基于特征增强的 RGB-D 显著性目标检测

刘译善,孙 涵
(南京航空航天大学 计算机科学与技术学院/人工智能学院/软件学院,江苏 南京 211100)

摘 要:显著性目标检测方法中,深度(Depth)信息的引入能弥补 RGB 图像缺失的空间信息,有助于从复杂的背景中检测显著目标,提升检测精度。但如何有效融合跨模态特征、获取清晰的边界是值得研究的问题。该文设计了一个基于特征增强的 RGB-D 显著性目标检测网络 FENet(Feature Enhancement Network),首先由特征融合增强模块(Feature Fusion Enhancement Model,FFEM),通过交叉融合和混合空间/通道注意力充分利用跨模态特征的相关性和互补性提取高级语义信息,然后通过边界特征增强模块(Boundary Feature Enhancement Model,BFEM)对浅层细节信息进行补充,并引入门控避免低质量底层信息的干扰,最后通过混合增强损失函数来完成模型对显著区域和边界的学习。FENet 模型在五个公开数据集上和当前较为先进的模型相比,有效提升了检测性能,尤其在显著物体的边缘细化和完整性检测上。

关键词:显著性目标检测;深度学习;边界特征增强;特征融合增强;多模态
中图分类号:TP391 **文献标识码**:A **文章编号**:1673-629X(2023)11-0028-07
doi:10.3969/j.issn.1673-629X.2023.11.005

Feature Enhancement Based RGB-D Salient Object Detection

LIU Yi-shan,SUN Han
(School of Computer Science and Technology,School of Artificial Intelligence,
School of Software,NUAA,Nanjing 211100,China)

Abstract:The addition of depth information to salient object detection can fill in the spatial information that RGB images lack,enabling the accurate detection of a salient object in the complex background. But how to fuse cross-modal features and obtain distinct boundaries is the challenge in RGB-D salient object detection. We design a feature-enhanced RGB-D salient object detection network (FENet). Firstly,high-level semantics are extracted using the feature fusion enhancement module (FFEM),which fully exploits the correlation and complementarity of RGB and depth information through cross-fusion and spatial/channel attention. The boundary feature enhancement module (BFEM) is then employed to enrich the shallow details information,and the saliency map depurator unit is used to prevent the entry of poor data sources. Finally,the learning of the salient regions and edges of the model is completed by the hybrid enhancement loss function. The FENet model proposed effectively improves the detection performance compared with the current advanced models on four public datasets,particularly in edge refinement and integrity of the predicted salient regions.

Key words:salient object detection;deep learning;boundary feature enhancement;feature fusion enhancement;multi-modal

0 引 言

显著性目标检测旨在模拟人类视觉注意系统,检测场景中最为显著的物体。作为计算机视觉任务中非常重要的预处理步骤之一,在立体匹配[1]、图像理解[2]、动作识别[3]、视频检测及分割[4]、语义分割[5]、医学图像分割[6]、目标跟踪[7]、行人重识别[8]、伪装目标检测[9]、图像检索[10]等领域中发挥着非常重要的作用[11-12]。早期基于 RGB 图像的显著性目标检测在面对复杂背景、光照变化等挑战性因素时难以取得理想效果,随着 Microsoft Kinect 等深度传感器的广泛使用,研究人员将深度图像引入,在检测中起到了较好地区分前景和背景的作用。但在跨模态特征融合、边界细化等问题上还需进一步探索。近几年,越来越多的研究工作采用中期融合策略实现跨模态特征融合,以此提升检测模型性能[13-15],考虑到只对边界进行增强容易导致检测的显著目标不完整,只对语义进行增强则会导致边界不准确。受文献[16-20]等相关工作的启发,该文提出一种基于特征增强的网络结构,同时增

收稿日期:2022-11-25 **修回日期**:2023-03-28
基金项目:中央高校基本科研业务费专项资金(NZ2019009)
作者简介:刘译善(1996-),女,硕士研究生,CCF 会员(N2104G),通讯作者,研究方向为显著性目标检测;孙 涵(1978-),男,博士,副教授,CCF 高级会员(33361S),研究方向为图像处理、计算机视觉等。

强语义和边界,以此获得边界清晰、完整的显著目标,设计模块单独捕捉边界信息的同时引入门控机制,选择丢弃或者保留引入了边界信息的显著图,以避免当边界信息捕捉效果不佳时破坏显著图质量的情况。首先特征融合增强模块(FFEM)交叉融合后通过混合注意力提取跨模态特征,提升模型对高层语义信息的捕捉。然后,考虑到深度信息有更明确的边界特征[21],通过边界特征增强模块(BFEM)对包含丰富细节信息的底层特征进行提取,为了避免噪声的引入,进一步设计门控,对低质量边界信息进行舍弃。最后通过混合增强损失对模型进行优化。所提出的模型在五个具有挑战性的数据集上进行实验,与当前主流的 RGB-D 显著性目标检测方法进行对比,达到了良好的检测效果。

