

DOI:10.16136/j.joel.2023.10.0443

基于提取双选紧密特征的 RGB-D 显著性检测网络

化春键^{1,2*}, 邹新童^{1,2}, 蒋毅^{1,2}, 俞建峰^{1,2}, 陈莹³

(1. 江南大学 机械工程学院, 江苏 无锡 214122; 2. 江苏省食品先进制造装备技术重点实验室, 江苏 无锡 214122; 3. 江南大学 物联网工程学院, 江苏 无锡 214122)

摘要:针对现有算法对不同来源特征之间的交互选择关注度欠缺以及对跨模态特征提取不充分的问题,提出了一种基于提取双选紧密特征的 RGB-D 显著性检测网络。首先,为了筛选出能够同时增强 RGB 图像显著区域和深度图像显著区域的特征,引入双向选择模块(bi-directional selection module, BSM);为了解决跨模态特征提取不充分,导致算法计算冗余且精度低的问题,引入紧密提取模块(dense extraction module, DEM);最后,通过特征聚合模块(feature aggregation module, FAM)对密集特征进行级联融合,并将循环残差优化模块(recurrent residual refinement aggregation module, RAM)配合深度监督实现粗显著图的持续优化,最终得到精确的显著图。在4个广泛使用的数据集上进行的综合实验表明,本文提出的算法在4个关键指标方面优于7种现有方法。

关键词:图像处理; RGB-D 显著性检测; 跨模态特征; 循环残差

中图分类号: TP391 **文献标识码:** A **文章编号:** 1005-0086(2023)10-1026-10

RGB-D visual saliency detection network based on extracting bi-directional selection dense features

HUA Chunjian^{1,2*}, ZOU Xintong^{1,2}, JIANG Yi^{1,2}, YU Jianfeng^{1,2}, CHEN Ying³

(1. School of Mechanical Engineering, Jiangnan University, Wuxi, Jiangsu 214122, China; 2. Jiangsu Key Laboratory of Advanced Food Manufacturing Equipment & Technology, Wuxi, Jiangsu 214122, China; 3. School of Internet of Things Engineering, Jiangnan University, Wuxi, Jiangsu 214122, China)

Abstract: In order to solve the problem that the existing algorithms pay less attention to the interactive selection between features from different sources and the extraction of cross modal features is insufficient, a RGB-D visual saliency detection network based on extracting bi-directional selection dense features is proposed. First, in order to filter out the features that can enhance the saliency areas of RGB images and depth images at the same time, a bi-directional selection module (BSM) is introduced. In order to solve the problem of insufficient cross modal feature extraction, which leads to redundant calculation and low accuracy, a dense extraction module (DEM) is introduced. Finally, the dense features are cascaded and fused through the feature aggregation module (FAM), and the recurrent residual refinement aggregating module (RAM) is combined with the deep supervision to achieve the continuous optimization of the coarse saliency maps, and finally the accurate saliency maps are obtained. Comprehensive experiments on four widely used datasets show that the proposed algorithm is superior to seven existing methods in four key indicators.

Key words: image processing; RGB-D visual saliency detection; cross modal feature; recurrent residual

0 引言

显著性检测的目的是通过模仿人类视觉注意

的特征来识别图像中最明显的区域或对象^[1]。显著性检测算法将有限的计算资源分配给场景中的重要区域,不仅为后续的视觉任务提供有效的信

* E-mail: cjhua@jiangnan.edu.cn

收稿日期: 2022-06-16 修订日期: 2022-09-26

基金项目: 国家自然科学基金(62173160)资助项目

息,而且可以消除冗余区域的干扰,降低计算成本,有利于各种计算机视觉任务,包括交通场景检测^[2]、合成孔径雷达图像处理^[3]、海面船舶检测^[4,5]等。

得益于深度学习技术的快速发展,显著性目标检测近年来取得了显著的进展。BRUNO 等^[6]将图像中兴趣点的分布与多尺度分析相结合,利用感知一致的颜色空间来研究颜色对显著性提取的影响。凌艳^[7]提出一种基于紧密聚合特征的循环残差网络,改善了目前显著性检测算法在复杂场景中精度下降的问题。近年来,随着深度信息的引入,有不少学者开始研究 RGB-D 显著性检测算法并卓有成效,现有的 RGB-D 显著性检测算法按照其提取 RGB 特征和深度特征的策略不同,大体上分为两类:并行独立提取和提取深度特征输入 RGB 流。

并行独立提取策略首先并行地从 RGB 和深度图像中提取特征,然后通过其他模块将其融合,随后得到显著图。IMAMOGLU 等^[8]引入了一个多模式显著对象检测框架,先对 RGB 图像和深度图像进行自下而上和自上而下的双重特征提取,再进行像素级加法,从而获得融合显著线索。ARIVAZHAGAN 等^[9]从 RGB 和深度动作序列生成显著信息图,并从中提取了代表完全局部二值模式的符号、幅值和中心描述符,然后通过典型相关分析结合降维进行深度特征和 RGB 特征的融合。LI 等^[10]提出了一种跨模态特征调制模块,通过将深度特征作为先验来增强特征表示,其次,提出了一种自适应特征选择模块来选择与显著性相关的特征并抑制劣质特征。SINGH 等^[11]提出了一种具有基于相互注意区分窗口的复合骨干网络,在每个编码器阶段插入基于通道、空间和特征

级注意的区分窗口,以增强显著性特征。

提取深度特征时输入 RGB 流的特征提取策略,即利用深度线索作为显著性检测的指导信息或对 RGB 特征进行增强。ZHU 等^[12]提出了一种用于 RGB-D 显著目标检测的鲁棒性先验模型引导深度增强网络,该网络由处理 RGB 值的主网络和充分利用深度线索并将基于深度的特征合并到主网络中的子网络组成。DAS 等^[13]使用深度估计网络来查找二维图像的深度图,将该深度图用于训练深度引导显著性网络,从而生成中间深度显著性图,最后,将深度显著图与粗糙显著图进行融合,得到最终显著图。

