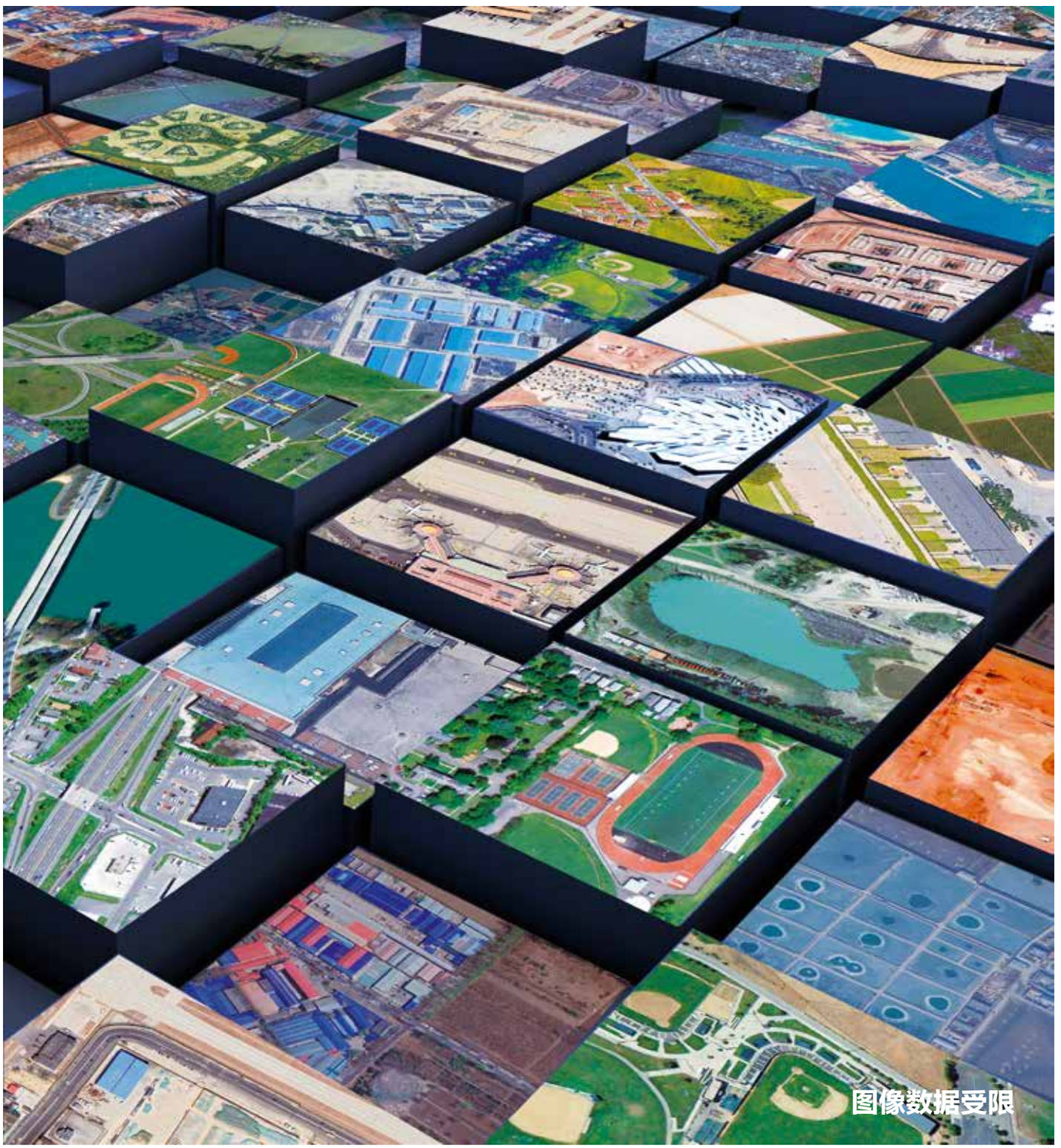
JOURNAL OF IMAGE AND GRAPHICS

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

2022
10
VOL.27

ISSN1006-8961
CN11-3758/TB



中国图象图形学报

刊名题字：宋健 | 月刊（1996年创刊）

第27卷第10期（总第318期）
2022年10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》编辑出版委员会
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@aircas.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 and Publishing Board 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@aircas.ac.cn
Telephone 010-58887035
Website www.cjig.cn

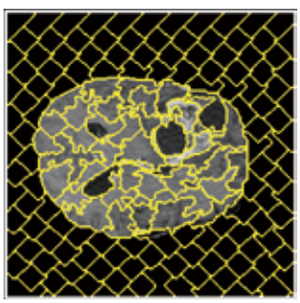
Distributed by Beijing Bureau for Distribution of Newspapers and Journals
Domestic All Local Post Offices in China
Overseas China International Book Trading Corporation
(P.O.Box 399, Beijing 100048,P.R.China))
Printed by Beijing Kexin Printing Co., Ltd.

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

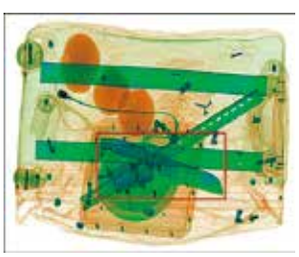
国外发行代号 M1406
国内邮发代号 82-831
国内定价 60.00元

2022年10月 第27卷 第10期（总第318期）
中国图象图形学报 Zhongguo Tuxiang Tuxing Xuebao

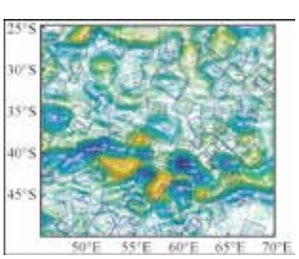
目次



MRI脑肿瘤图像的超像素/体素分割及发展现状(第2897页)



融合多尺度特征与全局上下文信息的X光违禁物品检测(第3043页)



融合多尺度旋转锚机制的海洋中尺度涡自动检测(第3092页)

图像数据受限

《中国图象图形学报》图像数据受限专栏简介
刘怡光，孙显，赵启军，魏秀参，王琦，陈秀妍 …… 2801

数据受限条件下的多模态处理技术综述
王佩瑾，闫志远，容雪娥，李俊希，路晓男，胡会扬，严启炜，孙显 …… 2803

图像数据受限下的处理与分析
刘怡光 …… 2835

面向跨模态行人重识别的单模态自监督信息挖掘
吴岸聪，林城栋，郑伟诗 …… 2843

小样本条件下的RGB-D显著性物体检测
何静，傅可人 …… 2860

综述

面向目标检测的对抗样本综述
袁珑，李秀梅，潘振雄，孙军梅，肖蕾 …… 2873

MRI脑肿瘤图像的超像素/体素分割及发展现状
方玲玲，王欣 …… 2897

图网络层级信息挖掘分类算法综述
魏文超，蔺广逢，廖开阳，康晓兵，赵凡 …… 2916

个性化图像美学评价的研究进展与趋势
祝汉城，周勇，李雷达，赵佳琦，杜文亮 …… 2937

视盘和视杯分割在计算机辅助青光眼诊断中的应用综述
方玲玲，张丽榕 …… 2952

图像处理和编码

多监督损失函数光滑化图像超分辨率重建
孟志青，张晶，邱健数 …… 2972

轻量级注意力约束对齐网络的视频超分重建
靳雨桐，宋慧慧，刘青山 …… 2984

面向图像修复的增强语义双解码器生成模型
王倩娜，陈燚 …… 2994

图像分析和识别

动态模态交互和特征自适应融合的RGBT跟踪
王福田，张淑云，李成龙，罗斌 …… 3010

采用Transformer网络的视频序列表情识别
陈港，张石清，赵小明 …… 3022

对抗型半监督光伏面板故障检测
卢芳芳，牛然，杜海舟，杨振辰，陈菁菁 …… 3031

融合多尺度特征与全局上下文信息的X光违禁物品检测
李晨，张辉，张邹铨，车爱博，王耀南 …… 3043

高速公路场景的车路视觉协同行车安全预警算法
汪长春，高尚兵，蔡创新，陈浩霖 …… 3058

结合卷积神经网络与曲线拟合的人体尺寸测量
马燕，殷志昂，黄慧，张玉萍 …… 3068

医学图像处理

U-Net支气管超声弹性图像纵膈淋巴结分割
刘羽，吴蓉蓉，唐璐，宋宁宁 …… 3082

遥感图像处理

融合多尺度旋转锚机制的海洋中尺度涡自动检测
杜艳玲，刘倩倩，王丽丽，徐鑫，魏泉苗，宋巍 …… 3092

集成注意力机制和扩张卷积的道路提取模型
王勇，曾祥强 …… 3102

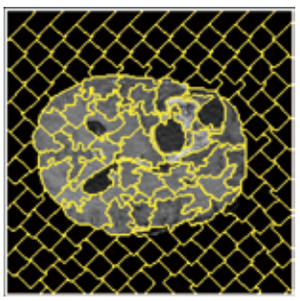
空域协同自编码器的高光谱异常检测
樊港辉，马泳，梅晓光，黄珺，樊凡，李晖 …… 3116

自适应权重金字塔和分支强相关的SAR图像舰船检测
郭伟，申磊，曲海成，王雅萱，林畅 …… 3127

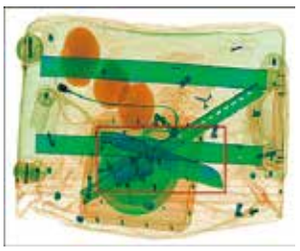
封面素材来自中国科学院空天信息创新研究院二部孙显研究员团队

Volume 27, Number 10
Published October 16, 2022

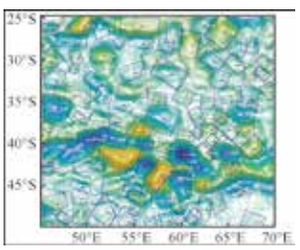
CONTENTS
JOURNAL OF IMAGE AND GRAPHICS



The review of superpixel/voxel segmentation of MRI brain tumor images(P2897)