1　相关工作

传统 RGB-D 显著性目标检测研究工作依赖于手工提取的特征。2012 年,首个 RGB-D 显著性目标检测模型 DM[22]将深度先验集成到显著性检测模型中,并提出了从 2D 和 3D 场景中收集的包含 600 张图像的 NUS-3D 数据集。此后各类研究方法陆续出现,如基于对比度[23-24]、形状[25]等手工特征,通过马尔可夫随机场[26]、高斯差分[27]和图知识[28]等方式进行建模的检测模型。除此之外,一些研究还尝试将传统方法组合来集成 RGB 和深度特征,如随机森林回归器[29]、角密度[30]等。但受到低水平显著性线索的限制,传统方法在复杂场景下的泛化性能较弱。随着深度学习在计算机视觉领域的应用,RGB-D 显著性目标检测也取得突破进展。

2017 年,Qu 等人[31]首次将卷积神经网络应用到 RGB-D 显著性目标检测模型中,将传统方法基于超像素的拉普拉斯传播框架与训练后的 CNN 相结合,通过利用输入图像的内在结构来提取空间一致的显著图。早期基于深度学习的显著性目标检测方法简单使用全连接层[32-33],容易破坏数据的空间结构信息。目前更多的研究方法使用全卷积神经网络网络[34-36],能够缓解这一问题。根据跨模态特征融合阶段的不同,常常将相关研究方法分为早期融合[36]、中期融合[37-38]和后期融合[39]三个类别,中期融合是对另外两者的补足,能够从两种模态中学习高层语义,因此也是最常用的特征融合策略。尽管 RGB-D 显著性目标检测当前已经取得了突破性进展[16,31,34,39-41],但仍在以下两个方面存在一定的提升空间。

一是显著物体检测的完整性。目前已有方法无法在有效进行跨模态特征提取和融合的同时捕捉两种模态的相互作用,且鲜有检测模型明确利用两种模态的特异性,导致最终显著图不能够完整、正确地描述显著目标。该文设计的 FFEM 模块通过交叉融合和混合注意力,在利用跨模态特征互补性的同时充分利用了二者的相关性,消融实验部分验证了该模块的有效性。

二是显著物体的边界清晰度。当前研究大多集中在区域精度上不在边界质量上,且通过一个步骤同时捕捉图片的语义信息和边界细节,导致最终显著图边界模糊。针对这一问题,该文设计的 BFEM 模块对边界特征进行单独提取和增强,设计门控避免低质量信息干扰。除此之外,显著性目标检测方法中常用的损失函数交叉熵损失在判别边界像素点时,通常置信度都比较低,容易导致边界模糊。通过对区域和边界进行约束,以获得最终最优的检测结果。相关设计同样在消融实验部分验证了其有效性。

2　文中方法

该文提出的 FENet 网络结构如图 1 所示,采用端到端的模型。首先,使用两个 ResNet-50 残差网络分别提取 RGB 信息流和深度信息流的特征,表示为 $r_i(i=0,1,\cdots,4)$ 和 $d_i(i=0,1,\cdots,4)$;然后,由特征融合增强模块 FFEM 实现不同尺度的跨模态特征的逐级融合,同时充分利用跨模态特征的差异性对强化后的跨模态特征进行信息补充和完善;最后,通过边界特征增强模块 BFEM, 从前三层浅层特征中获取更精确的边界信息,通过门控来抑制低质量深度图信息的影响,以生成最终高质量的显著图。所设计的特征融合增强模块 FFEM 和边界特征增强模块 BFEM 在 2.1 和 2.2 两个小节进行详细介绍。

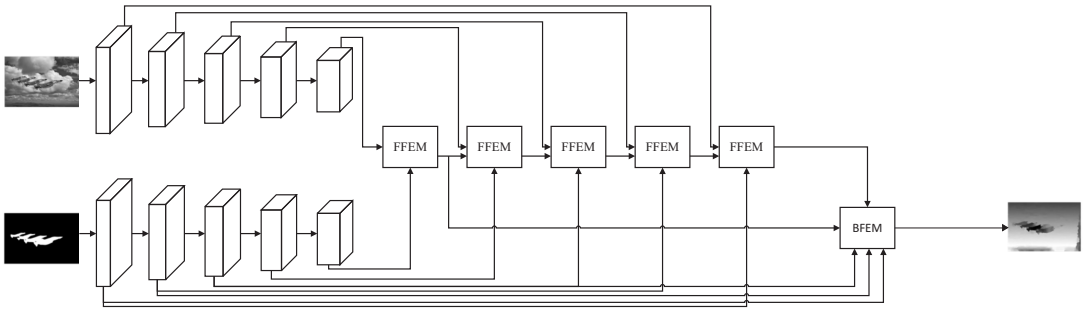

图 1　FENet 网络框架示意图

2.1 特征融合增强模块(FFEM)