上述两种常用的特征提取策略都是直接对从 RGB 图像和深度图像中提取的特征进行传输利用,而忽略了不同来源特征之间的交互选择问题,即应筛选出有利于增强 RGB 图像中显著部分的深度特征,以及有利于增强深度图像中显著部分的 RGB 特征,并在较早阶段完成跨模态特征的融合和细化提取,以求在获得高质量特征的同时控制计算成本。

为解决上述问题,本文提出一种基于提取双选紧密特征的 RGB-D 显著性检测网络,双向选择的特征提取策略即在特征提取阶段,使 RGB 特征和深度特征进行相互校正,随后进行紧密的特征提取,为后续模块提供较高质量的特征,再通过特征聚合模块(feature aggregation module, FAM)得到粗显著图,最后由循环残差对粗显著图进行持续优化,得到更精确的显著图,本文算法模块均可拆分迁移,具有广阔的适用范围。

1 基本原理

本文算法的总体框架如图1所示,选用去除全

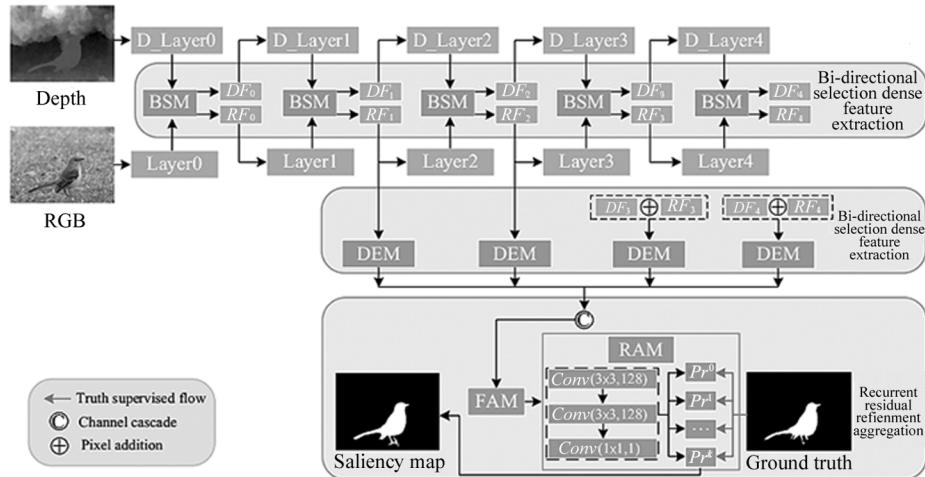


图 1 基于提取双选紧密特征的 RGB-D 显著性检测网络总体框架

Fig. 1 The overall structure of RGB-D visual saliency detection network based on extracting bi-directional selection dense features

连接层的 ResNeXt-101^[14] 作为主干网络,并按照输入图像不同,将其分为 RGB 流和深度流。算法主体由双选紧密特征提取和聚合循环残差优化两部分构成,其中,双选紧密特征提取部分包括双向选择模块(bi-directional selection module, BSM)和密集提取模块(dense extraction module, DEM),聚合循环残差优化部分包括 FAM 和循环残差优化模块(recurrent residual refinement aggregating module, RAM)。主干网络提取到的初级特征经过 RGB 信息和深度信息的双向选择后进行紧密提取,得到质量较高的特征后对其进行聚合,最后通过循环残差对粗显著图进行持续优化并对中间过程进行深度监督后得到最终显著图。

1.1 双选紧密特征提取

双选紧密特征提取部分的主要任务是完成对 RGB 特征和深度特征的双向选择,获得跨模态特征,然后再对跨模态特征进行紧密提取,在控制计算成本的同时为后续处理提供质量良好的特征。

1.1.1 BSM

BSM 结构与执行过程如图 2 所示,该模块以经典的空间-通道注意力机制^[15]为基础,先通过空间注意力辨别显著对象的位置,再通过通道注意力选择特征通道来辨别哪些特征是有意义的^[16],然后对 RGB 信息和深度信息进行交叉传输和跨模态信息融合,从而过滤掉不利于显著性目标检测的特征。与文献[16]直接将交叉选择后的特征反馈回主干网络不同,本文算法在反馈主干网络的同时,还对交叉选择特征进行跨模态融合,并通过紧密连接的方式对融合后的特征进行进一步提取,有效保留了显著信息并提高了模型的运算效率。

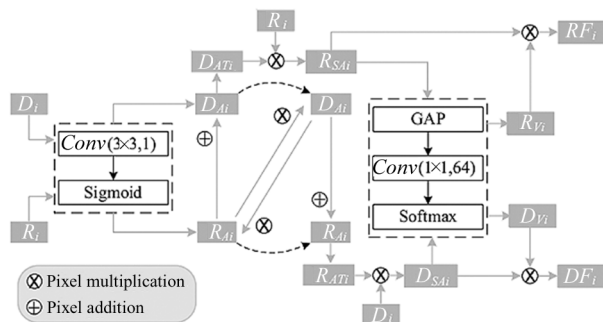


图 2 双向选择模块结构与执行过程

Fig. 2 The structure and execution process of BSM module

BSM 的输入 $D_i (i = 0, 1, 2, 3, 4)$ 和 $R_i (i = 0, 1, 2, 3, 4)$ 分别为图 1 中深度流主干网络 D_Lay-

er0—D_Layer4 和 RGB 流主干网络 Layer0—Layer4 的各层输出。首先,对 D_i 和 R_i 分别进行核为 3×3 且输出通道数为 1 的卷积和 Sigmoid 非线性激活后,得到相应的空间注意线索 $D_{Ai} (i = 0, 1, 2, 3, 4)$ 和 $R_{Ai} (i = 0, 1, 2, 3, 4)$ 。接下来,将得到的深度空间注意线索 D_{Ai} 转换到 RGB 模态,得到由深度信息补偿的 RGB 特征 $R_{SAi} (i = 0, 1, 2, 3, 4)$, 如式 (1) 所示^[16]:

$$\begin{cases} D_{ATi} = D_{Ai} \times R_{Ai} + D_{Ai} \\ R_{SAi} = D_{ATi} \times R_i, i = 0, 1, 2, 3, 4, \end{cases} \quad (1)$$

