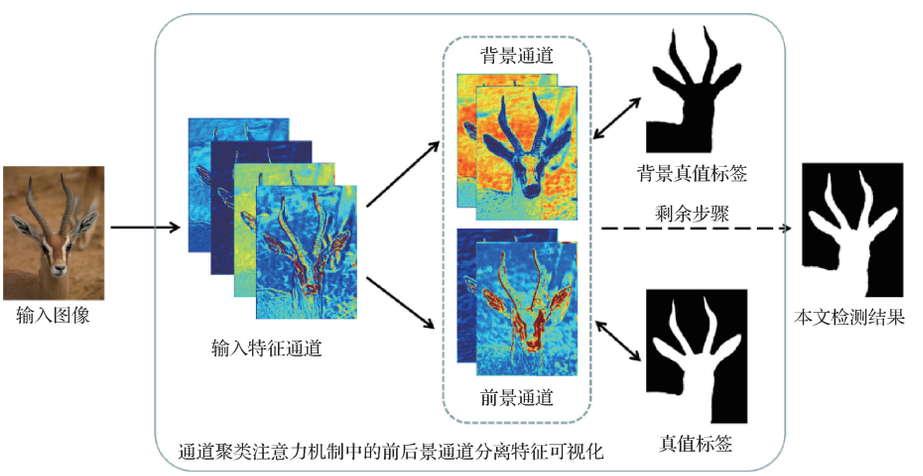


图 4 通道聚类注意力机制中的前后景通道分离特征可视化示意图

Fig. 4 Separation of foreground and background channels in channel deep clustered attention

转置交换后的特征为 $\boldsymbol{B}_i^{\mathrm{T}}$，对其进一步执行类间注意力加权，以抑制背景类，增强前景类。具体为

$$\boldsymbol{B}(i)^{\mathrm{CDCA}} = \mathrm{softmax}(GMP(\boldsymbol{B}_i^{\mathrm{T}})) \cdot \boldsymbol{B}_i^{\mathrm{T}} \tag{8}$$

式中，GMP 表示全局最大池化(global max pooling)，softmax 为归一化函数。使用 softmax 函数是为了保证所有通道的权重之和为 1，即重点关注少数几个前景敏感通道。

2.1.2　空间聚类注意力 Spatial-DCA

图 5 是空间聚类注意力机制示意图。通道聚类加权后，在空间维度进行注意力求解。首先，在全局范围求解每个像素点的上下文关联，即在执行全局平均池化后利用 softmax 函数求解每个位置的显著得分。具体为

$$S_G(i,j) = \frac{\mathrm{e}^{-AP_s(x(i,j))}}{\sum_{i,j}\mathrm{e}^{-AP_s(x(i,j))}} \tag{9}$$

式中，(i,j) 为像素点坐标，$x(i,j)$ 为像素点，$S_G(\cdot)$ 为像素点对应的全局显著得分，$AP_s(\cdot)$ 应用于空间聚类注意力的全局平均池化操作。

接着，利用金字塔池化构建不同大小的局部像素块，以充分求解当前像素点所在像素块与其他像素块之间的关联，从而得到充分的局部显著得分。具体为

$$S_L^{2\times2} = \delta(F^1_{\langle3\times3\rangle}(AP_{\langle2\times2\rangle}(x);[\boldsymbol{\omega},b])) \tag{10}$$

$$S_L^{4\times4} = \delta(F^1_{\langle3\times3\rangle}(AP_{\langle4\times4\rangle}(x);[\boldsymbol{\omega},b])) \tag{11}$$

$$S_L^{8\times8} = \delta(F^1_{\langle3\times3\rangle}(AP_{\langle8\times8\rangle}(x);[\boldsymbol{\omega},b])) \tag{12}$$

$$S_L = S_L^{2\times2} + S_L^{4\times4} + S_L^{8\times8} \tag{13}$$

最后，通过联合局部和全局分量求解得到最终的空间注意力得分。即

$$\boldsymbol{B}(i)_S = (S_L(i,j) + S_G(i,j) + \boldsymbol{I}) \cdot \boldsymbol{B}(i)_C \tag{14}$$

式中，$\boldsymbol{I}$ 表示值为 1 的所有矩阵，采用残差结构来构造多条通路，保证信息流在前馈过程中自适应融合局部和全局分量，得到鲁棒的加权特征。