目前已有方法融合 RGB 和深度信息流特征时,在考虑二者相关性的同时常常容易忽略差异性,导致融合过程中容易丢失细节信息。该文设计的 FFEM 模块充分利用跨模态特征相关性进行特征自增强,即 RGB 和深度信息流特征通过交叉相乘和混合注意力,在互补特征的引导下进行自增强,再通过原始特征信息的补充完善特征,将自增强后跨模态特征拼接融合后通过 3 × 3 卷积进行特征提取,跨模态特征逐级融合以不断强化特征信息,如图 2 所示。

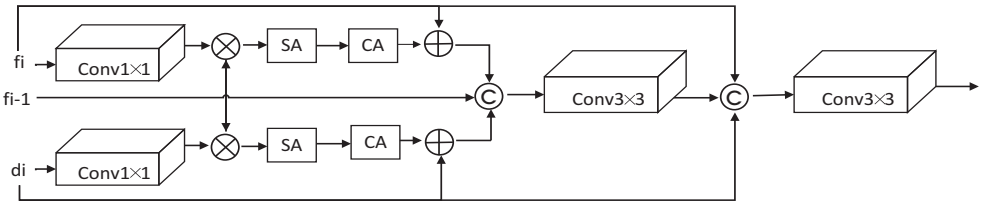

图 2 特征融合增强模块结构

具体来说,首先通过 1 × 1 的卷积对通道进行压缩,之后采用跨模态特征两两交叉相乘的方式放大 RGB 和深度特征的相关性,抑制不相关特征,进而达到突出显著特征的目的。如下公式所示,Conv(·)表示卷积操作:

$$F_r = \mathrm{Conv}_{1\times1}(r_i) \otimes \mathrm{Conv}_{1\times1}(d_i) \tag{1}$$

$$F_r = \mathrm{Conv}_{1\times1}(d_i) \otimes \mathrm{Conv}_{1\times1}(r_i) \tag{2}$$

通过混合使用空间注意力(SA)和通道注意力(CA),同时在空间维度和通道维度增强特征表达;之后跳跃连接原始跨模态特征,并与上一层的融合特征 F_{i-1} 拼接,以实现特征的逐级增强,公式如下所示:

$$F_r^{'} = \mathrm{CA}(\mathrm{SA}(F_r)) \oplus r_i \tag{3}$$

$$F_d^{'} = \mathrm{CA}(\mathrm{SA}(F_d)) \oplus d_i \tag{4}$$

自增强后的跨模态特征 $F_r^{'}$、$F_d^{'}$ 以及上一层融合特征 F_{i-1} 通过 Concat 函数进行拼接,最后通过一个 3 ×3 卷积来获取高质量显著区域,公式如下所示:

$$F_i^{'} = \mathrm{Conv}_{3\times3}(\mathrm{Concat}(F_r^{'}, F_d^{'}, F_{i-1})) \tag{5}$$

为了进一步利用跨模态特征的差异性,弥补原始跨模态特征在融合过程中的损耗,将 r_i 和 d_i 进行补充,公式如下所示:

$$F_i = \mathrm{Conv}_{3\times3}(\mathrm{Concat}(F_i^{'}, r_i, d_i)) \tag{6}$$

2.2 边界特征增强模块(BFEM)

将细节特征分开提取,针对浅层的低级特征设计了边界特征增强模块 BFEM,以提取清晰边界特征,如图 3 所示。

考虑到高级语义特征能够准确定位图片中显著目标的位置,而深度图边缘更突出,因此提取深度图(d_0、d_1、d_2)的细节特征。

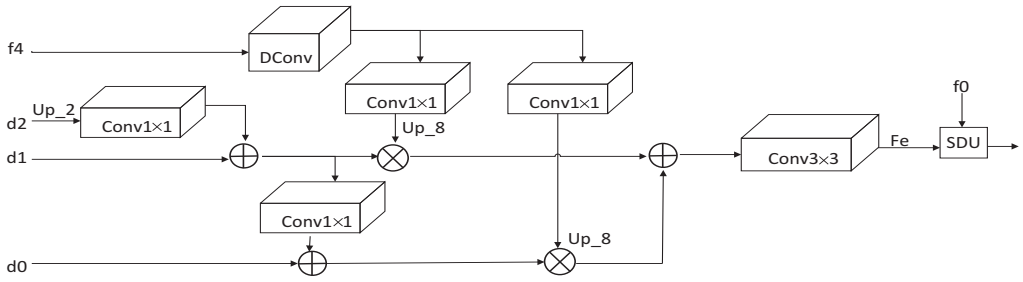

图 3 边界特征增强模块结构

不同层级的深度图特征二倍上采样后两两相加,与空洞卷积提取的多尺度特征进行相乘,增强边界的细节特征。两两增强后的细节特征相加后送入 3 × 3 卷积获取融合后的高质量显著区域。公式如下:

$$F_e = \mathrm{Conv}_{3\times3}(\mathrm{DConv}(F_i) \otimes ((d_0, d_1) \oplus (d_0, d_1, d_2))) \tag{7}$$

在过往的研究工作中发现,底层特征往往包含一定的噪声,为避免噪声干扰,本模块还设计了门控 SDU,将本模块获得的显著图 S 与前序阶段获得的显著图 S_M 和真值图对比,计算各自的 MAE 值完成比较,取得分高者作为最终的显著性目标检测图输出。

2.3 损失函数

该网络结构的损失函数由两部分构成,结构损失和边界损失。二元交叉熵(BCE)是应用最广泛的损失函数,但 BCE 损失独立计算每个像素的损失,忽略图像全局结构,同时在背景占优势的图片中,前景像素的损失会被稀释。因此,针对高级感受野提取的区域显著性将更关注于困难像素点的二进制交叉熵损失 BCE 和全局结构的加权交并比损失 IoU 相结合,即:

$$L_r = L_{\mathrm{wbce}} + L_{\mathrm{wIoU}} \tag{8}$$

为了进一步增强对边缘的监管力度,对边缘附近区域进行了约束和优化。公式如下:

$$L_e = \frac{\sum_{h}^{H}\sum_{w}^{W}(e * |S - G|)}{\sum_{h}^{H}\sum_{w}^{W} e} \tag{9}$$

$$e = \begin{cases} 0, \text{if}(G - P(G)_{[h,w]}) = 0 \\ 1, \text{if}(G - P(G)_{[h,w]}) \neq 0 \end{cases} \tag{10}$$

其中,H、W 分别表示图片的高和宽,L_e 表示边缘增强损失,$P(\cdot)$ 表示具有 5×5 滑动窗口的平均池化操作,通过 e 来获取真值图轮廓附近局部区域,以达到优化显著物体轮廓的目的。S 为获得的显著图,G 为真值图。综上,总的损失函数 L 为:

$$L = L_r + L_e \tag{11}$$

3　实验和分析

3.1　数据集和评估指标

在 NJU2k[27]、NLPR[42]、DES[23]、STERE[43]、SIP[16]五个公开的 RGB-D 数据集上验证模型的有效性。其中选择 NJU2K 的 1 485 个样本和 NLPR 的 700 个样本作为训练数据集,NJU2K 和 NLPR 剩余 800 个样本以及 DES、STERE、SIP 五个数据集的样本作为测试集。实验过程中采用 F 指标[44]、平均绝对误差[45]、S 指标[46]和 E 指标[47]进行评估。F 指标对准确度和完整度进行综合判断,计算公式如下:

$$F_\beta = \frac{(1+\beta^2)\,\text{Precision} \times \text{Recall}}{\beta^2\,\text{Precision} + \text{Recall}} \tag{12}$$

其中,β^2 根据很多显著性目标检测工作经验设置为 0.3,Precision 为正确率,Recall 为召回率。平均绝对误差(MAE)用来评估显著图 S 和真值图 G 之间的逐像素平均绝对误差,计算公式如下:

$$\text{MAE} = \frac{1}{W \times H}\sum_{x=1}^{W}\sum_{y=1}^{H}|S(x,y) - G(x,y)| \tag{13}$$

其中,W 和 H 分别表示显著图的宽和高,$S(x,y)$ 为模型检测得到的显著图,$G(x,y)$ 为真值图。MAE 的值越小,模型的性能越好。S 指标用来评估区域感知(S_r)和目标感知(S_o)之间的结构相似性,定义为:

$$S_\alpha = \alpha S_0 + (1-\alpha)S_r \tag{14}$$

其中,α 是取自区间[0,1]的平衡参数,在文中设置为 0.5。E 指标在认知视觉研究的基础上提出,用于获取图像级统计信息和局部像素匹配信息,计算公式如下:

$$E_\varphi = \frac{1}{W \times H}\sum_{x=1}^{W}\sum_{y=1}^{H}\varphi\text{FM}(x,y) \tag{15}$$