Integrated multi-scale features and global context in x-ray detection for prohibited items(P3043)



Multi-scale rotating anchor mechanism based automatic detection of ocean mesoscale eddy(P3092)

Limited Image Data

Review of multimodal data processing techniques with limited data
Wang Peijin, Yan Zhiyuan, Rong Xuee, Li Junxi, Lu Xiaonan, Hu Huiyang, Yan Qiwei, Sun Xian ··· 2803

The processing and analyzing derived of limited image data
Liu Yiguang ··· 2835

Single-modality self-supervised information mining for cross-modality person re-identification
Wu Ancong, Lin Chengzhi, Zheng Weishi ··· 2843

RGB-D salient object detection of using few-shot learning
He Jing, Fu Keren ··· 2860

Review

Review of adversarial examples for object detection
Yuan Long, Li Xiumei, Pan Zhenxiong, Sun Junmei, Xiao Lei ··· 2873

The review of superpixel/voxel segmentation on MRI brain tumor images
Fang Lingling, Wang Xin ··· 2897

Survey of graph network hierarchical information mining for classification
Wei Wenchao, Lin Guangfeng, Liao Kaiyang, Kang Xiaobing, Zhao Fan ··· 2916

The review of personalized image aesthetics assessment
Zhu Hancheng, Zhou Yong, Li Leida, Zhao Jiaqi, Du Wenliang ··· 2937

The review of optic disc and optic cup segmentation applications in computer-aided glaucoma diagnosis
Fang Lingling, Zhang Lirong ··· 2952

Image Processing and Coding

Multi-supervision loss function based smoothed super-resolution image reconstruction
Meng Zhiqing, Zhang Jing, Qiu Jianshu ··· 2972

Super-resolution Video frame reconstruction through lightweight attention constraint alignment network
Jin Yutong, Song Huihui, Liu Qingshan ··· 2984

Enhanced semantic dual decoder generation model for image inpainting
Wang Qianna, Chen Yi ··· 2994

Image Analysis and Recognition

RGBT tracking based on dynamic modal interaction and adaptive feature fusion
Wang Futian, Zhang Shuyun, Li Chenglong, Luo Bin ··· 3010

Video sequence-based human facial expression recognition using Transformer networks
Chen Gang, Zhang Shiqing, Zhao Xiaoming ··· 3022

Generative adversarial networks based semi-supervised fault detection for photovoltaic panel
Lu Fangfang, Niu Ran, Du Haizhou, Yang Zhenchen, Chen Jingjing ··· 3031

Integrated multi-scale features and global context in x-ray detection for prohibited items
Li Chen, Zhang Hui, Zhang Zouquan, Che Aibo, Wang Yaonan ··· 3043

Vehicle-road visual cooperative driving safety early warning algorithm for expressway scenes
Wang Changchun, Gao Shangbing, Cai Chuangxin, Chen Haolin ··· 3058

The convolution neural network and curve fitting based human body size measurement
Ma Yan, Yin Zhiang, Huang Hui, Zhang Yuping ··· 3068

Medical Image Processing

U-Net-based mediastinal lymph node segmentation method in bronchial ultrasound elastic images
Liu Yu, Wu Rongrong, Tang Lu, Song Ningning ··· 3082

Remote Sensing Image Processing

Multi-scale rotating anchor mechanism based automatic detection of ocean mesoscale eddy
Du Yanling, Liu Qianqian, Wang Lili, Xu Xin, Wei Quanmiao, Song Wei ··· 3092

Road extraction model derived from integrated attention mechanism and dilated convolution
Wang Yong, Zeng Xiangqiang ··· 3102

Spatial-coordinated autoencoder for hyperspectral anomaly detection
Fan Ganghui, Ma Yong, Mei Xiaoguang, Huang Jun, Fan Fan, Li Hao ··· 3116

Ship detection in SAR images based on adaptive weight pyramid and branch strong correlation
Guo Wei, Shen Lei, Qu Haicheng, Wang Yaxuan, Lin Chang ··· 3127

E-mail: jig@aircas.ac.cn
Website: www.cjig.cn
Tel: 010-58887035

中国图象图形学报
JOURNAL OF IMAGE AND GRAPHICS

中图法分类号:TP391.4　　**文献标识码**:A　　**文章编号**:1006-8961(2022)10-2860-13

论文引用格式:He J and Fu K R. 2022. RGB-D salient object detection of using few-shot learning. Journal of Image and Graphics,27(10):2860-2872
(何静,傅可人. 2022. 小样本条件下的 RGB-D 显著性物体检测. 中国图象图形学报,27(10):2860-2872)[DOI:10.11834/jig.211068]

小样本条件下的 RGB-D 显著性物体检测

何静[1],傅可人[1,2*]

1. 四川大学视觉合成图形图像技术国防重点学科实验室,成都　610065;2. 四川大学计算机学院,成都　610065

摘　要:**目的**　现有基于 RGB-D(RGB-depth)的显著性物体检测方法往往通过全监督方式在一个较小的 RGB-D 训练集上进行训练,因此其泛化性能受到较大的局限。受小样本学习方法的启发,本文将 RGB-D 显著性物体检测视为小样本问题,利用模型解空间优化和训练样本扩充两类小样本学习方法,探究并解决小样本条件下的 RGB-D 显著性物体检测。**方法**　模型解空间优化通过对 RGB 和 RGB-D 显著性物体检测这两种任务进行多任务学习,并采用模型参数共享的方式约束模型的解空间,从而将额外的 RGB 显著性物体检测任务学习到的知识迁移至 RGB-D 显著性物体检测任务中。另外,训练样本扩充通过深度估计算法从额外的 RGB 数据生成相应的深度图,并将 RGB 图像和所生成的深度图用于 RGB-D 显著性物体检测任务的训练。**结果**　在 9 个数据集上的对比实验表明,引入小样本学习方法能有效提升 RGB-D 显著性物体检测的性能。此外,对不同小样本学习方法在不同的 RGB-D 显著性物体检测模型下(包括典型的中期融合模型和后期融合模型)进行了对比研究,并进行相关分析与讨论。**结论**　本文尝试将小样本学习方法用于 RGB-D 显著性物体检测,探究并利用两种不同小样本学习方法迁移额外的 RGB 图像知识,通过大量实验验证了引入小样本学习来提升 RGB-D 显著性物体检测性能的可行性和有效性,对后续将小样本学习引入其他多模态检测任务也提供了一定的启示。

关键词:多模态检测;RGB-D 显著性检测;小样本学习;多任务学习;深度估计

RGB-D salient object detection of using few-shot learning

He Jing[1], Fu Keren[1,2*]

1. National Key Laboratory of Fundamental Science on Synthetic Vision, Sichuan University, Chengdu　610065, China;
2. College of Computer Science, Sichuan University, Chengdu　610065, China

Abstract: **Objective** Salient object detection is mainly used in computer vision pre-processing like video/image segmentation, visual tracking and video/image compression. Current RGB-depth (RGB-D) salient object detection (SOD) can be categorized into fully supervision and self-supervision. Fully supervised RGB-D SOD can effectively fuse the complementary information of two different modes for RGB images input and the corresponding depth maps by means of the three types of fusion (early/middle/late). To capture contextual information, self-supervised salient object detection uses a small number of unlabeled samples for pre-training. However, existing RGB-D salient object detection methods are mostly trained on a small RGB-D training set in a fully supervised manner, so their generalization ability is greatly restricted. Thanks to the emerging few-shot learning methods, our RGB-D salient object detection uses model hypothesis space optimization and training sample augmentation to explore and solve RGB-D salient object detection with few-shot learning. **Method** For mod-

收稿日期:2021-11-24;**修回日期**:2022-04-08;**预印本日期**:2022-04-15

*通信作者:傅可人　fkrsuper@scu.edu.cn

基金项目:国家自然科学基金项目(62176169, 61703077);四川大学泸州战略合作项目(2020CDLZ-10)

Supported by: National Natural Science Foundation of China (62176169, 61703077); Sichuan University-Luzhou Municipal People's Government Strategic Cooperation Project (2020CDLZ-10)