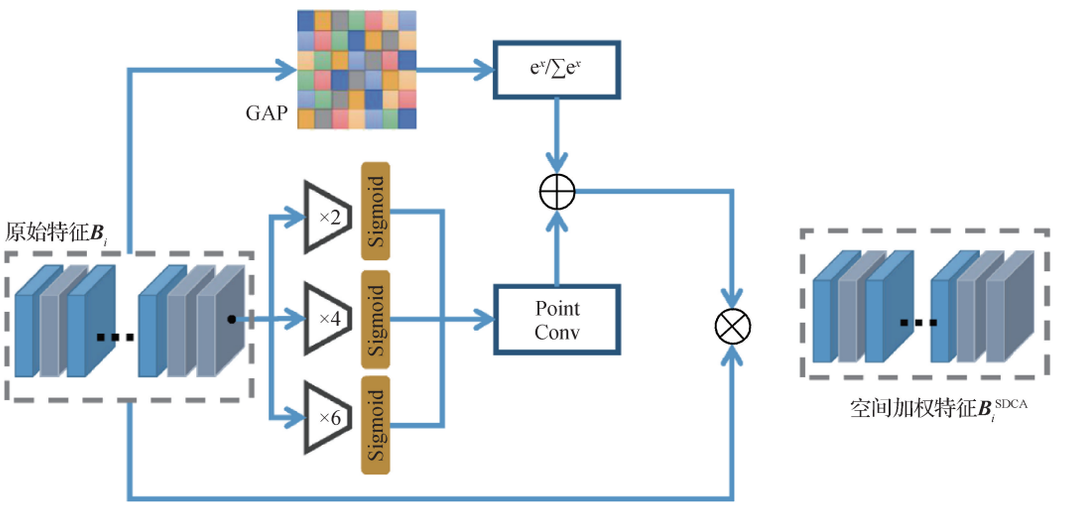


图 5　空间聚类注意力机制

Fig. 5　Spatial deep clustered attention

2.2　显著性推断

将深度聚类注意力机制用于多尺度特征优化过程，并通过栈式堆叠的多层反卷积结构完成最终的特征解码。

上层特征经过执行 DCA 之后，使用卷积核大小为 4 × 4、步长为 2 × 2 的反卷积运算进行特征上采样。此时，特征的空间分辨率变大，但是由于上采样可能会导致棋盘伪影，对其构建横向连接，与基于 DCA 处理的下层特征进行逐像素求和。以此类推，得到最终融合多尺度加权特征信息的显著特征。即

$$\boldsymbol{X}_{\mathrm{fuse}}^i = \begin{cases} A(F_{\mathrm{de}}^i(\boldsymbol{B}_{\mathrm{DCA}}^i;[\boldsymbol{\varpi};b]),\boldsymbol{B}_{\mathrm{DCA}}^{i-1}) \\ \delta(F_{\mathrm{conv}}^{\langle3\times3\rangle}(\boldsymbol{X}_{\mathrm{fuse}}^{i-1};[\boldsymbol{\omega},b])) \end{cases} \tag{15}$$

式中，$\boldsymbol{X}_{\mathrm{fuse}}^i$ 表示最终获得的融合了多尺度加权特征信息的显著特征，$A(\cdot)$ 代表像素级加法操作，$\delta(\cdot)$ 为非线性函数，$F_{\mathrm{de}}^i(\cdot)$ 表示权重为 $\boldsymbol{\varpi}$、偏置项为 b 的反卷积操作，$\boldsymbol{B}_{\mathrm{DCA}}^i$ 为基于 DCA 处理的特征，$F_{\mathrm{conv}}^{\langle3\times3\rangle}(\cdot)$ 表示卷积核大小为 3 × 3 的卷积操作。

2.3　损失函数

DCANet 在混合损失的监督下进行训练，混合损失 L_{total} 包括交叉熵损失 L_{BCE} 和边缘损失 L_{BD}。

1)交叉熵损失 L_{BCE}。将显著对象分割任务视为二分类任务,并执行逐像素类别求解,0 表示背景,1 表示前景。交叉熵损失 L_{BCE} 为

$$L_{BCE} = -\sum_{i=1}^{H}\sum_{j=1}^{W}[G_{ij}\log(\boldsymbol{S}_{ij}) + (1-G_{ij})\log(1-\boldsymbol{S}_{ij})] \tag{16}$$

