

基于 BC^2 FNet 网络的 RGB-D 显著性目标检测

王峰^{1,2}, 程咏梅³

(1.渭南师范学院 物理与电气工程学院, 陕西 渭南 714099;
2.X 射线成像与检测陕西省高校工程研究中心, 陕西 渭南 714099; 3.西北工业大学 自动化学院, 陕西 西安 710072)

摘要:面对复杂的场景图像,深度信息的引入可以大大提高显著性目标检测的性能。然而,神经网络的上采样和下采样操作会模糊显著图中目标的边界,从而降低显著性目标检测性能。针对此问题,提出了一种基于边界驱动跨模态跨层融合网络(BC^2 FNet)的 RGB-D 显著性目标检测方法。该网络在跨模态和跨层融合中分别加入边界信息引导来保持目标区域。设计了边界生成模型,分别从 RGB 和深度模态的低层特征中提取 2 种边界信息;设计边界驱动的特征选择模块,在 RGB 与深度模态融合过程中,聚焦重要特征信息并保留边界细节;提出了一种边界驱动的跨层融合模块,在相邻层的上采样融合过程中加入 2 种边界信息。通过将该模块嵌入到自顶向下的信息融合流中,预测出包含精确目标和清晰边界的显著性图。在 5 种标准 RGB-D 数据集上进行仿真实验,结果证明所提出的模型具有较好的性能。

关键词:显著性目标检测;边界驱动;跨模态融合;跨层融合

中图分类号:TP391

文献标志码:A

文章编号:1000-2758(2024)06-1135-09

显著性目标检测(SOD)的目的是在给定的图像中定位和选择最具有视觉吸引力的对象或区域。它是计算机视觉领域(CV)的一个热门话题,具有很高的研究价值和广泛的应用场景。SOD 检测到的显著区域包含更多有用的语义信息,因此,它们经常被用于其他计算机视觉任务的预处理,如视觉跟踪^[1]、图像分割、图像质量评估、物体识别等^[2]。

早期 RGB-D 显著性目标检测主要研究 RGB 和深度图像中的有用信息,设计手工制作特征,如中心先验、边界先验和对比度。这些手工特征信息表达能力较低,导致在面对复杂场景图像时检测结果不满意。随着卷积神经网络(CNN)的快速发展,由于其在提取特征方面的优势,各种方法都使用 CNN 对 RGB-D 显著性目标进行检测,大大提高了检测性能。然而,在具体检测过程中也存在一些问题。

CNN 中上采样和下采样操作会丢失目标边界细节,影响 SOD 的整体性能。因此,如何在检测显著目标的同时保留目标边界信息是一个关键问题。

针对此问题,提出了许多 RGB-D SOD 方法以保持物体清晰的边界。根据解决方案的不同可将这些方法分为两大类:①设计不同类型的损失函数激励网络模型在训练过程中约束目标的边界,以此保证显著目标具有明确的边界^[3];②设计各种边界检测网络学习边界特征以辅助 SOD^[4]。这些方法虽然在一定程度上保留了显著目标的边界信息,但只使用了网络模型某一阶段(如跨模态融合或显著性图预测)学习到的边界信息,并没有充分利用边界信息辅助显著性目标检测,使得目标边界不够清晰。因此,本文认为提高目标精度并保留目标边界信息可以从以下 2 个方面考虑:①如何从 RGB 和深度模态中提取准确的边界细节信息;②如何有效地利用提取的边界信息加入 SOD。

基于以上 2 个方面,本文提出了一种边界驱动 BC^2 FNet 的 RGB-D SOD 方法。该网络将学习到的边界特征融入到整个 SOD 过程中,使预测的显著性图包含清晰边界的显著性目标。首先,考虑到主干