式中, $D_{ATi} (i = 0, 1, 2, 3, 4)$ 的作用是通过 R_{Ai} 保持 RGB 模态的固有属性。由 RGB 信息补偿的深度特征 $D_{SAi} (i = 0, 1, 2, 3, 4)$ 采用相同方式获得, 如式 (2) 所示^[16]:

$$\begin{cases} R_{ATi} = R_{Ai} \times D_{Ai} + R_{Ai} \\ D_{SAi} = R_{ATi} \times D_i, i = 0, 1, 2, 3, 4, \end{cases} \quad (2)$$

式中, $R_{ATi} (i = 0, 1, 2, 3, 4)$ 的作用是通过 $D_{Ai} (i = 0, 1, 2, 3, 4)$ 保持深度模态的固有属性。

然后沿着通道维度对获得的跨模态补偿特征 R_{SAi} 和 D_{SAi} 进行选择,具体操作是使 R_{SAi} 和 D_{SAi} 依次经过全局平均池化层、核为 1×1 、输出通道为 64 的卷积层以及 Softmax 函数,得到通道权重向量 $R_{Vi} (i = 0, 1, 2, 3, 4)$ 和 $D_{Vi} (i = 0, 1, 2, 3, 4)$, 最后通过式 (3) 和式 (4) 得到双向选择后的 RGB 特征 $RF_i (i = 0, 1, 2, 3, 4)$ 和深度特征 $DF_i (i = 0, 1, 2, 3, 4)$ ^[16]:

$$RF_i = R_{Vi} \times R_{SAi}, i = 0, 1, 2, 3, 4, \quad (3)$$

$$DF_i = D_{Vi} \times D_{SAi}, i = 0, 1, 2, 3, 4. \quad (4)$$

BSM 能使不同来源的特征相互细化,以获得低噪的纯净初始特征,为获得高精度的显著图提供良好基础。

1.1.2 DEM

经过双向细化选择后的特征,还需进行进一步提取,从而避免信息分散,导致检测结果产生误差,为增强跨模态特征连续性,引入多个 DEM,每个 DEM 均为 4 层,关于该模块的深度问题,将在“2.3.3 DEM 深度实验”中进行详细讨论,DEM 的每层输入特征均为之前所有层输出特征的级联,以实现跨模态特征的紧密提取。

DEM 的结构和执行过程如图 3 所示。该模块的输入为经双向选择后的 RGB 特征或 RGB 和深度的融合特征,关于深度信息融合层数问题,将在“2.3.1 深度信息融合层数实验”中进行详细讨论。DEM 由一个核为 3×3 、输出通道数为 128 的降维卷积和 4 个核为 3×3 、输出通道数为 64 的特征提取卷

积构成^[7],该模块的输出是所有卷积层输出特征的级联。该模块通过紧密提取特征的方式,有效扩大感受野的同时实现了对多层信息的重复利用,另外,引入该模块后,模型计算量减少了 2 672 M,可以有效提高计算效率。

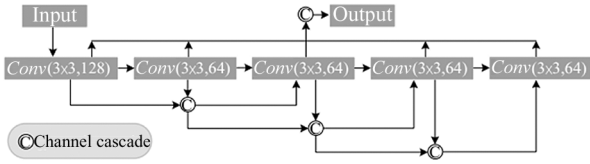


图 3 紧密提取模块的结构和执行过程

Fig. 3 The structure and execution process of DEM module

1.2 聚合循环残差优化

聚合循环残差优化部分的主要任务是对输入的级联紧密跨模态特征在多个尺度上进行聚合,获得粗显著图,然后再通过循环残差对粗显著图进行持续优化,最终得到较为精确的显著图。

1.2.1 FAM

FAM 即为图 1 中 FAM,该模块受空洞空间金字塔池化(atrous spatial pyramid pooling, ASPP)^[17]启发,先将级联特征降维至 256 后输入 ASPP,在 ASPP 中,所有特征并行通过一个核为 1×1 、输出通道数为 128、膨胀率分别为 2、4、6 的膨胀卷积以及一个全局平均池化层,得到了充分覆盖上下文信息和图像级全局信息的聚合特征,从而实现不同尺度间信息的交叉融合以及对显著性信息的整合。FAM 的输出为本文模型的粗显著图。

1.2.2 RAM

RAM 即为图 1 中 RAM,该模块的输入为 DEM 输出的粗显著图,经点卷积运算得到初始预测图 Pr^0 ,将 Pr^0 与 DEM 输出的粗显著图级联后依次输入两个核为 3×3 、输出通道数为 128 的卷积层和一个点卷积层^[7],得到预测图 Pr^1 ,以此类推循环 K 次,直至得到预测图 Pr^K , Pr^K 即为本文算法得到的最终显著图,RAM 执行过程的公式化体现如式(5)所示:

$$Pr^k = \begin{cases} \text{Conv}(AF), & k = 0 \\ \text{Conv}(\text{Conv3}(\text{Conv3}(\text{Conv3}(Pr^c))) + Pr^{k-1}), & k > 0 \end{cases}, \quad (5)$$

式中, $\text{Conv}(\cdot)$ 表示一个点卷积, AF 为 DEM 输出的粗显著图, $\text{Conv3}(\cdot)$ 表示核为 3×3 的卷积层, $Pr^c = \text{Cat}(Pr^{k-1}, AF)$ 表示预测图与粗显著图的级联, k 为残差循环次数。关于 k 的取值问题将在“2.3.2 RAM