式中,$\boldsymbol{S}$ 为显著图,G 为对应的真值标签。

2)边缘损失 L_{BD}。对真值标签以及检测结构分别应用拉普拉斯算子,得到显著边缘真值 $\boldsymbol{S}'$ 和显著检测结果边缘图 $\boldsymbol{G}'$,然后采用二值交叉熵损失求解边缘损失,进一步改善边缘部分的检测结果。边缘损失 L_{BD} 为

$$L_{BD} = L_{BCE}(\boldsymbol{S}',\boldsymbol{G}') \tag{17}$$

3)混合损失 L_{total}。最终的混合损失为

$$L_{total} = \lambda_1 \times L_{BCE} + \lambda_2 \times L_{BD} \tag{18}$$

式中,λ_1 和 λ_2 为两个权重因子,$\lambda_1 : \lambda_2 = 5 : 1$。

3 实验分析

3.1 模型部署细节

整个模型基于 Keras 深度学习框架部署,使用 DUTS-TR(DUTS-Train)(Wang 等,2017)数据集进行训练。DUTS-TR 数据集包含 10 553 幅图像,通过反转、旋转和镜像等图像增强操作,最终的训练数据量达到 70 000 幅。DCANet 的前 5 个卷积模块基于 ImageNet(Deng 等,2009)训练得到的 VGGNet16 初始化,其余层的参数均采用随机初始化。实验在 NVIDIA Geforce GTX 1080Ti(11 GB 内存)和 i7-8700k CPU 的硬件环境上完成。在训练过程中,输入数据调整为 224 × 224 像素。为了加速训练,本文使用 Adam(Kingma 和 Ba,2014)优化器进行参数优化,初始学习率设置为其默认值,即 1×10^{-3}。

3.2 数据集

为了有效评估本文提出的 DCANet,选取 DUT-OMRON(Dalian University of Technology and OMRON Corporation)(Yang 等,2013)、ECSSD(extended complex scene saliency dataset)(Shi 等,2016)、HKU-IS(Li 等,2016)、PASCAL(pattern analysis, statistical modeling and computational learning)-S(subset)(Li 等,2014)、SOD(salient object detection)(Movahedi 和 Elder,2010)和 DUTS-TE(DUTS-Test)(Wang 等,2017)等 6 个公开基准数据集进行性能验证。上述数据集的统计信息如表 1 所示,其中对象数量和显著对象面积占比信息取自相关文献(Wang 等,2019)。

表 1 各数据集的统计信息

Table 1 Summary of different datasets

数据集	大小	对象数量	对象面积占比/%
DUTS-OMRON	5 168	1～4+	14.85±12.15
ECSSD	1 000	1～4+	23.51±14.02
HKU-IS	4 447	1～4+	19.13±10.90
PASCAL-S	850	1～4+	24.23±16.70
DUTS-TE	5 019	1～4+	23.17±15.52
SOD	300	1～4+	27.99±19.36

注:"4+"表示对象数量在 4 个以上。

3.3 评价指标

为了充分说明 DCANet 的性能并获得客观结论,采用结构度量 S-measure(Fan 等,2017)、平均绝对误差(mean absolute error, MAE)(Borji 等,2015)和 F-measure 值等指标进行评估。

S-measure 更注意结构上的相似性,具体为

$$S_\lambda = \lambda \times S_O + (1-\lambda) \times S_R \tag{19}$$

式中,S_O 和 S_R 分别是对象感知结构相似性和区域感知结构相似性,$\lambda = 0.5$。

MAE 衡量了检测结果 $\boldsymbol{S}$ 与真值标签 G 之间的相似性,具体为

$$f_{MAE} = \frac{1}{W\times H}\sum_{x=1}^{W}\sum_{y=1}^{H}|S(x,y) - G(x,y)| \tag{20}$$