el hypothesis space optimization, it can transfer the learned knowledge from extra RGB salient object detection task to RGB-D salient object detection task based on multi-task learning of RGB and RGB-D salient object detection tasks, and the hypothesis space of the model is constrained by sharing model parameters. Model-oriented, takeing into account middle and late fusions can add additional supervision to the network, therefore, the JL-DCF model is selected for middle fusion and the DANet† model is optioned for late fusion. To improve the effectiveness and generalization of RGB-D salient object detection tasks, RGB-D and RGB are simultaneously input into the network for online training and optimization in terms of JL-DCF, and the coarse prediction of RGB is supervised to optimize the network. In view of the commonality between the semantic segmentation and the saliency detection, the dual attention network for scene segmentation (DANet) model is transferred to the RGB-D salient object detection network, named DANet†. Similar to JL-DCF joint training, additional RGB supervision is added to the RGB branch of the two-stream DANet†. Furthermore, the training sample augmentation generates the related depth map based on the additional RGB data in terms of the depth estimation algorithm, and uses the RGB and the synthesized depth map for the training of the RGB-D salient object detection task. We adopt ResNet-101 as our network backbone. The scale of input image is $320 \times 320 \times 3$ in JL-DCF network, and the scale of DANet† network input image is fixed to $480 \times 480 \times 3$. The depth map is transformed into three-channel map by gray scale mapping. Our training set is composed of data from NJU2K, NLPR and DUTS, and the test set is NJU2K, NLPR, STERE, RGBD135, LFSD, SIP, DUT-RGBD, ReDWeb-S, DUTS (it is worth noting that, DUT-RGBD and ReDWeb-S are tested in the completed dataset based on 1 200 samples and 3 179 samples, respectively). The evaluation metrics are demonstrated as following: S measure (S_α), maximum F measure ($F_\beta^{\max}$), maximum E measure ($E_\varphi^{\max}$) and MAE (M). Our experiment is based on the Pytorch framework. The momentum parameter is 0.99, the learning rate is 0.000 05, and the weight decay is set to 0.000 5. Stochastic gradient descent learning technique is used to accelerate on NVIDIA RTX 2080S GPU. 1) Modeling: it takes about 20 hours to train 50 epochs. 2) Sampling: it takes about 100 hours to train 50 epochs and a weighting coefficient $\alpha = 2\ 200/10\ 553 \approx 0.21$ is illustrated to guarantee the roughly balanced in learning using the two different strategies. **Result** Our comparative experiments show that the introduction of few-shot learning methods on nine datasets can effectively improve the performance of RGB-D salient object detection. In addition, we compare different few-shot learning methods under different RGB-D salient object detection models (including typical middle-fusion model and late-fusion model), and draws relevant analysis and discussion. In addition, the visual saliency map shows its potential of our few-shot RGB-D saliency object detection method. **Conclusion** We facilitate the few-shot learning method for RGB-D salient object detection. It develops two different few-shot learning methods for transferring additional knowledge. Our research is beneficial to develop the subsequent introduction of few-shot learning towards more multi-modal detection tasks.

Key words: multi-modal detection; RGB-D saliency detection; few-shot learning; multi-task learning; depth estimation

0 引 言

显著性物体检测(salient object detection, SOD)旨在定位图像或视频中最吸引人注意力的物体,并将其从背景中分离出来。显著性物体检测主要应用于计算机视觉任务中的预处理,如视频/图像分割、视觉追踪、视频/图像压缩等。在早期,显著性物体检测主要基于 RGB 图像进行检测,从输入的 RGB 图像中提取有用信息用于物体显著程度的估计。近年来,随着深度传感器的发展和普及,基于 RGB-D(RGB-depth)的多模态显著性物体检测受到研究者们广泛的关注。

现有的 RGB-D SOD 方法按监督方式可以分为全监督和自监督两种。全监督 RGB-D SOD(Fu 等,2020;Zhang 等,2020)对输入的 RGB 图像以及相应的深度图通常采用早期融合、中期融合和后期融合的方式将两种不同模态的互补信息进行有效融合。自监督 RGB-D SOD(Zhao 等,2021)用少量无标记 RGB-D 数据集进行预训练,使网络捕获丰富的上下文语义信息,从而为下游任务提供有效初始化。

目前大多数 RGB-D SOD 采用全监督的方式在一个较小的 RGB-D SOD 训练集上进行训练,然而,此方式的泛化性能局限于较少的训练样本,难以泛化到真实场景。因此,本文提出将 RGB-D SOD 视为小样本学习问题。受 Wang 等人(2021)综述的启

发,本文应用两类小样本学习方法,第1类为基于模型解空间优化的方法,通过多任务训练以及参数共享的方式将训练样本数量较多的RGB SOD任务学习到的知识迁移至训练样本数量较少的RGB-D SOD任务,从模型角度约束特征解空间;第2类为基于训练样本扩充的方法,利用单图深度估计算法将额外的RGB图像生成相应的深度图,再将得到的RGB-D图像对用于训练样本扩充。通过对以上两类方法的结果进行对比分析,证明了引入小样本学习来提升RGB-D SOD性能的可行性和有效性。本文的主要贡献如下:

1)提出将RGB-D SOD视为小样本学习问题,根据小样本学习方法的分类,从模型解空间优化角度和训练样本扩充角度研究如何从RGB SOD任务迁移额外的先验知识,以提高小样本条件下的RGB-D SOD的性能和泛化性。与之前方法不同,本文从"训练样本少"的角度出发,利用小样本学习方法进行显著性物体检测的研究工作。

2)针对不同小样本学习方法,研究并实验了不同的显著性检测策略(包括典型的中期融合模型和后期融合模型),并在9个常用基准数据集上进行定量、定性的实验和分析,结果表明将RGB-D SOD视为小样本学习问题具有有效性和可行性。

1 相关工作

1.1 RGB-D显著性物体检测

近年来,RGB-D SOD在性能上取得了质的飞跃。传统的RGB-D SOD主要采用提取的手工特征将RGB图像信息与深度图信息进行融合。Niu等人(2012)提出第1个传统的基于RGB-D的显著性物体检测,利用全局视差对比和立体规则进行显著性估计。传统的RGB-D SOD模型,往往通过深度线索探索有用的属性,如边界线索、区域对比度、深度对比度和形状属性等。其中,Peng等人(2014)采用多阶段的RGB-D算法将深度和外观线索结合用于显著性物体的分割。值得一提的是,他们构建了第1个大规模的RGB-D SOD基准数据集,即NLPR。虽然传统的RGB-D SOD取得了不错的效果,但它们在复杂场景、低对比度和强光照等环境缺乏鲁棒性和泛化性。

Qu等人(2017)首次提出基于卷积神经网络的RGB-D显著性物体检测,利用卷积神经网络有效地学习输入图像的低级特征和深度线索,并通过卷积神经网络整合以获得最终的显著性检测结果,开启了基于深度神经网络的RGB-D SOD新方向。为充分利用RGB图与深度图的互补信息,CTMF方法(Han等,2018)利用卷积神经网络(convolutional neural network, CNN)学习RGB图像和深度图中的高级表示,将模型结构从RGB图像转移到深度图。Zhao等人(2019)提出一种流体金字塔集成模块,通过分层的方式有效融合跨模态信息。MMCI(Chen等,2019)利用多尺度多路的融合方式捕获RGB图像与多层深度线索之间的相关性。UC-Net(Zhang等,2020)提出通过条件变分自编码器对人的注释不确定性进行建模以产生不同的显著性预测,最终通过投票机制预测准确的显著性图。JL-DCF(Fu等,2020)将深度图与RGB图像进行级联输入到共享卷积神经网络进行特征提取,并提出一种密集协作融合策略,有效地融合不同模态学习到的特征。D3Net(Fan等,2021)通过判断深度图是否应该与RGB图像串联作为输入信号,设计网络以减少低质量深度图引入的噪声,并构造了一个新的RGB-D SOD基准数据集(SIP)。

由此可见,基于RGB-D的显著性物体检测在过去几年得到了快速发展,并获得较好的性能。但这些方法往往注重RGB与深度特征的有效融合(李贝等,2021),如设计早期融合、中期融合、晚期融合和多尺度融合等策略。而本文关注RGB-D SOD的训练样本较少,导致网络泛化能力具有一定局限性的问题。因此提出将RGB-D SOD视为小样本学习问题,研究如何将RGB SOD任务学习到的知识迁移到RGB-D SOD任务,并基于JL-DCF模型(Fu等,2020)和DANet(dual attention network for scene segmentation)模型(Fu等,2019),探讨引入小样本学习方法后,对RGB-D SOD带来的性能提升。

1.2 小样本学习

小样本学习任务旨在解决如何在监督信息有限的样本条件下增强目标任务的学习,通常见于小样本分类问题(徐鹏帮 等,2021),即N-way-K-shot问题。与小样本分类任务不同,本文利用RGB SOD任务与RGB-D SOD任务间的共性,解决RGB-D SOD监督信息有限的问题,增强RGB-D SOD任务的特征学习和泛化性。

目前，鲁棒的机器学习算法模型离不开大量的训练数据，但实际中训练样本的获取往往较难，小样本问题广泛存在于深度学习领域，因此近年来小样本学习方法成为热门方向，研究者们尝试探索小样本学习方法在不同领域的应用。小样本学习在特征识别（Finn 等，2017；Munkhdalai 和 Yu，2017；Snell 等，2017）和图像分类（Ravi 和 Larochelle，2017；Tsai 等，2017；Wang 和 Hebert，2016）的应用较广，在 Ominiglot 和 miniImageNet 两个基准数据集均取得较高的准确率。在视频方向也有较多应用，如视频分类（Zhu 和 Yang，2018）、动作预测（Gui 等，2018）、行人重识别（Wu 等，2018）、目标分割（Caelles 等，2017）等。尽管小样本学习方法应用于众多领域，但目前尚未有工作将小样本学习方法应用于显著性物体检测。与现有 RGB-D SOD 文献不同，本文发现并尝试解决 RGB-D SOD 的小样本问题。

2　基于小样本学习的 RGB-D SOD

本文在 Wang 等人（2021）综述的启发下，探索小样本条件下的 RGB-D SOD，研究两类不同的小样本学习方法在 RGB-D SOD 领域的综合性能表现，对基于两类小样本学习方法的 RGB-D SOD 进行对比分析。首先，从模型解空间优化角度，使 RGB-D SOD 任务和 RGB SOD 任务进行多任务学习共享权重参数，利用两个关联任务学习任务之间的共性，从模型角度约束参数，从而实现小样本条件下的 RGB-D SOD。从训练样本扩充角度，使用现有的单目深度估计算法生成相应的深度图，即直接利用 RGB SOD 数据集中的先验知识对数据进行增强，从而扩充小样本条件下的 RGB-D SOD 有监督数据。