其中,φFM 表示增强对角矩阵[47]。

3.2　实施细节

所提出的模型基于 PyTorch 网络框架,主干网络 Res2Net-50[48]在 ImageNet[49]上进行预训练。GPU 为 NVIDIA TITAN XP,显存大小为 12 GB。训练过程中学习率设置为 1e-4,迭代次数 200。训练阶段通过随机翻转、旋转等策略进行数据增强,测试阶段最终输出的显著图重新调整到原来的大小。

3.3　与前沿方法对比

将所提出的方法与多种显著性目标检测方法,即基于深度方法的 DMRA[50]、ICNet[41]、HDFNet[40]、UC-Net[51]、D3Net[16]、DQSP[52]、DSA2F[53]、SPSN[54],进行比较。表 1 列出了上述方法在五个数据集上 F 指标、平均绝对误差、S 指标和 E 指标的对比情况。其中 F 指标、S 指标和 E 指标数值越大表示模型性能越好,MAE 则是数值越小表示模型性能越好。从对比结果可以看出,FENet 模型在五个数据集上均取得了较好的检测结果,尤其在图片场景多以日常真实场景为主的 NLPR 和 STERE 数据集上,相较于其他基于深度学习的方法,F 指标均提升了近 1%,模型的泛化性能得到加强。在 MAE 和 E 指标上,总体也得到了提升,虽然在 DES 和 SIP 两个数据集上的结果要略低于 UC-Net 模型和 SPSN 模型,但 FENet 模型在这两个数据集上的 F 指标和 S 指标分别更高于两个模型,这也契合在设计该模型时更聚焦于跨模态特征相关性、特异性进而提升检测结果完整性的探索,达到最终显著图在准确度和完整度上的综合判断。

表 1　FENet 模型与不同深度方法基准测试结果对比

Method		DMRA	ICNet	HDFNet	UCNet	D3Net	DQSD	DSA2F	SPSN	Ours
NJU2k	$F_\beta\uparrow$	0.904	0.903	0.922	0.886	0.910	0.899	0.916	0.927	0.936
	MAE ↓	0.052	0.052	0.038	0.043	0.047	0.052	0.039	0.033	0.032
	$S_\alpha\uparrow$	0.886	0.903	0.908	0.897	0.900	0.897	0.904	0.918	0.920
	$E_\varnothing\uparrow$	0.906	0.906	0.939	0.930	0.928	0.906	0.916	0.950	0.950
NLPR	$F_\beta\uparrow$	0.889	0.919	0.926	0.920	0.907	0.909	0.915	0.918	0.929
	MAE ↓	0.031	0.028	0.023	0.023	0.030	0.029	0.024	0.023	0.021
	$S_\alpha\uparrow$	0.899	0.923	0.923	0.922	0.912	0.916	0.919	0.925	0.931
	$E_\varnothing\uparrow$	0.932	0.944	0.957	0.955	0.942	0.939	0.950	0.955	0.962
DES	$F_\beta\uparrow$	0.898	0.925	0.938	0.943	0.909	0.939	0.930	0.942	0.946
	MAE ↓	0.030	0.027	0.021	0.017	0.031	0.021	0.023	0.017	0.016
	$S_\alpha\uparrow$	0.900	0.920	0.931	0.931	0.897	0.935	0.917	0.937	0.941
	$E_\varnothing\uparrow$	0.928	0.948	0.963	0.968	0.923	0.962	0.950	0.973	0.974

续表 1

Method		DMRA	ICNet	HDFNet	UCNet	D3Net	DQSD	DSA2F	SPSN	Ours
STERE	F_β ↑	0.908	0.907	0.910	0.913	0.904	0.901	0.898	0.908	0.916
	MAE ↓	0.047	0.045	0.042	0.035	0.046	0.051	0.036	0.043	0.037
	S_α ↑	0.886	0.903	0.900	0.906	0.899	0.892	0.904	0.901	0.907
	$E_\varnothing$ ↑	0.923	0.927	0.939	0.943	0.925	0.920	0.933	0.935	0.942
SIP	F_β ↑	0.831	0.873	0.909	0.903	0.880	0.890	0.891	0.910	0.917
	MAE ↓	0.085	0.069	0.048	0.045	0.063	0.065	0.057	0.043	0.044
	S_α ↑	0.816	0.854	0.886	0.882	0.860	0.864	0.862	0.892	0.894
	$E_\varnothing$ ↑	0.860	0.892	0.924	0.925	0.897	0.890	0.908	0.932	0.930