式中,W 和 H 是图像的宽和高,(x, y) 是像素点。

F-measure 是对平均查准率 P 和平均查全率 R 的加权平均值。具体为

$$F_\beta = \frac{(1+\beta^2)\times P\times R}{\beta^2(P+R)} \tag{21}$$

式中,β 为常量,取值 0.3。

3.4 实验结果分析

3.4.1 对比方法

为了验证本文算法的有效性,与 HS(hierarchical saliency)(Yan 等,2013)、MCDL(multi-context deep learning)(Zhao 等,2015)、LEGS(local estimation and global search)(Wang 等,2015)、MDF(multi-scale deep features)(Li 和 Yu,2015)、ELD(encoded low level distance)(Lee 等,2016)、DHSNet(deep hierarchical saliency network)(Liu 和 Han,2016)、DCL

(deep contrast learning)(Li 和 Yu,2016)、MAP(maximum a posterior)(Zhang 等,2016)、RFCN(recurrent fully convolutional networks)(Wang 等,2018a)、MDS(multi-task deep saliency)(Li 等,2016)、DSS(deeply supervised salient object detection)(Hou 等,2017)、WSS(weakly supervised saliency)(Wang 等,2017)、Amulet(Zhang 等,2017a)、UCF(uncertain convolutional features)(Zhang 等,2017b)、PiCANet(pixel-wise contextual attention network)(Liu 等,2018)、PAGRN(progressive attention guided recurrent network)(Zhang 等,2018a)、DGRL(detect globally, refine locally)(Wang 等,2018b)、C2SNet(contour-to-saliency)(Li 等,2018)和 MWSNet(multi-source weak supervision network)(Zeng 等,2019)等 19 种代表性显著性检测算法进行性能比较。上述模型的详细信息如表 2 所示。

表 2　不同模型的详细信息

Table 2　Summary of different models

模型	文献	骨干网络	训练集
HS	Yan 等人(2013)	–	–
MCDL	Zhao 等人(2015)	GoogLeNet	MSRA10K
LEGS	Wang 等人(2015)	–	MSRA-B + PASCAL-S
MDF	Li 和 Yu(2015)	–	MSRA-B
ELD	Lee 等人(2016)	VGGNet	MSRA10K
DHSNet	Liu 和 Han(2016)	VGGNet	MSRA10K + DUT-OMRON
DCL	Li 和 Yu(2016)	VGGNet	MSRA-B
MAP	Zhang 等人(2016)	VGGNet	SOS (Zhang 等,2015)
RFCN	Wang 等人(2018a)	VGGNet	MSRA10K
MDS	Li 等人(2016)	VGGNet	MSRA10K
DSS	Hou 等人(2017)	VGGNet	MSRA-B + HKU-IS
WSS	Wang 等人(2017)	VGGNet	ImageNet
Amulet	Zhang 等人(2017b)	VGGNet	MSRA10K
UCF	Zhang 等人(2017b)	VGGNet	MSRA10K
PiCANet	Liu 等人(2018)	VGGNet/ResNet50	DUTS
PAGRN	Zhang 等人(2018a)	VGGNet	DUTS
DGRL	Wang 等人(2018b)	ResNet50	DUTS
C2SNet	Li 等人(2018)	VGGNet	MSRA-B + Web
MWSNet	Zeng 等人(2019)	–	DUTS

注:“–”表示对应数据无法获取。

3.4.2　实验结果分析

本文算法与其他 19 种对比算法在 6 个数据集上的测试结果如表 3 所示。可以看出,本文模型在所有数据集上的综合排名最高。表明本文算法能够有效过滤多尺度特征中的无关项和噪声信息,有效抑制了负面信息对显著预测结果的影响。同时基于边界的混合损失也促进了模型生成清晰的显著对象边界。

图 6 展示了本文算法与 WSS、RFCN、MSRNet、UCF 和 Amulet 等代表算法的可视化结果。可以看出,与其他代表算法相比,本文模型的显著预测图最接近真实标签,能够准确定位到显著对象,同时具有平滑的边界。

3.4.3　模型拆分实验

为了验证深度聚类注意力机制对整体性能的贡献,将DCANet进行拆分,分别验证通道聚类注意力

表 3　不同算法在 6 个数据集上的定量实验结果对比

Table 3　Comparison of experimental results in six datasets among different algorithms