RGB SOD 与 RGB-D SOD 的多任务学习方法需为额外的 RGB 图像进行监督，因此选择中期融合模型与后期融合模型作为本文框架。原因有：1）早期融合将 RGB 图像与深度图像在通道维度进行级联输入网络，或者将 RGB 图像与深度图像的浅层表示合并后输入网络进行显著性预测，在输入阶段将 RGB 图像与深度图进行级联，因此不能分别对 RGB 图像和深度图进行监督；2）中期融合将 RGB 图像与深度图像分别输入相应的网络，通过双流网络的方式获得特征，再将特征融合后输入深度神经网络解码器进行显著性预测，可为网络添加额外的监督信号；3）后期融合则利用双流网络分别提取 RGB 图像特征以及深度图像特征，将提取的特征联合用于最终的显著性预测。因此，由于采用了双流网络结构，中期融合和后期融合均可作为两类小样本学习方法的基本框架。

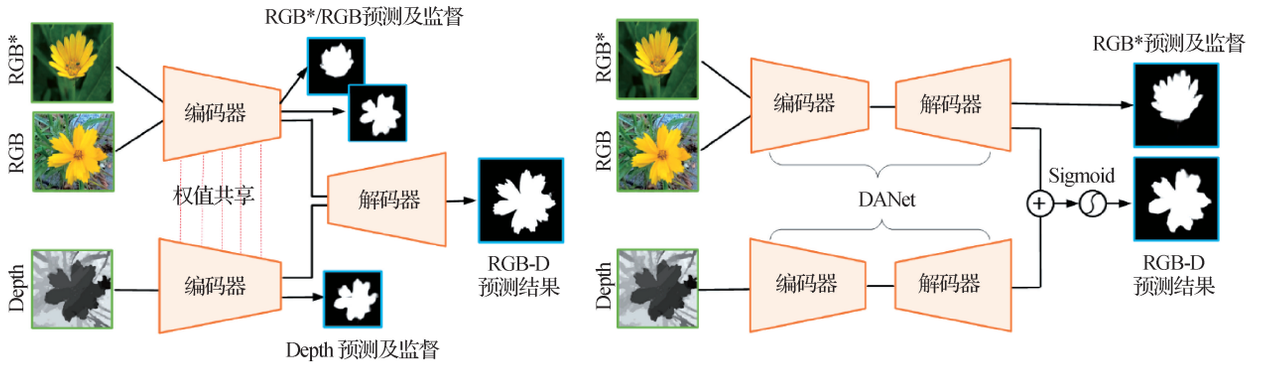
鉴于 RGB SOD 与 RGB-D SOD 的任务相似性以及 RGB SOD 的数据可用性，将 RGB SOD 任务学习到的知识迁移至 RGB-D SOD 任务中，解决 RGB-D SOD 的小样本问题。首先将该问题定义如下：视 RGB-D SOD 任务为任务 T，T 对应的数据集为 $\boldsymbol{D}=\{\boldsymbol{D}_{\text{train}},\boldsymbol{D}_{\text{test}}\}$，其中 $\boldsymbol{D}_{\text{train}}=\{(\boldsymbol{X}_{\text{RGB},D},\boldsymbol{Y})\}^{I}$，$\boldsymbol{D}_{\text{test}}=\{\boldsymbol{X}^{\text{test}}\}$，$\boldsymbol{Y}$ 表示 Ground Truth，I 表示数据集大小（即训练样本数量），此处 I 值较小，亦即训练样本较少。给定一个大得多的 RGB SOD 训练数据集 $\boldsymbol{D}^{*}_{\text{train}}$，$\boldsymbol{D}^{*}_{\text{train}}$ 的样本量远大于 $\boldsymbol{D}_{\text{train}}$。小样本条件下的 RGB-D SOD 问题则是探究如何使用小样本学习方法从 $\boldsymbol{D}^{*}_{\text{train}}$ 中引入额外的知识增强任务 T 的学习。

2.1　基于模型解空间优化

从模型解空间优化角度，小样本学习方法可以通过增加先验知识限制模型假设空间，使经验风险最小化的结果更可靠，并且降低过拟合风险（Wang 等，2021）。根据先验知识的利用方法，将基于模型的小样本学习方法分为多任务学习、嵌入学习和生成式模型（Wang 等，2021）。采用多任务学习方法，将两个相似任务进行参数共享，从而将 RGB SOD 任务的知识迁移至 RGB-D SOD 模型中。

考虑到从模型解空间优化角度进行多任务学习需要加入额外的监督信号，选择中期融合和后期融合模型对小样本 RGB-D SOD 进行探究。在中期融合模型中，Fu 等人（2020）提出的 JL-DCF 是具有代表性的中期融合模型，同时，JL-DCF 对 RGB 图像和深度图两种模态均有单独的监督；另外，JL-DCF 共享了 RGB 分支和深度分支的权重，使额外的 RGB 图像信息更好地增强两种模态的学习。对于后期融合模型，参考 Fu 等人（2021）将 DANet（Fu 等，2019）构造为双流后期融合模型 DANet†，用于多任务学习，框架图如图 1 所示。

图 1（a）表示基于中期融合的小样本条件下的 RGB-D SOD，网络主干部分为 JL-DCF 模型。JL-DCF（Fu 等，2020）通过孪生网络提取 RGB 图像与深度图像的特征，并提出密集协作融合策略有效地



(a) 基于中期融合的多任务学习　　(b) 基于后期融合的多任务学习

图 1　将多任务学习用于基于中期融合和后期融合的 RGB-D SOD 模型(RGB* 表示额外的 RGB 图像)

Fig. 1　Applying multi-task learning to RGB-D SOD models that are based on middle fusion and late fusion (RGB* denotes the extra RGB images) ((a) multi-task learning based on middle fusion; (b) multi-task learning based on late fusion)

融合不同模态的特征。本文在编码模块对 RGB-D SOD 任务和 RGB SOD 任务进行参数共享,为引导网络更好地学习多任务特征,将特征编码器输出的粗略显著图进行监督从而优化编码模块以提高模块的泛化能力。解码器将 RGB-D 数据编码的各级特征与解码部分的各级特征进行跨模块融合,最后输出精确的显著预测图。其中,RGB-D SOD 任务的训练数据远小于 RGB SOD 任务的训练数据。

图 1(b)表示基于后期融合的小样本条件下的 RGB-D SOD,网络主干部分为 DANet 模型。DANet(Fu 等,2019)为基于 RGB 的语义分割模型,通过多尺度特征融合捕获上下文信息,同时采用双注意力网络以自适应地将局部特征与其全局依赖性相结合,分别对空间和通道维度的语义相互依赖性进行建模。本文参考 Fu 等人(2021)将语义分割模型 DANet 的分类预测头卷积层($1\times1, C$)(输出通道数 C 表示语义分割类别数)替换为($1\times1, 1$)的预测卷积层以用于显著性物体检测。由于 DANet 为基于 RGB 的单流模型,因此将 DANet 修改为输入为 RGB 图像和深度图像的双流后期融合模型,即 DANet†。对输入的 RGB 图像与深度图像进行编解码操作得到单通道激活特征图,再在输入 Sigmoid 函数前进行相加融合操作得到最终的显著性图。

对图 1(a)(b),以 $\boldsymbol{D}_{\text{train}}^{\text{RGB}}$、$\boldsymbol{D}_{\text{train}}^{\text{Depth}}$ 分别表示输入的 RGB-D 数据集中的 RGB 图像集与深度图像集(即 $\boldsymbol{D}_{\text{train}}=\{\boldsymbol{D}_{\text{train}}^{\text{RGB}},\boldsymbol{D}_{\text{train}}^{\text{Depth}}\}$)。为实现 RGB-D SOD 与 RGB SOD 的多任务学习,将 $\boldsymbol{D}_{\text{train}}^{*}$(即额外的 RGB 数据集)与 $\boldsymbol{D}_{\text{train}}$ 同时输入编码器,同步训练两个不同来源的数据,具体见图 1。

综上,基于中期融合和后期融合的多任务学习方法的总体损失分别表示为

$$L_1=L(S_r,G)+L(S_d,G)+L(S_{r^*},G^*)+L(S_f,G) \tag{1}$$

$$L_2=L(S_{r^*},G^*)+L(S_f,G) \tag{2}$$

式中,S_r、S_d、S_{r^*} 分别表示数据来源为 $\boldsymbol{D}_{\text{train}}^{\text{RGB}}$、$\boldsymbol{D}_{\text{train}}^{\text{Depth}}$ 以及 $\boldsymbol{D}_{\text{train}}^{*}$ 的显著性预测结果,S_f 为最终的 RGB-D 显著性预测结果,$L(\cdot,\cdot)$ 表示交叉熵损失函数,G 为 $\boldsymbol{D}_{\text{train}}$ 的 ground truth,G^* 表示 $\boldsymbol{D}_{\text{train}}^{*}$ 的 ground truth。

2.2　基于训练样本扩充

目前,在显著性物体检测领域,RGB SOD 数据集的大小约为 RGB-D SOD 数据集的 5 ~ 10 倍(Wang 等,2022),本文从数据扩充角度,通过单目深度估计算法生成 RGB 图像相应的深度图,并将 RGB-D 图像对用于显著性物体检测。选择两种前沿方法 DPT(vision transformers for dense prediction)(Ranftl 等,2021)和 MegaDepth(Li 和 Snavely,2018)用于单目深度估计,在 Ranftl 等人(2021)的对比实验中表明,DPT 方法相比于 MegaDepth 方法能生成更精细、视觉效果更好的视差图。DPT 利用视觉 Transformer 代替卷积网络作为密集预测任务的主干网络,通过视觉 Transformer 和卷积解码器产生全分辨率的预测。MegaDepth 采用 SFM + MVS(structure-from-motion + multi-view stereo)(Li 和 Snavely,2018)训练网络产生新的位置和场景,最终产生稠密的相

对深度数据。DPT 与 MegaDepth 方法均以视差图的形式生成深度信息，因此实验时，将生成的视差图进行归一化和反转得到相应的深度图。将生成的深度图与 $\boldsymbol{D}^{*}_{\text{train}}$ 中的图像联合作为训练数据（简称为合成数据），在 JL-DCF 网络和 DANet† 网络上进行对比实验。