循环次数实验”中进行详细讨论。RAM 通过残差循环的方式对多组预测图进行反复级联净化,力求降低中间预测图含有的噪声,从而达到提高检测准确率的效果。

1.3 监督机制

本文算法的训练过程采用标准交叉熵损失,同时,考虑到 RAM 涉及多重输出,为了能对各级输出进行全面细致的监督,加速算法收敛,引入深度监督机制^[7],先从高级特征中学习残差信息,有利于快速修正结果,提高模型学习效率,本文算法总损失的计算方法如式(6)所示:

$$L = L(Pr^0, GT) + \sum_{k=1}^K L(Pr^k, GT). \quad (6)$$

当总损失 L 值收敛至最小时,即可得到本文算法训练的最终模型。

2 实验结果分析

2.1 实验参数说明

本文实验共用到 5 个广泛使用的 RGB-D 数据集,其中,训练集包括来自 NJU2K^[18] 的 1 500 个样本和来自 NLPR^[19] 的 700 个样本,为了增强网络对图像变换的鲁棒性,同时缓解过度拟合问题,使用了数据增强,对训练数据集中的图像应用随机旋转、随机裁剪和水平翻转。测试集为 NLPR 剩余的 300 个样本以及 STEREO^[20]、SIP^[21] 和 DES^[22]。以下是关于所涉及的 5 个数据集的简介。

NJU2K 由从互联网、3D 电影和照片中收集的图像组成,其中深度图像是根据立体图像估计的;NLPR 包含 1 000 对在不同照明条件下拍摄的图像,其中深度图像由 Microsoft Kinect 捕获;STEREO 也称为 SSB1000,由 1 000 对立体图像和从互联网下载的相应真值图组成,其中深度图像由立体图像估计得到;SIP 包括 929 对高分辨率室外真人图像,其中深度图像由华为 Meta10 捕获;DES 由 135 对室内图像和真值图组成,其中深度图由 Microsoft Kinect 捕获。可知本文所选用数据集涉及人物、景物、室内、室外以及不同光照条件下的图像,具有广泛的覆盖范围。

实验运行环境是 64 位 Ubuntu 16.04 操作系统,32 GB 内存和 1 块 Geforce GTX 1080 GPU。本文算法基于的深度框架是 Pytorch^[23],并将输入 RGB 深度图像对调整为 352×352 分辨率,使用 Adam 优化模型^[24]。初始的学习率设置为 1×10^{-4} 并且每隔 60 轮下降 10 倍。

2.2 实验评价指标

本文主要使用4个评价指标来评价模型的性能,包括准确率-召回率(precision-recall, PR)曲线、 F 值^[25]、平均绝对误差(mean absolute error, MAE)^[26]和 S 值^[27]。

PR 曲线用于计算特征图的准确率和召回率。当对输出图像进行二值化时,阈值从0—255进行选择,并且每次获取阈值时,可以为所有输出图像计算一组相应的精度值和召回值。最后,分别对该阈值下的所有图像的精度值和召回值进行平均,得到256对 P 值和 R 值。召回值是横坐标,精度值是纵坐标,绘制曲线图以获得 PR 曲线, PR 曲线越靠近右上角区域则表明算法性能越优。

F 值是非负权重下召回率和准确率的加权调和平均值,其计算式如式(7)所示^[25]:

$$F = \frac{(1 + \beta^2)PR}{\beta^2 P + R}, \quad (7)$$

式中, β 是衡量召回率和准确率的非负权值,根据显著目标检测任务的经验,通常设置 $\beta^2 = 0.3$,即增加准确率的权重值。本文采用的是最大的 F 值表示算法能够获得的最好的性能。

MAE 表示预测值和真值之间绝对误差的平均值,范围是 $[0, +\infty)$,当预测值与真值完全一致时,它等于0,这是一个完美的模型;预测值与真值之间的误差越大, MAE 值越大。 MAE 计算公式如式(8)所示^[26]:

$$MAE = \frac{1}{m} \sum_{i=1}^m |y_i - f(x_i)|, \quad (8)$$

式中, m 表示样本数, $f(x)$ 表示模型的预测值, y 表示真值。 MAE 值越小,表明两者相似度越高并且背景噪声越少,算法整体性能更加优越。

S 值侧重于评估显著图的结构信息,它比 F 值更接近人类视觉系统。它主要计算预测值与真值之间的物体感知和区域感知的结构相似性。 S 值的计算式如式(9)所示^[27]:

$$S_\lambda = \lambda \times S_o + (1 - \lambda) \times S_a, \quad (9)$$

式中, $\lambda \in [0, 1]$ 是平衡参数,通常值取为0.5, S_o 和 S_a 分别表示目标感知和区域感知的结构相似性。 S 值越大,表示显著图和真值图在结构上的误差越小,算法的性能更优。

2.3 消融实验

2.3.1 深度信息融合层数实验

位于低层网络的低维深度特征可能含有一定噪声,从而对检测结果造成不良影响,为确定融合几层

深度信息能使本文算法达到最佳性能,进行了关于深度信息融合层数的实验,具体操作是在DEM中,分别加入4层深度信息、3层深度信息、2层深度信息和1层深度信息,输入的深度信息来源即图1中深度流主干网络的D_Layer1—D_Layer4的4层输出,此处记为d1—d4。

以涉及检测目标类别最多的STEREO数据集为例进行消融实验,实验所得 PR 曲线图如图4所示,由图4可知,当深度信息融合层数为d3—d4,即加入图1中D_Layer3和D_Layer4的输出时, PR 曲线最靠近右上角,说明算法性能达到最优。表1给出了融合不同层数深度信息时STEREO数据集上得到的客观评价指标值,最优值加粗表示,可以看出,当融入深度信息层数为d3—d4时,得到了最佳的 F 值、 MAE 和 S 值。所以,确定本文算法的深度信息融合层数为d3—d4。

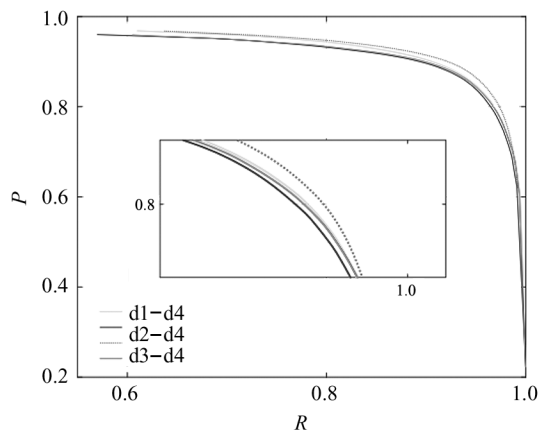


图4 STEREO数据集深度信息融合层数的 PR 曲线图
Fig. 4 PR curves of depth information fusion layers on STEREO dataset