收稿日期:2023-10-11

基金项目:国家自然科学基金面上项目(61135001)、渭南市科技计划(ZDYF-JCYJ-196)与渭南师范学院校级人才项目(2020RC11)资助

作者简介:王峰(1981—),讲师

通信作者:王峰(1981—) e-mail:wangfeng81113@163.com

网提取高层特征包含更多的语义信息,而低层特征包含更多细节信息,提出了一种边界生成模块(BGM),分别从 RGB 和深度模态的低层特征中提取 2 种类型的边界信息。其次,为了充分利用提取的边界信息,分别在 SOD 的跨模态和跨层融合过程中加入边界信息。具体而言,本文设计了一种边界驱动特征选择模块(BFSM),实现 RGB 特征与深度特征的有效融合,该模块能够在保留目标边界信息的同时,重点关注重要特征。此外,本文还设计了一种边界驱动跨层融合模块(BCFM),有效整合相邻层特征,并将该模块嵌入自顶向下的信息融合流中,预测最终的显著性图。本文的贡献总结如下:

1) 针对现有 SOD 方法中目标边界不清晰问题,设计了一种边界驱动 BC²FNet 的 SOD 网络,该网络既能准确预测显著目标,又能保留目标边界。

2) 提出了一种边界生成模块,从 RGB 和深度模态的低层特征中提取 2 种边界信息,以辅助显著目标检测。

3) 分别设计了边界驱动特征选择融合模块和边界驱动跨层融合模块。在跨模态融合和跨层融合中加入 2 种边界信息,使得网络在检测显著目标过程中始终保持清晰的边界信息。

4) 在 5 种标准数据集上进行仿真实验,证明了本文提出的网络模型优于其他对比方法。

1 相关工作

1.1 RGB-D 显著性目标检测

许多 RGB SOD 方法能实现较好的检测效果。然而,当面对复杂背景和低对比度复杂图像时,其检测结果会受到目标缺失和背景干扰的影响。

考虑到易于获取的深度图像中含有利于显著目标定位的三维空间结构信息,许多学者采用 RGB 图像和深度图像进行 SOD,获得了较好的检测结果。

传统 RGB-D 显著性方法,主要利用不同类型的手工特征,如颜色、边界、全局对比度、空间先验信息等,从 RGB 和深度图像中提取有用的特征信息,并利用特征信息预测显著性图。Cong 等^[5]提出使用颜色和深度信息计算显著性图。Ren 等^[6]将区域对比度与背景、深度和方向先验相结合,得到显著性图。Liang 等^[7]设计了一种基于对比度和深度引导背景先验的 3D 立体显著性图。

随着深度学习的快速发展,与传统手工特征在

信息传递上的局限性相比,基于 CNN 提取的特征信息包含了丰富的语义信息,具有较强的表征能力。因此,许多基于 CNN 的 RGB-D SOD 方法研究取得了巨大的进展。根据 RGB 图像与深度图像的融合方法,可以分为 3 类:

1) 早期融合。将通道维度上的深度图像和 RGB 图像进行拼接,形成 4 个通道,然后,输入网络预测显著性图。

2) 后期融合。将 RGB 图像和深度图像输入到网络中生成不同类型的特征,然后将这些特征整合到网络中预测最终的显著性图。

3) 中间融合。从 RGB 图像和深度图像提取不同层次的信息,然后融合这些层次特征来预测最终的显著性图。

Chen 等^[8]提出了一种替代的细化策略来逐步融合 RGB 特征和深度特征。Li 等^[9]提出了 3 种 RGB 的深度交互模型,分别处理低层、中层和高层的跨模态信息融合。Zhang 等^[10]提出了 RGB 诱导的细节增强模块和深度诱导的语义增强模块,分别将其应用于低层特征和高层特征融合,实现差异化的跨模态融合。Zhang 等^[11]提出了一种双向传递选择网络,逐步双向融合 RGB 信息和深度信息。Li 等^[12]提出了一种分层交替交互模型,以有效增强 RGB 特征与深度特征之间的跨模态交互。这些方法预测的显著性图存在边界模糊问题。因此,本文将边界信息整合到 RGB-D SOD 中,提出了一种边界驱动的跨模态跨层融合网络。

1.2 边界显著性目标检测

CNN 中的上采样和下采样操作容易破坏显著目标的边界信息,进而降低 SOD 的性能。因此许多学者提出了不同的解决方案激励突出的物体保持清晰的边界。Qin 等^[2]通过融合二元交叉熵(BCE)、结构相似性(SSIM)和交联(IOUS)损失设计了一种混合损失函数,协助网络模型预测清晰边界的显著性图。Pang 等^[3]提出了一种包含边缘增强损失和区域增强损失的混合增强损失函数,以优化显著目标的边界细节,增强显著区域一致性。Niu 等^[1]在多尺度密集融合网络架构中加入边缘敏感损失(ESI),以保持检测到的突出区域边界。此外,也有一些方法以边界信息为中心设计不同的网络结构。Wang 等^[13]提出了一种显著边界检测模块探索边界信息。Liu 等^[14]在主网络中增加了一种预测分支评估显著目标的边界。Song 等^[15]提出了一种同时学