方法	DUTS-TE			SOD			ECSSD			DUT-OMRON			PASCAL-S			HKU-IS		
	F_β	S_λ	MAE	F_β	S_λ	MAE	F_β	S_λ	MAE	F_β	S_λ	MAE	F_β	S_λ	MAE	F_β	S_λ	MAE
HS	0.504	0.601	0.243	0.756	0.711	0.222	0.673	0.685	0.228	0.561	0.633	0.227	0.569	0.624	0.262	0.652	0.674	0.215
MCDL	0.634	0.713	0.105	0.689	0.651	0.182	0.816	0.803	0.101	0.670	0.752	0.089	0.706	0.721	0.143	0.787	0.786	0.092
LEGS	0.612	0.696	0.137	0.685	0.658	0.197	0.805	0.786	0.118	0.631	0.714	0.133	–	–	–	0.736	0.742	0.119
MDF	0.657	0.728	0.114	0.736	0.674	0.160	0.797	0.776	0.105	0.643	0.721	0.092	0.704	0.696	0.142	0.839	0.810	0.129
ELD	0.697	0.754	0.092	0.717	0.705	0.155	0.849	0.841	0.078	0.677	0.751	0.091	0.782	0.799	0.111	0.868	0.868	0.063
DHSNet	0.776	0.818	0.067	0.790	0.749	0.129	0.893	0.884	0.060	–	–	–	0.799	0.810	0.092	0.875	0.870	0.053
DCL	0.742	0.796	0.149	0.786	0.747	0.195	0.882	0.868	0.075	0.699	0.771	0.086	0.787	0.796	0.113	0.885	0.877	0.055
MAP	0.453	0.583	0.181	0.509	0.557	0.236	0.556	0.611	0.213	0.448	0.598	0.159	0.521	0.593	0.207	0.552	0.624	0.182
RFCN	0.755	0.829	0.090	0.769	**0.794**	0.170	0.875	0.852	0.107	0.707	0.764	0.111	0.800	0.798	0.132	0.881	0.859	0.089
DSS	0.796	0.824	0.057	0.805	0.751	0.122	0.906	0.882	0.052	0.737	0.790	0.063	0.805	0.798	0.093	–	–	–
WSS	0.778	0.822	0.079	0.807	0.675	0.170	0.879	0.811	0.104	0.725	0.730	0.110	0.804	0.744	0.139	0.878	0.822	0.079
Amulet	0.750	0.804	0.085	0.773	0.757	0.142	0.905	0.894	0.059	0.715	0.780	0.098	0.805	0.818	0.100	0.870	0.886	0.051
UCF	0.742	0.782	0.112	0.763	0.753	0.165	0.890	0.883	0.069	0.698	0.760	0.120	0.787	0.805	0.115	0.874	0.875	0.062
NLDF	0.777	0.816	0.065	**0.808**	0.789	0.125	0.889	0.875	0.063	0.699	0.770	0.080	0.795	0.805	0.098	0.888	0.879	0.048
PiCANet	0.804	0.844	0.053	0.770	0.732	0.130	0.887	0.888	0.059	0.710	0.789	0.067	0.792	0.822	0.850	0.871	0.877	0.051
PAGRN	0.817	0.838	0.056	–	–	–	0.904	0.889	0.061	0.707	0.775	0.071	0.814	0.822	0.089	0.897	0.887	0.048
DGRL	0.805	0.842	0.050	0.801	0.774	**0.103**	**0.914**	**0.903**	**0.041**	0.739	0.806	0.062	0.827	0.836	0.072	0.900	**0.895**	0.036
C2SNet	0.784	0.831	0.062	0.783	0.763	0.122	0.902	0.896	0.054	0.722	0.800	0.072	0.827	**0.839**	0.801	0.888	0.889	0.046
MWSNet	0.702	0.754	0.102	0.768	0.702	0.166	0.859	0.827	0.097	0.677	0.756	0.109	0.753	0.768	0.133	0.835	0.818	0.084
本文	**0.822**	**0.849**	**0.046**	0.800	0.769	0.104	**0.914**	0.900	0.043	**0.743**	**0.808**	**0.060**	**0.834**	0.832	**0.069**	**0.903**	0.893	**0.035**

注：加粗字体为各列最优结果；“–”表示对应数据无法获取。

(a) 输入图像　(b) 真值标签　(c) WSS　(d) RFCN　(e) MSRNet　(f) UCF　(g) Amulet　(h) 本文

图 6　本文算法与其他代表算法的显著预测图对比

Fig. 6　Comparison of saliency prediction images among our algorithm and others

((a) input images; (b) ground truth; (c) WSS; (d) RFCN; (e) MSRNet; (f) UCF; (g) Amulet; (h) ours)