表1 STEREO数据集深度信息融合层数评价指标
Tab. 1 Evaluation criteria of depth information fusion layers on STEREO dataset

Evaluation criteria	Depth information fusion layers			
	d1—d2	d2—d4	d3—d4	d4
$S \uparrow$	0.898	0.897	0.904	0.896
$F \uparrow$	0.909	0.902	0.914	0.903
$MAE \downarrow$	0.047	0.047	0.042	0.049

2.3.2 RAM循环次数实验

本文将RAM的循环次数分别设置为1—5进行消融实验,记为R1—R5,以STEREO数据集为例的 PR 曲线图如图5所示,由图5可知,当循环次数为3时, PR 曲线最靠近右上角,算法性能达到最佳。表2为不同循环次数下STEREO数据集上得到的客观

评价指标值,可以看出,当循环次数为 3 时,得到了最佳的 F 值、 MAE 和 S 值。所以,确定本文算法 RAM 的循环次数为 3。

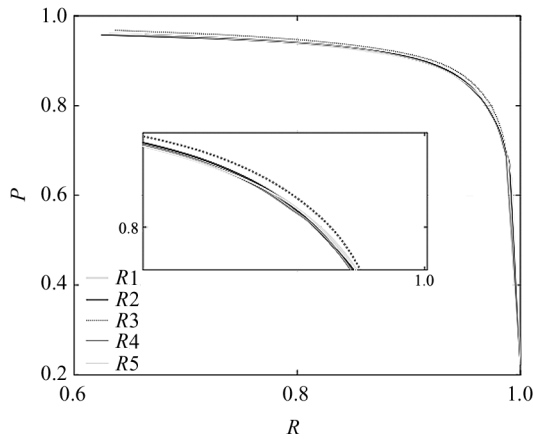


图 5 STEREO 数据集 RAM 循环次数的 PR 曲线图

Fig. 5 PR curves of recurrences number of RAM on STEREO dataset

表 2 STEREO 数据集 RAM 的循环次数评价指标

Tab. 2 Evaluation criteria of recurrences number of RAM on STEREO dataset

Evaluation criteria	Number of recurrences				
	R1	R2	R3	R4	R5
$S \uparrow$	0.901	0.901	0.904	0.899	0.898
$F \uparrow$	0.906	0.911	0.914	0.909	0.908
$MAE \downarrow$	0.045	0.045	0.042	0.046	0.046

2.3.3 DEM 深度实验

本文将 DEM 的深度分别设置为 1—5 进行消融实验,记为 Dense1—Dense5,以 STEREO 数据集为例的 PR 曲线图如图 6 所示,由图 6 可知,当紧密提取深度为 4 时, PR 曲线最靠近右上角,算法性能达到最佳。表 3 为不同紧密特征提取深度时 STEREO 数据集上得到的客观评价指标值,可以看出,当紧密提取深度为 4 时,得到了最佳的 F 值、 MAE 和 S 值。所以,确定本文算法 DEM 的深度为 4。

表 3 STEREO 数据集 DEM 深度评价指标

Tab. 3 Evaluation criteria of depth of DEM on STEREO dataset

Evaluation criteria	Dense extraction depth				
	Dense1	Dense2	Dense3	Dense4	Dense5
$S \uparrow$	0.902	0.901	0.900	0.904	0.899
$F \uparrow$	0.911	0.910	0.909	0.914	0.908
$MAE \downarrow$	0.045	0.045	0.046	0.042	0.047

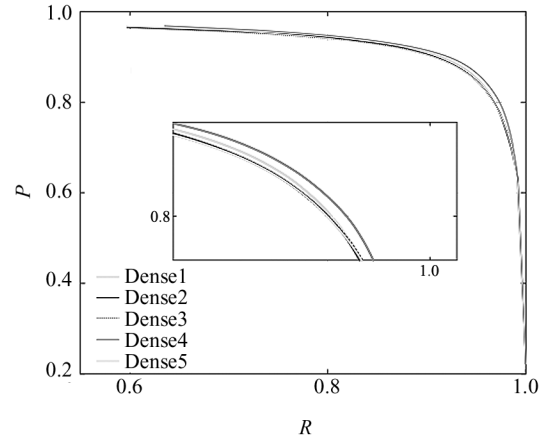


图 6 STEREO 数据集 DEM 的深度 PR 曲线图

Fig. 6 PR curves of depth of DEM on STEREO dataset

2.3.4 模块消融实验

本文分别验证了 4 个模块对算法性能的影响,具体操作为单独去掉某个模块,保持其他设置不变进行训练,如去掉 BSM,记为 w/o BSM,其他模块同理,Baseline 表示去掉所有模块的基础网络。以 STEREO 数据集为例的 PR 曲线图如图 7 所示,由图 7 可知,去掉 4 个模块中任意一个模块均会对算法性能造成影响。表 4 为去掉任意模块后 STEREO 数据集上得到的客观评价指标值与本文完整算法的评价指标对比,可以看出,去掉任一模块后,都会导致 F 值和 S 值下降、 MAE 值上升,因此可以证明本文算法涉及的 4 个模块对提升检测准确性均有一定帮助。