当加入了额外的合成数据，数据集的分布将发生改变，即小部分为原始 RGB-D 数据，大部分则为合成数据，因此为了减小数据量对网络训练的影响，深度生成时，将合成数据的训练损失按数据比例进行加权，因此深度生成方法的损失函数为

$$L_{\text{total}} = \sum_{x \in \boldsymbol{D}_{\text{train}}} L(x) + \alpha \sum_{x \in \boldsymbol{D}^{*}_{\text{train}}} L(x) \tag{3}$$

式中，L_{total} 表示总损失，$\boldsymbol{D}_{\text{train}}$、$\overline{\boldsymbol{D}}^{*}_{\text{train}}$ 分别表示原始 RGB-D 训练集与从 $\boldsymbol{D}^{*}_{\text{train}}$ 生成的合成训练集，在中期融合模型与后期融合模型中，$L(x)$ 分别对应式（1）和式（2），但均不包含其中的 $L(S_{r^*}, G^{*})$ 项。

3　实验结果分析

3.1　数据集与评估指标

为公平比较，在 8 个 RGB-D SOD 数据集以及 1 个 RGB SOD 数据集上进行实验，并对实验结果进行评价分析。RGB-D SOD 数据集包括：NJU2K（1 985 个样本）（Ju 等，2014）、NLPR（1 000 个样本）（Peng 等，2014）、STERE（1 000 个样本）（Niu 等，2012）、RGBD135（135 个样本）（Cheng 等，2014）、LFSD（100 个样本）（Li 等，2017）、SIP（929 个样本）（Fan 等，2021）、DUT-RGBD（800 个训练样本 + 400 个测试样本）（Piao 等，2019）、ReDWeb-S（2 179 个训练样本 + 1 000 个测试样本）（Liu 等，2021）。DUTS（10 553 个训练样本 + 5 019 个测试样本）（Wang 等，2017）为 RGB SOD 数据集。在本文中，RGB-D 训练集由 NJU2K 的 1 500 个样本、NLPR 的 700 个样本组成，额外的 RGB 训练数据由 DUTS 中 10 553 个样本组成。其余数据用于测试，值得一提的是，在 DUT-RGBD 和 ReDWeb-S 中，采用所有数据进行测试，即测试集分别包含 1 200 个和 3 179 个样本。

评价指标采用显著性物体检测领域中的常用指标，即 S-measure（S_{α}）（Fan 等，2017）、最大 F-measure（$F^{\max}_{\beta}$）（Borji 等，2015）、最大 E-measure（$F^{\max}_{\phi}$）（范登平 等，2021）和 MAE（M）（Perazzi 等，2012；Borji 等，2015）。

3.2　实验细节

本文方法的实现基于 JL-DCF 框架和 DANet† 框架。在基于 JL-DCF 的多任务学习实验中，将 RGB-D 数据与 DUTS 中的 RGB 数据同时输入编码器进行在线联合训练和优化。相似地，基于 DANet† 的多任务学习实验对两种不同来源的 RGB 数据同时进行联合训练和优化。JL-DCF 框架和 DANet† 框架的主干网络均为 ResNet-101，其中 JL-DCF 网络与 DANet† 网络输入图像的尺寸分别为 320 × 320 × 3、480 × 480 × 3，最终输出图像分辨率分别为 320 × 320 像素、480 × 480 像素，两个网络的输入均是将深度图通过简单的灰度映射转换为三通道图。

将 NJU2K 的 1 500 个 RGB-D 样本和 NLPR 的 700 个 RGB-D 样本作为 $\boldsymbol{D}_{\text{train}}$，将 DUTS 的 10 553 幅 RGB 图像作为 $\boldsymbol{D}^{*}_{\text{train}}$ 用于网络训练，将其余数据集作为 $\boldsymbol{D}_{\text{test}}$。训练时，对训练集中的样本数据采用镜面翻转的方法进行数据增强。实验方法基于 Pytorch 框架，动量参数设置为 0.99，学习率为 0.000 05，权重衰减为 0.000 5，训练时采用随机梯度下降法在 NVIDIA RTX 2080S GPU 上加速。基于模型解空间优化实验中，训练 50 个 Epoch 大约需要 20 h。基于训练样本扩充实验中，由于增加 10 553 个生成的 RGB-D 样本数据，在 Batch_size 为 1 的情况下，Epoch 为 50，大约需要 100 h（超 5 倍的数据量）。实验中，对式（3）中的权重系数设置为 $\alpha = 2\ 200/10\ 553 \approx 0.21$，以此来保证训练对数据分布的平衡性。

3.3　实验结果对比与分析

3.3.1　定量结果对比和分析

为直观地说明将 RGB-D SOD 视为小样本学习问题的有效性和泛化性，训练了 6 个不同模型验证两种不同的小样本学习方法（即 RGB SOD 与 RGB-D SOD 的多任务学习，以及训练样本深度生成）对 RGB-D SOD 的性能提升。如表 1 所示，其中 W/o FSL 表示原始 Pytorch 版本的模型性能（未采用小样本学习方法），Multi-task 表示 RGB SOD 与 RGB-D SOD 多任务学习优化的方法，DS-DPT、DS-MD 分别表示基于 DPT 和 MegaDepth 的深度生成的训练样本扩充。值得一提的是，表 1 中 DUTS 为 RGB SOD 数据集，因此在测试时通过 DPT 和 MegaDepth 两种方法生成深度图，分别表示为 DUTS（DPT）、DUTS（MD）。

表 1　在 9 个数据集上的定量分析

Table 1　Quantitative comparison on nine benchmark datasets

数据集	指标	SOTAs		JL-DCF				DANet†			
		UCNet	SSRNet	W/o FSL	multi-task	DS-DPT	DS-MD	W/o FSL	multi-task	DS-DPT	DS-MD
NJU2K	$S_\alpha\uparrow$	0.897	0.897	0.917	**0.922**	0.921	**0.922**	0.900	0.907	**0.914**	0.907
	$F_\beta^{\max}\uparrow$	0.895	0.893	0.919	0.925	0.924	**0.926**	0.897	0.908	**0.914**	0.904
	$E_\phi^{\max}\uparrow$	0.936	0.936	0.950	**0.956**	**0.956**	0.954	0.938	0.945	**0.950**	0.942
	$M\downarrow$	0.043	0.047	0.037	**0.034**	**0.034**	**0.034**	0.044	0.042	**0.038**	0.042
NLPR	$S_\alpha\uparrow$	0.920	0.915	0.931	**0.937**	0.935	0.934	0.912	0.920	**0.924**	0.923
	$F_\beta^{\max}\uparrow$	0.903	0.901	0.920	**0.927**	0.925	0.922	0.892	0.906	**0.909**	0.908
	$E_\phi^{\max}\uparrow$	0.956	0.953	0.964	**0.969**	**0.969**	0.964	0.949	0.960	**0.961**	0.958
	$M\downarrow$	0.025	0.028	0.022	**0.020**	**0.020**	0.021	0.027	0.027	**0.023**	0.026
STERE	$S_\alpha\uparrow$	0.903	0.892	0.906	0.908	**0.911**	0.903	0.889	0.895	0.901	**0.904**
	$F_\beta^{\max}\uparrow$	0.899	0.882	0.903	0.903	**0.908**	0.898	0.874	0.887	0.890	**0.892**
	$E_\phi^{\max}\uparrow$	0.944	0.930	0.946	0.947	**0.949**	0.942	0.930	**0.939**	0.938	**0.939**
	$M\downarrow$	0.039	0.048	0.040	**0.039**	**0.039**	0.042	0.048	0.044	**0.042**	**0.042**
RGBD135	$S_\alpha\uparrow$	0.934	0.905	0.934	0.935	**0.942**	0.938	0.896	**0.922**	0.921	0.912
	$F_\beta^{\max}\uparrow$	0.930	0.895	0.928	0.931	**0.938**	0.932	0.875	0.908	**0.911**	0.900
	$E_\phi^{\max}\uparrow$	0.976	0.958	0.967	0.965	**0.975**	0.973	0.935	**0.959**	0.957	0.942
	$M\downarrow$	0.019	0.028	0.020	0.018	**0.017**	0.018	0.027	0.023	**0.022**	0.026
LFSD	$S_\alpha\uparrow$	0.864	0.845	0.862	**0.876**	0.868	0.858	0.836	0.847	**0.849**	0.846
	$F_\beta^{\max}\uparrow$	0.864	0.846	0.861	**0.881**	0.870	0.864	0.829	**0.848**	0.838	0.839
	$E_\phi^{\max}\uparrow$	0.905	0.886	0.894	**0.913**	0.900	0.895	0.873	**0.893**	0.878	0.880
	$M\downarrow$	0.066	0.082	0.074	**0.063**	0.070	0.076	0.090	**0.079**	0.081	0.085
SIP	$S_\alpha\uparrow$	0.875	0.878	0.879	**0.900**	0.895	0.887	0.870	0.885	**0.888**	0.883
	$F_\beta^{\max}\uparrow$	0.879	0.884	0.889	**0.915**	0.906	0.896	0.865	0.885	**0.887**	0.882
	$E_\phi^{\max}\uparrow$	0.919	0.921	0.925	**0.943**	0.935	0.928	0.916	0.927	**0.930**	0.928
	$M\downarrow$	0.051	0.054	0.050	**0.039**	0.043	0.047	0.056	0.050	**0.048**	0.051
DUT-RGBD	$S_\alpha\uparrow$	0.847	0.838	0.876	**0.900**	0.895	0.897	0.821	**0.873**	0.860	0.862
	$F_\beta^{\max}\uparrow$	0.834	0.818	0.867	**0.897**	0.889	0.892	0.795	**0.863**	0.843	0.844
	$E_\phi^{\max}\uparrow$	0.889	0.876	0.911	**0.932**	0.926	**0.932**	0.860	**0.910**	0.892	0.893
	$M\downarrow$	0.069	0.072	0.056	**0.044**	0.047	0.045	0.086	**0.059**	0.066	0.067
ReDWeb-S	$S_\alpha\uparrow$	0.705	0.685	0.714	0.723	**0.726**	0.719	0.712	0.700	0.718	**0.719**
	$F_\beta^{\max}\uparrow$	0.702	0.674	0.710	**0.721**	0.716	0.719	0.692	0.702	0.705	**0.710**
	$E_\phi^{\max}\uparrow$	0.791	0.772	0.796	**0.804**	0.799	0.803	0.781	0.795	0.794	**0.799**
	$M\downarrow$	0.134	0.148	0.134	0.129	**0.128**	0.131	0.136	0.141	**0.130**	0.133
DUTS (DPT)	$S_\alpha\uparrow$	0.824	0.821	0.842	0.866	**0.880**	0.871	0.823	0.853	**0.878**	0.870
	$F_\beta^{\max}\uparrow$	0.778	0.765	0.800	0.834	**0.854**	0.841	0.764	0.817	**0.847**	0.835
	$E_\phi^{\max}\uparrow$	0.876	0.866	0.887	0.909	**0.918**	0.911	0.868	0.907	**0.922**	0.914
	$M\downarrow$	0.060	0.062	0.054	0.044	**0.042**	0.044	0.061	0.046	**0.041**	0.043
DUTS (MD)	$S_\alpha\uparrow$	0.813	0.803	0.824	0.855	0.854	**0.866**	0.816	0.852	**0.872**	0.869
	$F_\beta^{\max}\uparrow$	0.766	0.745	0.777	0.823	0.821	**0.834**	0.755	0.817	**0.840**	0.834
	$E_\phi^{\max}\uparrow$	0.866	0.854	0.873	0.905	0.897	**0.907**	0.861	0.907	**0.918**	0.912
	$M\downarrow$	0.064	0.068	0.062	**0.048**	0.053	**0.048**	0.065	0.048	**0.043**	0.044