注：↑ & ↓ 分别表示越大越好和越小越好。

基于深度学习方法的可视化结果对比如图 4 所示，对比第 1、3 行结果可以看到，在图片背景中存在干扰，如第 1 行的背景凹陷部分以及第 3 行人的左侧与背景中的树木衔接部分容易被误判为显著目标的一部分，FENet 模型相较于另外几个模型能够尽可能避免背景干扰，同时完整、准确地切割出显著目标；对比第 2 行结果可以看到，当面对显著目标中包含容易漏检的细小部分情况时，如图中蝴蝶的各个触角，相较于其他模型漏检触角、边界模糊等的问题，文中模型能够以较为清晰的边界较好地检测出显著目标；对比第 4、6 行可以看到，当面对光照和阴影变化等情况时，相较于其他模型对于显著目标内部检测不完整、阴影部分未完整检测出的情况，文中模型在检测的完整度和清晰度上要高于其他模型；对比第 5 行可以看到，当面对多个显著目标时，虽然图中存在多检测了背景中部分人影的情况，但实际的两个显著目标，文中模型相较于其他模型能够更完整地检测出来。可以看出，该文所设计的分层增强语义和边界特征的 FENet 模型在显著目标的完整性和边界清晰度上取得了较为理想的效果。

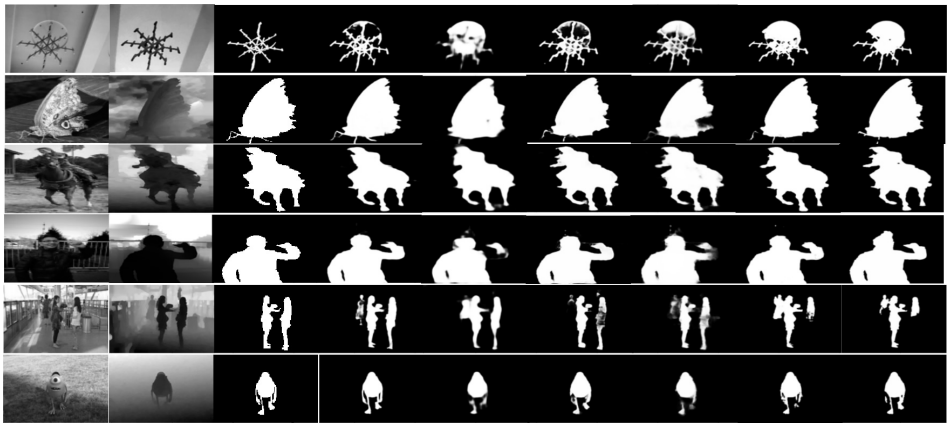

图 4 FENet 网络与前沿的 RGB-D 显著性目标检测模型的可视化比较

3.4 消融实验

为验证文中相应模块设计的有效性，进行了消融实验，相关数据对比见表 2。比较第 1、2 行可以看出，FFEM 模块增强了模型的性能，在四个指标上均有了不同幅度的提升，其中在 F 指标和 E 指标上提升了 0.5% 左右，在 DES 数据集上结构相似性指标也有了 1% 的提升；比较第 2、3 行可以看出，BFEM 模块的加入后在两个数据集的 F 指标和 E 指标上均提升了 0.5% 左右；比较第 3、4 行可以看到混合损失函数的使用在两个数据集上的四个指标上给模型性能带来了不同程度的提升，更契合预期。

表 2 FENet 模型在 STERE 和 DES 数据集上进行消融实验的结果对比

Method	STERE				DES			
	F_β ↑	MAE ↓	S_α ↑	$E_\varnothing$ ↑	F_β ↑	MAE ↓	S_α ↑	$E_\varnothing$ ↑
Baseline	0.902	0.045	0.896	0.928	0.934	0.023	0.917	0.950
Baseline+FFEM	0.906	0.041	0.901	0.934	0.938	0.020	0.928	0.962
Baseline+FFEM +BFEM	0.913	0.038	0.905	0.938	0.944	0.018	0.938	0.967
Baseline+FFEM+BFEM+loss	0.916	0.037	0.907	0.942	0.946	0.016	0.941	0.974

4　结束语

提出了一种 RGB-D 显著性目标检测框架,该框架通过特征融合增强模块和边界特征增强模块分别对高级语义信息和底层细节信息进行处理。实验结果表明,该框架是可行的,在主流的五个数据集上相较于前沿的方法取得了不错的效果,所设计的模块也通过消融实验进行了验证。

参考文献:

[1] NIE G Y, CHENG M M, LIU Y, et al. Multi-level context ultra-aggregation for stereo matching[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Long Beach: IEEE, 2019: 3283-3291.

[2] ZHU J Y, WU J, XU Y, et al. Unsupervised object class discovery via saliency-guided multiple class learning[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 37(4): 862-875.

[3] RAPANTZIKOS K, AVRITHIS Y, KOLLIAS S. Dense saliency-based spatiotemporal feature points for action recognition[C]//Proceedings of 2009 IEEE conference on computer vision and pattern recognition. Miami: IEEE, 2009: 1454-1461.