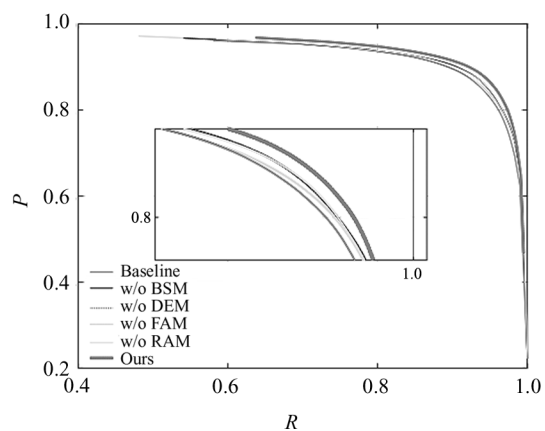


图 7 STEREO 数据集模块消融实验的 PR 曲线图

Fig. 7 PR curves of module ablation experiments on STEREO dataset

表4 STEREO数据集模块消融实验评价指标

Tab. 4 Evaluation criteria of module ablation experiments on STEREO dataset

Evaluation criteria	Module ablation					
	Baseline	w/o BSM	w/o DEM	w/o FAM	w/o RAM	Ours
$S \uparrow$	0.896	0.899	0.902	0.895	0.899	0.904
$F \uparrow$	0.903	0.907	0.907	0.904	0.907	0.914
$MAE \downarrow$	0.050	0.046	0.044	0.048	0.046	0.042

2.4 实验结果定量评价

本文分别验证了4个模块对算法性能的影响,图8—图11分别给出了近3年RGB-D显著性检测相关算法以及本文算法在4个测试数据集上的PR曲线图,可以看出,本文算法在高准确率的部分持续范围较广,并且较接近于坐标系的右上角,特别是当召回率处于0.8—1.0区间时,准确率较其他算法更高,因此,整体上本文算法是优于其他算法的,具有较高的检测可靠性。

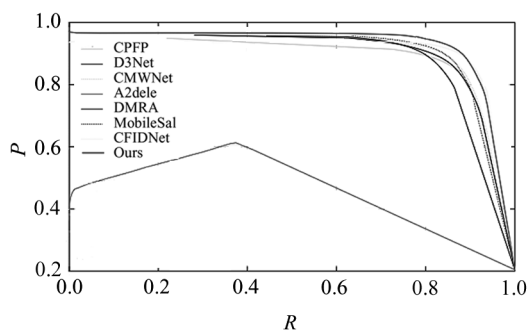


图8 SIP数据集对比算法的PR曲线图

Fig. 8 PR curves of comparison algorithms on SIP dataset

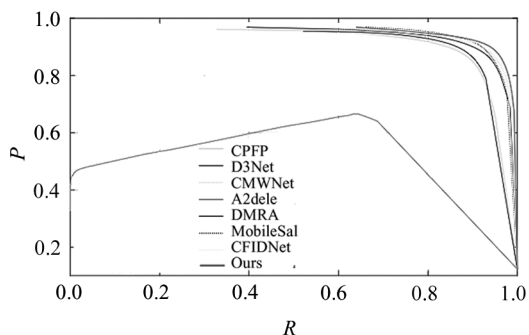


图9 NLPR数据集对比算法的PR曲线图

Fig. 9 PR curves of comparison algorithms on NLPR dataset

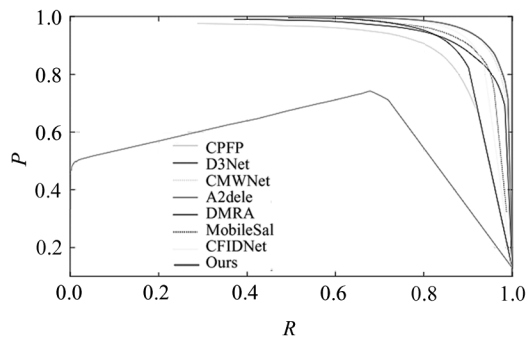


图10 DES数据集对比算法的PR曲线图

Fig. 10 PR curves of comparison algorithms on DES dataset

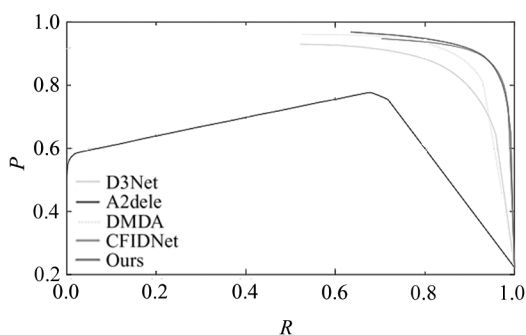


图11 STEREO数据集对比算法的PR曲线图

Fig. 11 PR curves of comparison algorithms on STEREO dataset

表5—表8给出了其他3个客观评价指标值,最佳值加粗表示,可以看出,本文算法的检测效果位于前列。由于MobileSal^[28]算法在训练时进行了多尺度预处理,即每张图像在训练中均被调整为[256, 288, 320],因此,在前景目标尺度范围较广的数据集中效果较好,故而在NLPR数据集中MAE值略优于本文算法,但从S值和F值来看依然是本文算法最优。从综合客观评价指标来看,本文算法预测的显著图与真值图空间结构更为接近。

2.5 实验结果定性评价

本文选用了7种经典算法进行了对比实验,视觉直观的比较结果如图12所示。图12(a)—(j)分别代表不同类型的图像,图(a)的背景复杂且背景包含颜色鲜艳区域,图(b)的前景目标边缘复杂,图(c)的图像整体对比度较低,图(d)为小尺度目标且背景复杂,图(e)的前景与背景相似度极高,图(f)在光照昏暗条件下,图(g)的复杂背景,图(h)的前景背景相似度较高且几乎在同一平面,图(i)为严重光照不均,图(j)为过度曝光。由图12中不同算法对各类图像的

检测结果可以看出,在复杂背景情况下,本文算法依然能直接定位最为显著的目标,并且有较高的准确度;在前景目标与背景相似度较高的情况下,本文算法的鲁棒性更高,得到的显著区域更加完整和平滑;在光照不均或图像整体昏暗的情况下,本文算法依然能够提取清晰完整的显著区域。综合以上各种情况,本文算法具有较强的抗背景干扰能力,受光照条