注：加粗字体分别表示基于 JL-DCF 和 DANet† 模型的小样本学习方法的最优值；↑/↓表示越大/越小性能越好；SOTAs 表示 state-of-the-arts，即前沿方法。

从表 1 可得：

1）由 JL-DCF 与 DANet†的结果可得，将 RGB-D SOD 视为小样本学习问题并引入小样本学习方法可提高模型性能。例如，对于 JL-DCF 模型，多任务学习方法（multi-task）在 SIP 和 DUT-RGBD 数据集性能表现总体最好，S_α 的提升分别为 2.1%、2.4%；对于 DANet†模型，基于 DPT 的深度生成方法（DS-DPT）在 SIP 和 DUT-RGBD 数据集上，S_α 的提升分别为 1.8%、3.9%。

2）在 JL-DCF 结果中，多任务学习方法（multi-task）的性能表现最佳，相比于未引入小样本学习（W/o FSL）的性能总体提升最高，这源于 JL-DCF 模型通过参数共享的方式将 RGB SOD 任务的知识迁移至 RGB-D SOD 任务，此方式更有利于显著性物体检测任务的特征学习。

3）在 DANet†结果中，基于 DPT 的深度生成（DS-DPT）性能总体提升最高，且高于多任务学习（multi-task）。原因在于采用 DPT 方法合成的高质量深度图进行网络训练，使基于 DPT 的深度生成方法（DS-DPT）性能提升最高。另一方面，DANet†通过双流网络分别学习 RGB 图像与深度图像的特征，并以后期融合的方式融合两支路（RGB 分支、深度图分支）的特征，在特征学习阶段额外的 RGB 图像信息仅有利于 RGB 分支的学习，而深度分支未能利用额外的 RGB 图像信息，因此多任务学习方法性能表现稍差。

4）分析 JL-DCF 和 DANet†数据中 DS-DPT 与 DS-MD 的性能表现，结果显示两类模型中 DS-DPT 的总体性能均要优于 DS-MD，可得出深度生成的质量对结果有一定的影响，即深度生成算法效果越好，引入额外的 RGB 图像知识所带来的性能提升越大。

5）JL-DCF 的整体性能优于 DANet†，因此小样本条件下的 RGB-D 显著性物体检测依赖于模型的选择。与 UCNet（Zhang 等，2020）、SSRNet（Zhao 等，2020）两种现有前沿方法进行对比，引入小样本学习方法后可获得优于 SOTA（state-of-the-art）的性能。同时，在 DUT-RGBD、ReDWeb-S 两个全数据集上的测试结果证明了在 RGB-D SOD 模型引入小样本学习方法的泛化性。

为直观表现采用小样本学习方法对 JL-DCF 与 DANet†模型的性能提升，对表 1 中数据进行统计归纳，得出将 RGB-D SOD 视为小样本学习问题后，多任务学习以及深度生成方法在 8 个通用数据集的性能提升（仅计算 NJU2K、NLPR、STERE、RGBD135、LFSD、SIP、DUT-RGBD、ReDWeb-S 数据集上的指标提升平均值），如表 2 所示，在 JL-DCF 模型中，多任务学习方法具有突出的性能表现，在 DANet†模型中，基于 DPT 方法的深度生成性能提升较为突出，如前所述，此结果与基础模型的结构相关。另外，JL-DCF 模型引入小样本学习方法的性能提升要小于 DANet† 模型引入小样本学习方法的提升，原因为原始 JL-DCF 性能表现已较好，而 DANet†模型性能稍差，说明小样本学习方法对模型带来的性能提升一定程度上取决于模型自身的基础性能。表 2 再次证明了将 RGB-D SOD 视为小样本学习问题的可行性和有效性。

表 2　小样本学习方法的平均性能提升

Table 2　Average improvement for few-shot learning methods

/%

指标	JL-DCF			DANet†		
	multi-task	DS-DPT	DS-MD	multi-task	DS-DPT	DS-MD
S_α ↑	**+1.0**	+0.9	+0.5	+1.4	**+2.1**	+1.5
$F_\beta^{\max}$ ↑	**+1.1**	+1.0	+0.7	**+2.4**	+2.2	+2.0
$E_\phi^{\max}$ ↑	**+1.0**	+0.7	+0.5	**+1.8**	+1.5	+1.2
M ↓	**−0.6**	−0.4	−0.2	−0.6	**−0.8**	−0.5

注：加粗字体分别表示基于 JL-DCF 和 DANet† 模型的小样本学习方法的最优值；↑/↓ 表示越大/越小性能越好。

表 3 展示了将模型用于 DUTS 数据集的性能提升，DUTS（DPT）、DUTS（MD）分别表示采用 DPT 方法和 MegaDepth 方法生成 DUTS 数据集的深度图用于测试。在 JL-DCF 模型中，采用 DPT 方法生成深度图训练的模型在 DPT 方法生成深度图的测试集上性能表现最佳，采用 MegaDepth 方法生成深度图训练

的模型在 MegaDepth 方法生成深度图的测试集上性能表现最佳。对于 DANet† 模型，DPT 方法生成深度图训练的模型（即 DS-DPT）在 DUTS（DPT）与 DUTS（MD）数据集上性能表现最佳，原因在于 DANet† 为后期融合模型，对 RGB-D 图像对的语义信息的利用较差，因此深度图的质量对此双流网络的影响较大，印证了基于训练样本扩充的小样本学习方法的性能依赖于深度生成算法的性能。另外，由于额外的 RGB 图像仅有利于 RGB 分支学习特征，而没有学习生成的深度图信息，因此多任务学习（即 multi-task）性能差于深度生成方法。总之，采用不同深度生成方法训练的模型能够在不同的测试集（DUTS（DPT）、DUTS（MD））取得较优的性能提升（S_α 最低提升 2.9%），证明引入小样本学习方法可提高模型的泛化性。

表 3　DUTS 数据集上小样本学习方法的性能提升

Table 3　Performance of few-shot learning method on DUTS

方法	数据集	指标	multi-task	DS-DPT	DS-MD
JL-DCF	DUTS（DPT）	S_α ↑	+0.024	**+0.038**	+0.029
		$F_\beta^{\max}$ ↑	+0.034	**+0.054**	+0.041
		$E_\phi^{\max}$ ↑	+0.022	**+0.031**	+0.024
		M ↓	−0.010	**−0.012**	−0.010
	DUTS（MD）	S_α ↑	+0.031	+0.030	**+0.042**
		$F_\beta^{\max}$ ↑	+0.046	+0.044	**+0.057**
		$E_\phi^{\max}$ ↑	+0.032	+0.024	**+0.034**
		M ↓	**−0.014**	−0.009	**−0.014**
DANet†	DUTS（DPT）	S_α ↑	+0.030	**+0.055**	+0.047
		$F_\beta^{\max}$ ↑	+0.053	**+0.083**	+0.071
		$E_\phi^{\max}$ ↑	+0.039	**+0.054**	+0.046
		M ↓	−0.015	**−0.020**	−0.018
	DUTS（MD）	S_α ↑	+0.036	**+0.056**	+0.053
		$F_\beta^{\max}$ ↑	+0.062	**+0.085**	+0.079
		$E_\phi^{\max}$ ↑	+0.046	**+0.057**	+0.051
		M ↓	−0.017	**−0.022**	−0.021