机制和空间聚类注意力机制的作用,测试结果如表 4 所示。实验数据显示,通道聚类注意力机制使模型的性能提升了 1.7%～3.6%;加入空间聚类注意力机制后,模型性能提升了 1.3%～4.1%。综上所述,本文提出的 DCA 机制能够有效学习特征的内部关联,在求解注意力得分之前先执行显式类别划分能够在一定程度上促使模型更好地适应当前任务,强制推动不同类别关注不同模式的特征。

表 4　DCANet 模型消融实验

Table 4　Ablation study of DCANet

模型	S_λ	F_β	MAE
En-De backbone	0.869	0.878	0.102
En-De + CDCA	0.884	0.889	0.081
En-De + SDCA	0.897	0.898	0.056
En-De + CDCA + SDCA(本文)	**0.900**	**0.914**	**0.043**

注:加粗字体为各列最优结果。

4　结　论

本文针对图像的显著对象检测问题,提出了基于深度聚类注意力机制的鲁棒算法 DCANet。通过对特征通道和空间区域进行显式划分,在通道上进行前后景通道分离,使得关注背景的通道被有效抑制,而关注前景的通道类则具有明确的显著对象区域感知意图,这种注意力建模机制有助于模型训练过程中快速感知和适应任务。而在空间上执行全局和局部显著得分求解则有助于得到兼具全局一致性和局部连续性的显著图。在不同领域的公开基准数据集上与多种显著对象检测算法进行对比实验,证明了本文算法的有效性。但是,由于通道划分方法固定,空间局部信息的获取也基于固定大小的几个感受野,因此本文算法仍然具有一定的局限性。在后续工作中,将进一步研究更自适应的特征划分方式,以提升模型对不同场景的适应能力。

参考文献(References)

Borji A, Cheng M M, Jiang H Z and Li J. 2015. Salient object detection: a benchmark. IEEE Transactions on Image Processing, 24(12): 5706-5722 [DOI: 10.1109/TIP.2015.2487833]

Borji A, Frintrop S, Sihite D N and Itti L. 2012. Adaptive object tracking by learning background context//Proceedings of 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. Providence, USA: IEEE: 23-30 [DOI: 10.1109/CVPRW.2012.6239191]

Cong R M, Lei J J, Fu H Z, Cheng M M, Lin W S and Huang Q M. 2019. Review of visual saliency detection with comprehensive information. IEEE Transactions on Circuits and Systems for Video Technology, 29(10): 2941-2959 [DOI: 10.1109/TCSVT.2018.2870832]

Craye C, Filliat D and Goudou J F. 2016. Environment exploration for object-based visual saliency learning//Proceedings of 2016 IEEE International Conference on Robotics and Automation (ICRA). Stockholm, Sweden: IEEE: 2303-2309 [DOI: 10.1109/ICRA.2016.7487379]

Deng J, Dong W, Socher R, Li L J, Li K and Li F F. 2009. ImageNet: a large-scale hierarchical image database//Proceedings of 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami, USA: IEEE: 248-255 [DOI: 10.1109/CVPR.2009.5206848]

Fan D P, Cheng M M, Liu Y, Li T and Borji A. 2017. Structure-measure: a new way to evaluate foreground maps//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE: 4548-4557 [DOI: 10.1109/ICCV.2017.487]

Feng M Y, Lu H C and Ding E R. 2019. Attentive feedback network for boundary-aware salient object detection//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 1623-1632 [DOI: 10.1109/CVPR.2019.00172]

Han J W, Zhang D W, Wen S F, Guo L, Liu T M and Li X L. 2015. Two-stage learning to predict human eye fixations via SDAEs. IEEE Transactions on Cybernetics, 46(2): 487-498 [DOI: 10.1109/TCYB.2015.2404432]

Harel J, Koch C and Perona P. 2007. Graph-based visual saliency//Proceedings of the 19th International Conference on Neural Information Processing Systems. Vancouver, Canada: MIT Press: 545-552

Hou Q B, Cheng M M, Hu X W, Borji A, Tu Z W and Torr P. 2017. Deeply supervised salient object detection with short connections//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 3203-3212 [DOI: 10.1109/CVPR.2017.563]