件影响较小,对前景目标的定位及检测精度较高,因而本文算法具有较好的显著性检测性能。

3 结 论

本文提出了一种基于提取双选紧密特征的 RGB-D 显著性检测网络,解决了现有算法对交叉选择的跨模态特征提取不充分的问题。利用BSM,筛

表 5 SIP 数据集对比算法的评价指标

Tab. 5 Evaluation criteria of comparison algorithms on SIP dataset

Evaluation criteria	Algorithms							
	CPFP ^[13]	D3Net ^[21]	CMWNet ^[29]	A2dele ^[30]	DMRA ^[31]	MobileSal ^[28]	CFIDNet ^[32]	Ours
$S \uparrow$	0.850	0.860	0.867	0.575	0.846	0.866	0.881	0.888
$F \uparrow$	0.870	0.861	0.874	0.537	0.879	0.893	0.900	0.910
$MAE \downarrow$	0.064	0.063	0.062	0.173	0.065	0.057	0.051	0.049

表 6 NLPR 数据集对比算法的评价指标

Tab. 6 Evaluation criteria of comparison algorithms on NLPR dataset

Evaluation criteria	Algorithms							
	CPFP ^[13]	D3Net ^[21]	CMWNet ^[29]	A2dele ^[30]	DMRA ^[31]	MobileSal ^[28]	CFIDNet ^[32]	Ours
$S \uparrow$	0.888	0.912	0.917	0.723	0.892	0.919	0.922	0.926
$F \uparrow$	0.888	0.897	0.903	0.660	0.896	0.917	0.914	0.919
$MAE \downarrow$	0.036	0.030	0.029	0.079	0.034	0.024	0.026	0.025

表 7 DES 数据集对比算法的评价指标

Tab. 7 Evaluation criteria of comparison algorithms on DES dataset

Evaluation criteria	Algorithms							
	CPFP ^[13]	D3Net ^[21]	CMWNet ^[29]	A2dele ^[30]	DMRA ^[31]	MobileSal ^[28]	CFIDNet ^[32]	Ours
$S \uparrow$	0.872	0.898	0.934	0.761	0.892	0.910	0.917	0.937
$F \uparrow$	0.882	0.885	0.930	0.726	0.913	0.921	0.934	0.943
$MAE \downarrow$	0.038	0.031	0.022	0.069	0.032	0.025	0.023	0.021

表 8 STEREO 数据集对比算法的评价指标

Tab. 8 Evaluation criteria of comparison algorithms on STEREO dataset

Evaluation criteria	Algorithms				
	D3Net ^[21]	A2dele ^[30]	DMRA ^[31]	CFIDNet ^[32]	Ours
$S \uparrow$	0.857	0.741	0.877	0.904	0.904
$F \uparrow$	0.863	0.752	0.898	0.910	0.914
$MAE \downarrow$	0.062	0.105	0.051	0.042	0.042

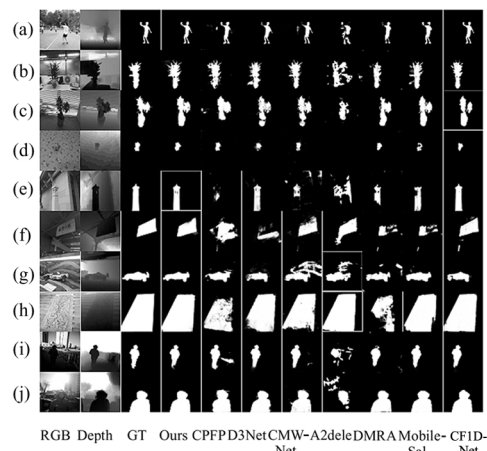


图12 不同算法视觉效果对比

Fig. 12 Comparison of visual effects of different algorithms

选出同时有利于 RGB 显著区域和深度显著区域的跨模态特征,然后对筛选出的跨模态特征进行紧密提取,在保留有效显著特征的同时减小了计算量。利用 FAM 对级联后的跨模态特征进行跨尺度聚合,使其能够全面覆盖上下文信息,最后在深度监督机制的配合下利用循环残差对聚合特征进行持续优化,最终获得分割效果良好且边缘清晰的显著图。在 4 个公开的数据集上,将本文方法与 7 种流行的 RGB-D 显著性检测方法进行定性和定量比较,实验结果表明,本文算法的显著图与真值图最为接近且在 4 个客观评价指标方面均有一定优势。后续工作将致力于提高算法在复杂场景下的检测精度以及尽可能减小模型大小从而节省计算资源。

参考文献:

- [1] WU D, LI T, WAN Q. Salient object region detection based on background-bias prior and center-bias prior [J]. Journal of Optoelectronics • Laser, 2022, 33(8): 799-806.
吴迪, 李婷, 万琴. 基于背景先验与中心先验的显著性目标检测[J]. 光电子·激光, 2022, 33(8): 799-806.
- [2] FANG Z M, CUI R Y, JIN J X. Static saliency region detection in traffic scenes [J]. Laser & Optoelectronics Progress, 2017, 54(5): 051501.
方志明, 崔荣一, 金璟璇. 交通场景静态显著性区域检测[J]. 激光与光电子学进展, 2017, 54(5): 051501.
- [3] YANG H X, GUO H, GAO Y, et al. SAR flow ice separation algorithm combined with saliency detection [J]. Laser & Optoelectronics Progress, 2020, 57(20): 201010.
杨红霞, 郭浩, 高岩, 等. 结合显著性检测的 SAR 流冰分离算法 [J]. 激光与光电子学进展, 2020, 57(20): 201010.
- [4] CHEN Y T, LI Y Y, YAO T T. Ship detection from remote sensing image under complex sea conditions [J]. Laser & Optoelectronics Progress, 2019, 56(18): 181007.
陈彦彤, 李雨阳, 姚婷婷. 复杂海况下遥感图像舰船目标检测方法研究 [J]. 激光与光电子学进展, 2019, 56(18): 181007.
- [5] YAO T T, ZHANG B, LI P F, et al. Salient ship detection based on robust background estimation [J]. Laser & Optoelectronics Progress, 2022, 59(8): 0810008.
姚婷婷, 张波, 李鹏飞, 等. 基于鲁棒背景估计的船舶显著性检测 [J]. 激光与光电子学进展, 2022, 59(8): 0810008.
- [6] BRUNO A, GUGLIUZZA F, PIRRONE R, et al. A multi-scale colour and keypoint density-based approach for visual saliency detection [J]. IEEE Access, 2020, 8: 121330-121343.
- [7] LING Y. Visual saliency detection based on convolutional features fusion [D]. Wuxi: Jiangnan University, 2020.
凌艳. 基于卷积特征融合的视觉显著性检测算法研究 [D]. 无锡: 江南大学, 2020.
- [8] IMAMOGLU N, SHIMODA W, ZHANG C, et al. An integration of bottom-up and top-down salient cues on RGB-D data: saliency from objectness versus non-objectness [J]. Signal Image & Video Processing, 2018, 12(4): 1-8.
- [9] ARIVAZHAGAN S, SHEBIAH R N, HARINI R, et al. Human action recognition from RGB-D data using complete local binary pattern [J]. Cognitive Systems Research, 2019, 58: 94-104.
- [10] LI C Y, CONG R, PIAO Y, et al. RGB-D salient object detection with cross-modality modulation and selection [C]//ECCV 2022. Lecture Notes in Computer Science, August 23-28, 2020, Glasgow, UK. Cham: Springer, 2020, 12353: 225-241.
- [11] SINGH S K, SRIVASTAVA R. SL-Net: self-learning and mutual attention-based distinguished window for RGBD complex salient object detection [J]. Neural Computing and Applications, 2023, 35: 595-609.
- [12] ZHU C, CAI X, HUANG K, et al. PDNet: Prior-model guided depth-enhanced network for salient object detection [C]//2019 IEEE International Conference on Multimedia and Expo (ICME), July 8-12, 2019, Shanghai, China. New York: IEEE, 2019: 199-204.
- [13] DAS D K, SHIT S, RAY D N. Depth-guided two-way saliency network for 2D images [M]//Advanced Computational Paradigms and Hybrid Intelligent Computing. Singapore, Springer, 2022: 61-71.
- [14] XIE S, GIRSHICK R, DOLLAR P, et al. Aggregated resid-

- ual transformations for deep neural networks[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017:1492-1500.
- [15] WOO S, PARK J, LEE J Y, et al. Cbam: Convolutional block attention module[C]//European Conference on Computer Vision (ECCV), September 8-14, 2018, Munich, Germany. Cham: Springer, 2018:3-19.
- [16] ZHANG W, JIANG Y, FU K, et al. BTS-Net: Bi-directional transfer-and-selection network for RGB-D salient object detection[C]//2021 IEEE International Conference on Multimedia and Expo (ICME), July 5-9, 2021, Virtual. New York: IEEE, 2021:1-6.
- [17] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(4): 834-848.
- [18] JU R, GE L, GENG W, et al. Depth saliency based on anisotropic center-surround difference[C]//2014 IEEE International Conference on Image Processing (ICIP), October 27-30, 2014, Paris, France. New York: IEEE, 2014: 1115-1119.
- [19] PENG H, LI B, XIONG W, et al. RGBD salient object detection: A benchmark and algorithms[C]//European conference on computer vision (ECCV), September 6-12, 2014, Zurich, Switzerland. Cham: Springer 2014:92-109.
- [20] NIU Y, GENG Y, LI X, et al. Leveraging stereopsis for saliency analysis[C]//2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 16-21, 2012, Providence, Rhode Island. New York: IEEE, 2012: 454-461.
- [21] FAN D P, LIN Z, ZHANG Z, et al. Rethinking RGB-D salient object detection: Models, data sets, and large-scale benchmarks[J]. IEEE Transactions on Neural Networks and Learning Systems, 2020, 32(5): 2075-2089.
- [22] CHENG Y, FU H, WEI X, et al. Depth enhanced saliency detection method[C]//ICIMCS'14: International Conference on Internet Multimedia Computing and Service, July 10-12, 2014, Xiamen, China. New York: ACM, 2014: 23-27.
- [23] PASZKE A, GROSS S, MASSA F, et al. Pytorch: an imperative style, high-performance deep learning library[C]//Advances in Neural Information Processing Systems, 2019, December 8-14, 2019, Vancouver, Canada. New York: IEEE, 2019:8026-8037.
- [24] KINGMA D, BA J. Adam: a method for stochastic optimization[EB/OL]. (2014-12-22) [2022-06-16]. <https://arxiv.org/abs/1412.6980>.
- [25] ACHANTA R, HEMAMI S, ESTRADA F, et al. Frequency-tuned salient region detection[C]//2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 20-25, 2009, Miami, FL. New York: IEEE, 2009:1597-1604.
- [26] BORJI A, CHENG M M, JIANG H, et al. Salient object detection: A benchmark[J]. IEEE Transactions on Image Processing, 2015, 24(12): 5706-5722.
- [27] FAN D P, CHENG M M, LIU Y, et al. Structure-measure: A new way to evaluate foreground maps[C]//2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE, 2017: 4548-4557.
- [28] WU Y H, LIU Y, XU J, et al. MobileSal: Extremely efficient RGB-D salient object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(12): 10261-1026.
- [29] LI G Y, LIU Z, YE L, et al. Cross-modal weighting network for RGB-D salient object detection[C]//Computer Vision-ECCV 2020, August 23-28, 2020, Glasgow, UK. Cham: Springer, 2020, 12362: 665-681.
- [30] PIAO Y, RONG Z, ZHANG M, et al. A² dele: Adaptive and attentive depth distiller for efficient RGB-D salient object detection[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE, 2020: 9060-9069.
- [31] JI W, YAN G, LI J, et al. DMRA: Depth-induced multi-scale recurrent attention network for RGB-D saliency detection[J]. IEEE Transactions on Image Processing, 2022, 31: 2321-2336.
- [32] CHEN T, HU X, XIAO J, et al. CFIDNet: cascaded feature interaction decoder for RGB-D salient object detection[J]. Neural Computing and Applications, 2022, 34: 7547-7563.

作者简介:

化春键 (1975—),男,博士,副教授,硕士生导师,主要从事机器视觉、深度学习方面的研究。