注：加粗字体为各行最优结果，↑/↓表示越大/越小性能越好。

为验证引入小样本学习对 RGB-D SOD 在训练样本数量极少时的优越性，本文将 RGB-D SOD 训练样本数按 1/4 进行指数式减少（即从 2 200 依次减少为 550、138、35 个 RGB-D 训练样本），而额外的 RGB SOD 训练样本数量则保持不变，以使 RGB SOD 数据量远大于 RGB-D SOD 数据量。本文选择基于 JL-DCF 的多任务学习方法进行验证。如表 4 所示，其中 Δ_1、Δ_2、Δ_3、Δ_4 分别表示样本数量为 2 200、550、138、35 时，多任务学习方法在 9 个数据集上的平均提升（基准模型 W/o FSL 也使用减少的样本进行了重新训练）。由表 4 给出的实验结果可知，当 RGB-D SOD 训练样本数为 2 200 和 550 时，引入小样本

表 4　指数减少 RGB-D 数据量时多任务学习方法的平均性能提升

Table 4　Performance improvement of multi-task learning after reduction of RGB-D SOD samples

指标	JL-DCF（multi-task）			
	Δ_1	Δ_2	Δ_3	Δ_4
S_α ↑	+0.014	+0.014	+0.036	**+0.066**
$F_\beta^{\max}$ ↑	+0.018	+0.018	+0.053	**+0.089**
$E_\phi^{\max}$ ↑	+0.013	+0.012	+0.034	**+0.058**
M ↓	−0.007	−0.011	−0.016	**−0.031**

注：加粗字体为各行最优结果，↑/↓表示越大/越小性能越好。

学习方法对该任务的性能提升相当，但随着样本数量的指数减少，多任务学习方法的性能提升越发显著。

3.3.2　定性结果对比和分析

由图 2 可得，小样本条件下的显著性物体检测准确率更高。同时，对于背景较复杂的图像也可以准确地检测出显著性物体。本文方法在 DANet† 模型与 JL-DCF 模型上均能明显提高检测准确率。由此证明了小样本条件下的 RGB-D SOD 的可行性，表现为所得到的显著性物体更加完整，置信度也更高。

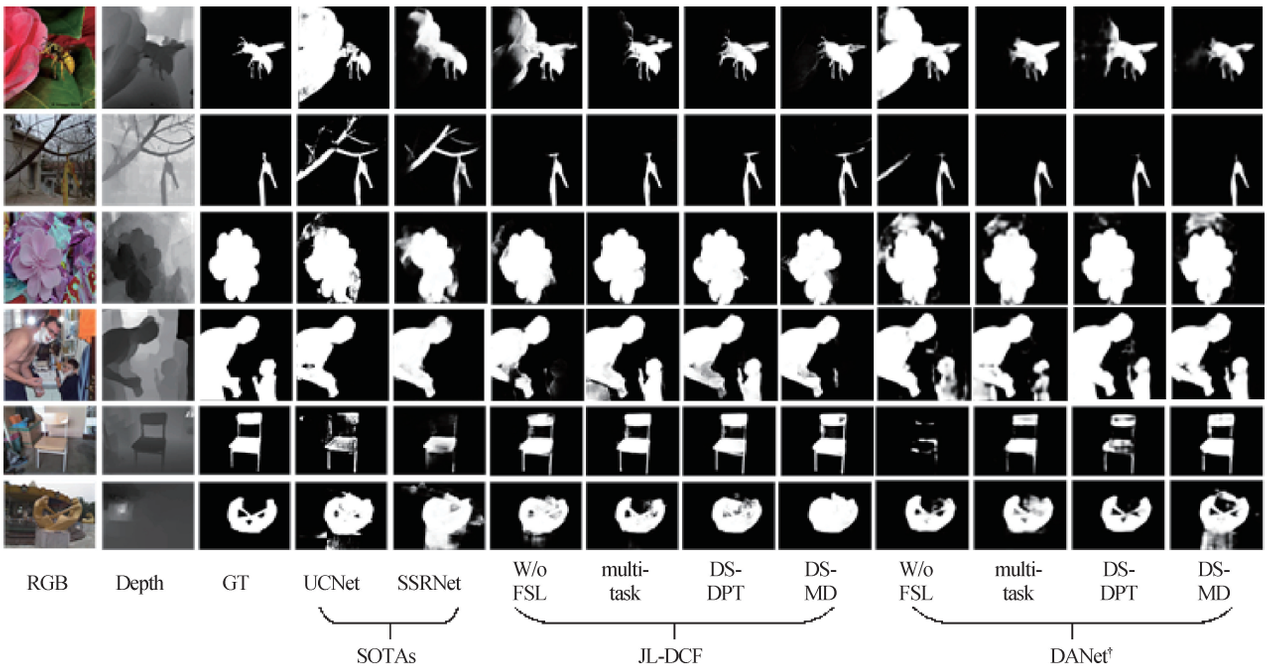
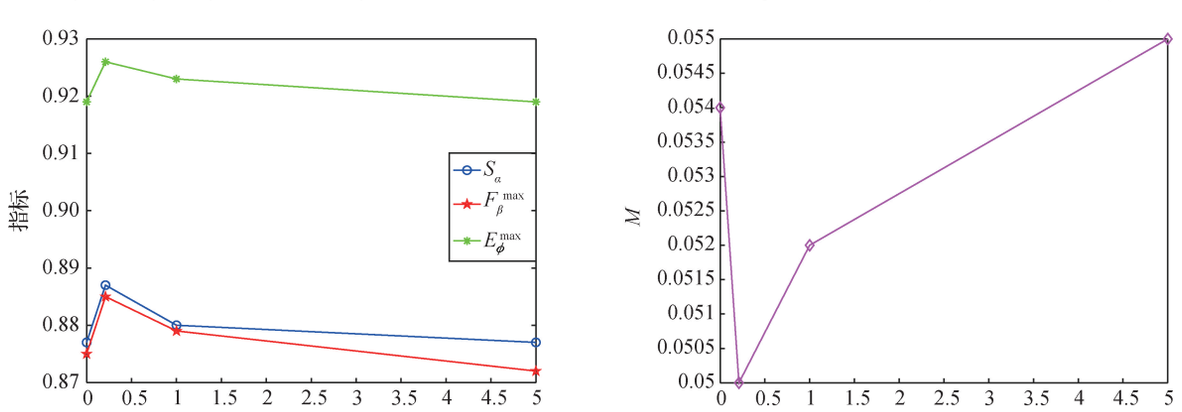


图 2　定性比较

Fig. 2　Qualitative comparison

3.3.3　参数 α 的敏感性分析

为分析基于训练样本扩充方法式(3)中权重系数的取值影响，设置 α 值分别为 0、0.21、1、5 在 JL-DCF 模型上进行实验，在 NJU2K、NLPR、STERE、RGBD135、LFSD、SIP、DUT-RGBD、ReDWeb-S 数据集上的各项平均指标如图 3 所示。值得一提的是，α = 0 即表示不采用合成训练数据的原始性能(对应表 1 中 JL-DCF 栏的 W/o FSL)；α = 1 意味着对所有训练样本同等对待，而 α = 5 表示增加了合成数据的训练权重。从图 3 可见，当 α = 0.21 时，网络取得



(a) S_α、$F_\beta^{\max}$、$E_\phi^{\max}$ 指标的平均值　　(b) M 指标的平均值

图 3　在 8 个数据集上对参数 α 的不同取值(0、0.21、1、5)进行敏感性分析

Fig. 3　Sensitivity analysis for parameters (α = 0, 0.21, 1, 5) on eight benchmark datasets

((a) average value of S_α, $F_\beta^{\max}$, $E_\phi^{\max}$; (b) average value of M)

的性能最好。随着 α 取值的增大,即扩大合成数据集对网络前向传播的影响,意味着网络逐渐偏向学习合成的深度信息,因此导致在真实的 RGB-D 数据上性能有所下降。上述实验表明,在训练样本扩充时控制好合成数据的权重更有益于模型性能的提升,也证明了本文对 α 取值的有效性。

3.3.4 其他讨论

基于以上对实验结果的定量、定性分析,可证明将 RGB-D SOD 视为小样本学习问题的可行性和有效性,但这两类小样本学习方法存在各自的优缺点。在适用性方面,多任务学习方法局限于模型的结构,仅可应用于中期融合模型与后期融合模型,如本文第 2 节所述,早期融合模型无法为网络加入额外的监督信号;而深度生成方法简单直接,理论上可应用于所有模型。另外,对于训练复杂度,深度生成方法受大量训练数据的影响,训练时间较长;而多任务学习方法训练时间较短,训练代价较低。此外,基于深度生成方法的小样本 RGB-D SOD 性能一定程度上依赖于深度生成算法的精度,低质量的深度图易给网络引入噪声,从而影响最终的训练结果。

4 结 论

针对 RGB-D SOD 训练数据集较小的问题,本文从小样本学习角度探讨 RGB-D SOD。鉴于 RGB SOD 任务与 RGB-D SOD 任务的相似性以及数据的可用性,利用小样本学习方法将 RGB SOD 任务的知识迁移到 RGB-D SOD 任务,从模型解空间优化和训练样本扩充对小样本条件下的 RGB-D SOD 进行研究。模型解空间优化将 RGB SOD 与 RGB-D SOD 进行多任务学习共享参数,通过引入 RGB SOD 任务的知识,使网络学习更具泛化性的特征。训练样本扩充利用单目深度生成算法生成相应的深度图,以实现 RGB-D SOD 训练数据集的增广。本文进行了大量实验,从不同角度证明小样本条件下的 RGB-D SOD 的有效性和可行性。总之,面向小样本条件下的 RGB-D SOD 的研究是一项重要任务,目前仅从模型解空间优化角度和训练样本扩充角度对小样本条件下的 RGB-D SOD 方法进行研究,未来将探索并应用更多小样本学习方法以提升 RGB-D 显著性物体检测,乃至其他显著性检测任务的性能。

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]