[4] WANG W, SHEN J, YANG R, et al. Saliency-aware video object segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 40(1): 20-33.

[5] HOYER L, MUNOZ M, KATIYAR P, et al. Grid saliency for context explanations of semantic segmentation[J]. Advances in Neural Information Processing Systems, 2019, 32: 6462-6473.

[6] WU Y H, GAO S H, MEI J, et al. JCS: an explainable covid-19 diagnosis system by joint classification and segmentation[J]. IEEE Transactions on Image Processing, 2021, 30: 3113-3126.

[7] HONG S, YOU T, KWAK S, et al. Online tracking by learning discriminative saliency map with convolutional neural network[C]//Proceedings of international conference on machine learning. [s. l.]: PMLR, 2015: 597-606.

[8] ZHAO R, OYANG W, WANG X. Person re-identification by saliency learning[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(2): 356-370.

[9] FAN D P, JI G P, SUN G, et al. Camouflaged object detection[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Seattle: IEEE, 2020: 2777-2787.

[10] LIU G, FAN D. A model of visual attention for natural image retrieval[C]//Proceedings of 2013 international conference on information science and cloud computing companion. Guangzhou: IEEE, 2013: 728-733.

[11] 金海燕,肖照林,蔡　磊,等. 显著性目标检测理论与应用研究综述[J]. 计算机技术与发展,2022,32(9):1-7.

[12] 丛润民,张　晨,徐　迈,等. 深度学习时代下的 RGB-D 显著性目标检测研究进展[J]. 软件学报,2023,34(4):1711-1731.

[13] ZHOU T, FAN D P, CHENG M M, et al. RGB-D salient object detection: a survey[J]. Computational Visual Media, 2021, 7(1): 37-69.

[14] 张卫明,史彩娟,任弼娟,等. 多尺度特征金字塔网格的显著性目标检测[J]. 小型微型计算机系统,2022,43(5):1068-1074.

[15] 崔丽群,陈晶晶,任茜钰,等. 融合多特征与先验信息的显著性目标检测[J]. 中国图象图形学报,2020,25(2):321-332.

[16] 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.

[17] ZHOU T, FU H, CHEN G, et al. Specificity-preserving rgb-d saliency detection[C]//Proceedings of the IEEE/CVF international conference on computer vision. Montreal: IEEE, 2021: 4681-4691.

[18] 何　伟,潘　晨. 注意力引导网络的显著性目标检测[J]. 中国图象图形学报,2022,27(4):1176-1190.

[19] 罗会兰,袁　璞,童　康. 基于深度学习的显著性目标检测方法综述[J]. 电子学报,2021,49(7):1417-1427.

[20] 郭继昌,岳惠惠,张　怡,等. 图像增强对显著性目标检测的影响研究[J]. 中国图象图形学报,2022,27(7):2129-2147.

[21] LIU Z, TAN Y, HE Q, et al. SwinNet: swin transformer drives edge-aware RGB-D and RGB-T salient object detection[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 32(7): 4486-4497.

[22] LANG C, NGUYEN T V, KATTI H, et al. Depth matters: influence of depth cues on visual saliency[C]//Computer vision - ECCV 2012: 12th European conference on computer vision. Florence: Springer, 2012: 101-115.

[23] CHENG Y, FU H, WEI X, et al. Depth enhanced saliency detection method[C]//Proceedings of international conference on internet multimedia computing and service. [s. l.]: ACM, 2014: 23-27.

[24] FAN X, LIU Z, SUN G, et al. Salient region detection for stereoscopic images[C]//2014 19th international conference on digital signal processing. Hong Kong: [s. n.], 2014: 454-458.

[25] CIPTADI A, HERMANS T, REHG J. An in depth view of saliency[C]//Proceedings of 2013 British machine vision conference. Bristol: BMVA Press, 2013: 112.1-112.11.

[26] REN J, GONG Xiaojin, YU L, et al. Exploiting global priors for RGB-D saliency detection[C]//2015 IEEE conference on computer vision and pattern recognition workshops (CVPRW). Boston: IEEE, 2015: 25-32.

[27] JU R, GE L, GENG W, et al. Depth saliency based on anisotropic center-surround difference[C]//Proceedings of 2014 IEEE international conference on image processing (ICIP).