Huang X, Shen C Y, Boix X and Zhao Q. 2015. SALICON: reducing the semantic gap in saliency prediction by adapting deep neural networks//Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago, Chile: IEEE: 262-270 [DOI: 10.1109/ICCV.2015.38]

Kingma D P and Ba J. 2014. Adam: a method for stochastic optimization [EB/OL]. [2020-02-25]. https://arxiv.org/pdf/1412.6980.pdf

Lee G, Tai Y W and Kim J. 2016. Deep saliency with encoded low level distance map and high level features//Proceedings of 2016 IEEE

Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 660-668 [DOI: 10.1109/CVPR.2016.78]

Li G B and Yu Y Z. 2015. Visual saliency based on multiscale deep features//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA: IEEE: 5455-5463 [DOI: 10.1109/CVPR.2015.7299184]

Li G B and Yu Y Z. 2016. Deep contrast learning for salient object detection//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 478-487 [DOI: 10.1109/CVPR.2016.58]

Li X, Yang F, Cheng H, Liu W and Shen D G. 2018. Contour knowledge transfer for salient object detection//Proceedings of the 15th European Conference on Computer Vision. Munich, Germany: Springer: 370-385 [DOI: 10.1007/978-3-030-01267-0_22]

Li X, Zhao L M, Wei L N, Yang M H, Wu F, Zhuang Y T, Ling H B and Wang J D. 2016. Deepsaliency: Multi-task deep neural network model for salient object detection. IEEE Transactions on Image Processing, 25(8): 3919-3930 [DOI: 10.1109/TIP.2016.2579306]

Li Y, Hou X D, Koch C, Rehg J M and Yuille A L. 2014. The secrets of salient object segmentation//Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, USA: IEEE: 280-287 [DOI: 10.1109/CVPR.2014.43]

Liu N and Han J W. 2016. DHSNet: deep hierarchical saliency network for salient object detection//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 678-686 [DOI: 10.1109/CVPR.2016.80]

Liu N, Han J W and Yang M H. 2018. PiCANet: learning pixel-wise contextual attention for saliency detection//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 3089-3098 [DOI: 10.1109/CVPR.2018.00326]

Movahedi V and Elder J H. 2010. Design and perceptual validation of performance measures for salient object segmentation//Proceedings of 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops. San Francisco, USA: IEEE: 49-56 [DOI: 10.1109/CVPRW.2010.5543739]

Qin X B, He S D, Yang X C, Dehghan M, Qin Q M and Martin J. 2018. Accurate outline extraction of individual building from very high-resolution optical images. IEEE Geoscience and Remote Sensing Letters, 15(11): 1775-1779 [DOI: 10.1109/LGRS.2018.2857719]

Shi J P, Yan Q, Xu L and Jia J Y. 2016. Hierarchical image saliency detection on extended CSSD. IEEE Transactions on Pattern Analysis and Machine Intelligence, 38(4): 717-729 [DOI: 10.1109/TPAMI.2015.2465960]

Simonyan K and Zisserman A. 2014. Very deep convolutional networks for large-scale image recognition [EB/OL]. [2020-02-13]. https://arxiv.org/pdf/1409.1556.pdf

Sun M J, Zhou Z Q, Hu Q H, Wang Z and Jiang J M. 2019. SG-FCN: a motion and memory-based deep learning model for video saliency detection. IEEE Transactions on Cybernetics, 49(8): 2900-2911 [DOI: 10.1109/TCYB.2018.2832053]

Wang L J, Lu H C, Ruan X and Yang M H. 2015. Deep networks for saliency detection via local estimation and global search//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA: IEEE: 3183-3192 [DOI: 10.1109/CVPR.2015.7298938]

Wang L J, Lu H C, Wang Y F, Feng M Y, Wang D, Yin B C and Ruan X. 2017. Learning to detect salient objects with image-level supervision//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA: IEEE: 136-145 [DOI: 10.1109/CVPR.2017.404]

Wang L Z, Wang L J, Lu H C, Zhang P P and Ruan X. 2018a. Salient object detection with recurrent fully convolutional networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 41(7): 1734-1746 [DOI: 10.1109/TPAMI.2018.2846598]

Wang T T, Zhang L H, Wang S, Lu H C, Yang G, Ruan X and Borji A. 2018b. Detect globally, refine locally: a novel approach to saliency detection//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 3127-3135 [DOI: 10.1109/CVPR.2018.00330]