Caelles S, Maninis K K, Pont-Tuset J, Leal-Taixé L, Cremers D and Van Gool L. 2017. One-shot video object segmentation//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 5320-5329 [DOI: 10.1109/CVPR.2017.565]

Chen H, Li Y F and Su D. 2019. Multi-modal fusion network with multi-scale multi-path and cross-modal interactions for RGB-D salient object detection. Pattern Recognition, 86: 376-385 [DOI: 10.1016/j.patcog.2018.08.007]

Cheng Y P, Fu H Z, Wei X X, Xiao J J and Cao X C. 2014. Depth enhanced saliency detection method//Proceedings of 2014 International Conference on Internet Multimedia Computing and Service. Xiamen, China: ACM: 23-27 [DOI: 10.1145/2632856.2632866]

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: 4558-4567 [10.1109/ICCV.2017.487]

Fan D P, Ji G P, Qin X B and Cheng M M. 2021. Cognitive vision inspired object segmentation metric and loss function. SCIENTIA SINICA Informationis, 51(9): 1475-1489 (范登平, 季葛鹏, 秦雪彬, 程明明. 2021. 认知规律启发的物体分割评价标准及损失函数. 中国科学: 信息科学, 51(9): 1475-1489) [DOI: 10.1360/ssi-2020-0370]

Fan D P, Lin Z, Zhang Z, Zhu M L and Cheng M M. 2021. Rethinking RGB-D salient object detection: models, data sets, and large-scale benchmarks. IEEE Transactions on Neural Networks and Learning Systems, 32(5): 2075-2089 [DOI: 10.1109/TNNLS.2020.2996406]

Finn C, Abbeel P and Levine S. 2017. Model-agnostic meta-learning for fast adaptation of deep networks//Proceedings of the 34th International Conference on Machine Learning. Sydney, Australia: JMLR.org: 1126-1135

Fu J, Liu J, Tian H J, Li Y, Bao Y J, Fang Z W and Lu H Q. 2019. Dual attention network for scene segmentation//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 3141-3149 [DOI: 10.1109/CVPR.2019.00326]

Fu K R, Fan D P, Ji G P and Zhao Q J. 2020. JL-DCF: joint learning and densely-cooperative fusion framework for RGB-D salient object detection//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 3049-3059 [DOI: 10.1109/CVPR42600.2020.00312]

Fu K R, Fan D P, Ji G P, Zhao Q J, Shen J B and Zhu C. 2021. Siamese network for RGB-D salient object detection and beyond. IEEE Transactions on Pattern Analysis and Machine Intelligence: #3073689 [DOI: 10.1109/TPAMI.2021.3073689]

Gui L Y, Wang Y X, Ramanan D and Moura J M F. 2018. Few-shot human motion prediction via meta-learning//Proceedings of the 15th European Conference on Computer Vision. Munich, Germany: Springer: 441-459 [DOI: 10.1007/978-3-030-01237-3_27]

Han J W, Chen H, Liu N, Yan C G and Li X L. 2018. CNNs-based RGB-D saliency detection via cross-view transfer and multiview fusion. IEEE Transactions on Cybernetics, 48(11): 3171-3183 [DOI: 10.1109/TCYB.2017.2761775]

Ju R, Ge L, Geng W J, Ren T W and Wu G S. 2014. Depth saliency based on anisotropic center-surround difference//Proceeding of 2014 International Conference on Image Processing. Paris, France: IEEE: 1115-1119 [DOI: 10.1109/ICIP.2014.7025222]

Li B, Yang Y and Liu Q. 2021. RGB-D video saliency detection via superpixel-level conditional random field. Journal of Image and Graphics, 26(4): 872-882 (李贝, 杨铀, 刘琼. 2021. 超像素条件随机场下的 RGB-D 视频显著性检测. 中国图象图形学报, 26(4): 872-882) [DOI: 10.11834/jig.200122]

Li N Y, Ye J W, Ji Y, Ling H B and Yu J Y. 2017. Saliency detection on light field. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(8): 1605-1616 [DOI: 10.1109/TPAMI.2016.2610425]

Li Z Q and Snavely N. 2018. MegaDepth: learning single-view depth prediction from internet photos//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 2041-2050 [DOI: 10.1109/CVPR.2018.00218]

Liu N, Zhang N, Shao L and Han J W. 2021. Learning selective mutual attention and contrast for RGB-D saliency detection. IEEE Transactions on Pattern Analysis and Machine Intelligence: #3122139 [DOI: 10.1109/TPAMI.2021.3122139]

Munkhdalai T and Yu H. 2017. Meta networks//Proceedings of the 34th International Conference on Machine Learning. Sydney, Australia: JMLR.org: 2554-2563

Niu Y Z, Geng Y J, Li X Q and Liu F. 2012. Leveraging stereopsis for saliency analysis//Proceedings of 2012 IEEE Conference on Computer Vision and Pattern Recognition. Providence, USA: IEEE: 454-461 [DOI: 10.1109/CVPR.2012.6247708]

Peng H W, Li B, Xiong W H, Hu W M and Ji R R. 2014. RGBD salient object detection: a benchmark and algorithms//Proceedings of the 13th European Conference on Computer Vision. Zurich, Switzerland: Springer: 92-109 [DOI: 10.1007/978-3-319-10578-9_7]

Perazzi F, Krähenbühl P, Pritch Y and Hornung A. 2012. Saliency filters: contrast based filtering for salient region detection//Proceedings of 2012 IEEE Conference on Computer Vision and Pattern Recognition. Providence, USA: IEEE: 733-740 [DOI: 10.1109/CVPR.2012.6247743]

Piao Y R, Ji W, Li J J, Zhang M and Lu H C. 2019. Depth-induced multi-scale recurrent attention network for saliency detection//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE: 7253-7262 [DOI: 10.1109/ICCV.2019.00735]

Qu L Q, He S F, Zhang J W, Tian J D, Tang Y D and Yang Q X. 2017. RGBD salient object detection via deep fusion. IEEE Transactions on Image Processing, 26(5): 2274-2285 [DOI: 10.1109/TIP.2017.2682981]

Ranftl R, Bochkovskiy A and Koltun V. 2021. Vision transformers for dense prediction//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: #01196 [DOI: 10.1109/ICCV48922.2021.01196]

Ravi S and Larochelle H. 2017. Optimization as a model for few-shot learning//Proceedings of the 5th International Conference on Learning Representations. Toulon, France: OpenReview.net: 1-11

Snell J, Swersky K and Zemel R. 2017. Prototypical networks for few-shot learning//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, USA: Curran Associates Inc.: 4080-4090

Tsai Y H H, Huang L K and Salakhutdinov R. 2017. Learning robust visual-semantic embeddings//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE: 3591-3600 [DOI: 10.1109/ICCV.2017.386]

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. Honolulu, USA: IEEE: 3796-3805 [DOI: 10.1109/CVPR.2017.404]

Wang W G, Lai Q X, Fu H Z, Shen J B, Ling H B and Yang R G. 2022. Salient object detection in the deep learning era: an in-depth survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(6): 3239-3259 [DOI: 10.1109/TPAMI.2021.3051099]

Wang Y Q, Yao Q M, Kwok J T and Ni L M. 2021. Generalizing from a few examples: a survey on few-shot learning. ACM Computing Surveys, 53(3): 63 [DOI: 10.1145/3386252]

Wang Y X and Hebert M. 2016. Learning from small sample sets by combining unsupervised meta-training with CNNs//Proceedings of the 30th International Conference on Neural Information Processing Systems. Barcelona, Spain: Curran Associates Inc.: 244-252

Wu Y, Lin Y T, Dong X Y, Yan Y, Ouyang W L and Yang Y. 2018. Exploit the unknown gradually: one-shot video-based person re-identification by stepwise learning//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 5177-5186 [DOI: 10.1109/CVPR.2018.00543]

Xu P B, Sang J T and Lu D Y. 2021. Few shot image recognition based

on class semantic similarity supervision. Journal of Image and Graphics, 26(7): 1594-1603 (徐鹏帮,桑基韬,路冬媛. 2021. 类别语义相似性监督的小样本图像识别. 中国图象图形学报, 26(7): 1594-1603) [DOI: 10.11834/jig.200504]

Zhang J, Fan D P, Dai Y C, Anwar S, Saleh F S, Zhang T and Barnes N. 2020. UC-Net: uncertainty inspired RGB-D saliency detection via conditional variational autoencoders//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 8579-8588 [DOI: 10.1109/CVPR42600.2020.00861]

Zhao J X, Cao Y, Fan D P, Cheng M M, Li X Y and Zhang L. 2019. Contrast prior and fluid pyramid integration for RGBD salient object detection//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 3922-3931 [DOI: 10.1109/CVPR.2019.00405]

Zhao X Q, Pang Y W, Zhang L H, Lu H C and Ruan X. 2021. Self-supervised pretraining for RGB-D salient object detection[EB/OL]. http://arxiv.org/pdf/2101.12482.pdf

Zhao X Q, Zhang L H, Pang Y W, Lu H C and Zhang L. 2020. A single stream network for robust and real-time RGB-D salient object detection//Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: Springer: 646-662 [DOI: 10.1007/978-3-030-58542-6_39]

Zhu L C and Yang Y. 2018. Compound memory networks for few-shot video classification//Proceedings of the 15th European Conference on Computer Vision. Munich, Germany: Springer: 782-797 [DOI: 10.1007/978-3-030-01234-2_46]

作者简介

何静,女,硕士研究生,主要研究方向为 RGB-D 显著性物体检测。E-mail: hej@stu.scu.edu.cn

傅可人,通信作者,男,副研究员,主要研究方向为图像视频中的目标检测识别、深度学习、计算机视觉。

E-mail:fkrsuper@scu.edu.cn