Paris: IEEE,2014:1115-1119.
[28] CONG R, LEI J, ZHANG C, et al. Saliency detection for stereoscopic images based on depth confidence analysis and multiple cues fusion[J]. IEEE Signal Processing Letters, 2016,23(6):819-823.
[29] DU H,LIU Z,SONG H,et al. Improving RGBD saliency detection using progressive region classification and saliency fusion[J]. IEEE Access,2016,4:8987-8994.
[30] FENG D,BARNES N,YOU S,et al. Local background enclosure for RGB-D salient object detection[C]//2016 IEEE conference on computer vision and pattern recognition (CVPR). Las Vegas:IEEE,2016:2343-2350.
[31] QU L,HE S,ZHANG J,et al. RGBD salient object detection via deep fusion[J]. IEEE Transactions on Image Processing, 2017,26(5):2274-2285.
[32] HAN J,CHEN H,LIU N,et al. CNNs-based RGB-D saliency detection via cross-view transfer and multiview fusion [J]. IEEE Transactions on Cybernetics,2017,48(11):3171-3183.
[33] 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). Shanghai:IEEE,2019:199-204.
[34] CHEN H,LI Y. Three-stream attention-aware network for RGB-D salient object detection[J]. IEEE Transactions on Image Processing,2019,28(6):2825-2835.
[35] LONG J,SHELHAMER E,DARRELL T. Fully convolutional networks for semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. [s. l.]:IEEE,2015:3431-3440.
[36] LIU Z,SHI S,DUAN Q,et al. Salient object detection for RGB-D image by single stream recurrent convolution neural network[J]. Neurocomputing,2019,363:46-57.
[37] FAN D P,ZHAI Y,BORJI A,et al. BBS-Net:RGB-D salient object detection with a bifurcated backbone strategy network[C]//Proceedings of European conference on computer vision. [s. l.]:Springer,2020:275-292.
[38] FU K,FAN D P,JI G P,et al. JL-DCF:joint learning and densely-cooperative fusion framework for RGB-D salient object detection[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Seattle: IEEE,2020:3052-3062.
[39] WANG N,GONG X. Adaptive fusion for RGB-D salient object detection[J]. IEEE Access,2019,7:55277-55284.
[40] PANG Y,ZHANG L,ZHAO X,et al. Hierarchical dynamic filtering network for RGB-D salient object detection[C]// Proceedings of European conference on computer vision. [s. l.]:Springer,2020:235-252.
[41] LI G,LIU Z,LING H. ICNet:information conversion network for RGB-D based salient object detection[J]. IEEE Transactions on Image Processing,2020,29:4873-4884.
[42] PENG H,LI B,XIONG W,et al. RGBD salient object detection:a benchmark and algorithms[C]//Proceedings of European conference on computer vision. [s. l.]:Springer,2014: 92-109.
[43] NIU Y,GENG Y,LI X,et al. Leveraging stereopsis for saliency analysis[C]//2012 IEEE conference on computer vision and pattern recognition. San Francisco:IEEE,2012:454-461.
[44] 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.
[45] PERAZZI F,KRÄHENBÜHL P,PRITCH Y,et al. Saliency filters:contrast based filtering for salient region detection [C]//Proceedings of 2012 IEEE conference on computer vision and pattern recognition. Providence:IEEE,2012:733-740.
[46] CHENG M M,FAN D P. Structure-measure:a new way to evaluate foreground maps[J]. International Journal of Computer Vision,2021,129(9):2622-2638.
[47] FAN D P,GONG C,CAO Y,et al. Enhanced-alignment measure for binary foreground map evaluation[J]. arXiv: 1805.10421,2018.
[48] GAO S H,CHENG M M,ZHAO K,et al. Res2net:a new multi-scale backbone architecture[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2019,43(2):652-662.
[49] RUSSAKOVSKY O,DENG J,SU H,et al. Imagenet large scale visual recognition challenge[J]. International Journal of Computer Vision,2015,115(3):211-252.
[50] PIAO Y,JI W,LI J,et al. Depth-induced multi-scale recurrent attention network for saliency detection [C]//2019 IEEE/CVF international conference on computer vision (ICCV). Seoul:IEEE,2019:7253-7262.
[51] ZHANG J. UC-Net:uncertainty inspired RGB-D saliency detection via conditional variational autoencoders[C]//2020 IEEE/CVF conference on computer vision and pattern recognition (CVPR). Seattle:IEEE,2020:8579-8588.
[52] CHEN C,WEI J,PENG C,et al. Depth-quality-aware salient object detection[J]. IEEE Transactions on Image Processing,2021,30:2350-2363.
[53] SUN P,ZHANG W,WANG H,et al. Deep RGB-D saliency detection with depth-sensitive attention and automatic multi-modal fusion[C]//2021 IEEE/CVF conference on computer vision and pattern recognition (CVPR). Nashville: IEEE, 2021:1407-1417.
[54] LEE M,PARK C,CHO S,et al. Spsn:superpixel prototype sampling network for rgb-d salient object detection[C]// Computer vision - ECCV 2022:17th European conference. Tel Aviv:Springer,2022:630-647.