Wang W G, Lai Q X, Fu H Z, Shen J B, Ling H B and Yang R G. 2019. Salient object detection in the deep learning era: an in-depth survey. IEEE Transactions on Pattern Analysis and Machine Intelligence: #9320524 [DOI: 10.1109/TPAMI.2021.3051099]

Yan Q, Xu L, Shi J P and Jia J Y. 2013. Hierarchical saliency detection//Proceedings of 2013 IEEE Conference on Computer Vision and Pattern Recognition. Portland, USA: IEEE: 1155-1162 [DOI: 10.1109/CVPR.2013.153]

Yang C, Zhang L H, Lu H C, Ruan X and Yang M H. 2013. Saliency detection via graph-based manifold ranking//Proceedings of 2013 IEEE Conference on Computer Vision and Pattern Recognition. Portland, USA: IEEE: 3166-3173 [DOI: 10.1109/CVPR.2013.407]

Yuan Y C, Li C Y, Kim J, Cai W D and Feng D D. 2018. Reversion correction and regularized random walk ranking for saliency detection. IEEE Transactions on Image Processing, 27(3): 1311-1322 [DOI: 10.1109/TIP.2017.2762422]

Zeng Y, Zhuge Y Z, Lu H C, Zhang L H, Qian M Y and Yu Y Z. 2019. Multi-source weak supervision for saliency detection//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 6074-6083 [DOI: 10.1109/CVPR.2019.00623]

Zhang J M, Ma S G, Sameki M, Sclaroff S, Betke M, Lin Z, Shen X H, Price B and Měch R. 2015. Salient object subitizing//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA: IEEE: 4045-4054 [DOI: 10.1109/CVPR.2015.7299031]

Zhang J M, Sclaroff S, Lin Z, Shen X H, Price B and Mech R. 2016. Unconstrained salient object detection via proposal subset optimization//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 5733-5742 [DOI: 10.1109/CVPR.2016.618]

Zhang L H, Yang C, Lu H C, Ruan X and Yang M H. 2017c. Ranking saliency. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39 (9): 1892-1904 [DOI: 10.1109/TPAMI.2016.2609426]

Zhang P P, Liu W, Lu H C and Shen C H. 2018a. Salient object detection by lossless feature reflection//Proceedings of the 27th International Joint Conference on Artificial Intelligence. Stockholm, Sweden: ACM: 1149-1155

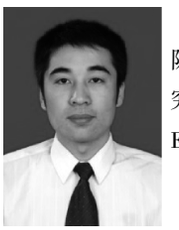
Zhang P P, Wang D, Lu H C, Wang H W and Ruan X. 2017a. Amulet: aggregating multi-level convolutional features for salient object detection//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE: 202-211 [DOI: 10.1109/ICCV.2017.31]

Zhang P P, Wang D, Lu H C, Wang H Y and Yin B C. 2017b. Learning uncertain convolutional features for accurate saliency detection//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE: 212-221 [DOI: 10.1109/ICCV.2017.32]

Zhao R, Ouyang W L and Wang X G. 2013. Unsupervised salience learning for person re-identification//Proceedings of 2013 IEEE Conference on Computer Vision and Pattern Recognition. Portland, USA: IEEE: 3586-3593 [DOI: 10.1109/CVPR.2013.460]

Zhao R, Ouyang W L, Li H S and Wang X G. 2015. Saliency detection by multi-context deep learning//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA: IEEE: 1265-1274 [DOI: 10.1109/CVPR.2015.7298731]

作者简介



陈庆文，1981 年生，男，高级工程师，主要研究方向为新能源和新型电力系统电力设计。
E-mail：18681880525@163.com

张雪，通信作者，女，硕士研究生，主要研究方向为计算机图像处理和深度学习。
E-mail：zhx95@tju.edu.cn

谢宏文，女，正高级工程师，主要研究方向为新能源和新型电力系统设计规划。E-mail：xhwgg@sina.com

查浩，男，正高级工程师，主要研究方向为新能源系统规划和消纳利用、综合能源微网和储能应用。
E-mail：26002834@qq.com

奚瑜，女，教授级高级工程师，主要研究方向为水电工程、新能源工程规划与设计。
E-mail：xiyucc@sina.com