习显著性对象区域、边界显著性预测网络和边界保持网络的方法。Yang等^[16]设计了一种基于连通性的边缘特征提取方法,该方法直接从网络输出中提取边缘特征信息。Zhai等^[17]提出了一种分叉骨干策略网络,利用低层特征中的细节信息优化显著目标的边界。Ji等^[4]提出了一种结合边界、深度和显著性的协作学习框架。Li等^[18]提出了显著性引导位置边缘机制,同时定位显著目标,缓解边界模糊问题。Jin等^[19]设计了一种两阶段的跨模态融合方法,使用高层和低层特征分别预测初始显著性图和目标边界。

与上述方法不同的是,本文首先提出边界生成模块,分别从RGB图和其对应的深度图中学习出边界特征,然后将边界特征引入跨模态和跨层融合中,预测包含目标清晰边界的显著性图。

2 本文方法

2.1 整体结构

本文提出的BC²FNet整体结构如图1所示,该图由2个边界生成模块(BGM)、4个边界驱动特征选择模块(BFSM)和3个边界驱动跨层融合模块(BCFM)组成。其中,BFSM和BCFM模块构成本文跨模态跨层融合策略。首先采用PVTv2骨干网分别从RGB图像和深度图像中提取4层特征,使用1×1卷积层来降低提取分层特征的通道维数,得到通道维数为32的RGB特征 $\{f_r^i, i=1,2,3,4\}$ 和深度特征 $\{f_d^i, i=1,2,3,4\}$ 。其次,对RGB特征和深度特征进行合成得到提取的2类边界信息图。然后,

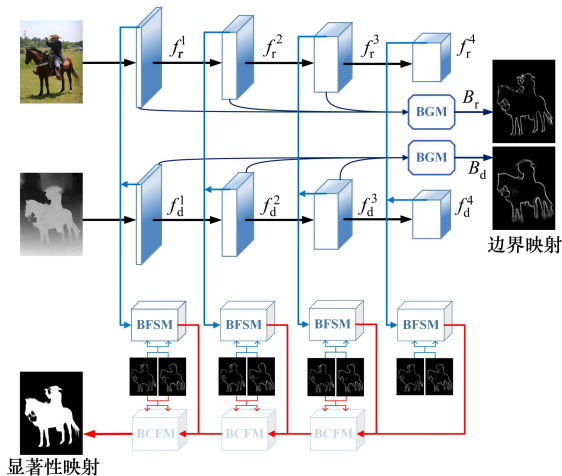


图1 本文提出BC²FNet的总体框图

通过BFSM将提取的2类边界信息结合通道和空间注意机制进行RGB特征和深度特征的跨模态融合。最后,在相邻层的上采样融合过程中加入2种边界信息,通过将BCFM模块嵌入到自顶向下的信息融合流中,得到目标边界清晰的显著性图。

2.2 边界生成模型

CNN中的上采样和下采样过程会丢失目标的边界信息,而边界信息也会深刻影响RGB-D SOD的性能。考虑到主干网络提取的高层特征包含较多的语义信息,而低层特征包含较多的边界细节,本文从低层特征中挖掘更多的边界信息以辅助SOD。如图2所示,在多目标场景下,骨干网提取4层边界信息。可以看出,前3层保留了更多的边界特征,而第4层几乎没有物体的边界信息。因此,本文设计了边界生成模块,分别从RGB和深度特征的前3层提取准确边界信息。BGM网络结构如图3所示。

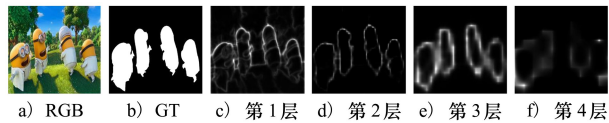


图2 利用骨干网络提取不同场景下的层次边界图

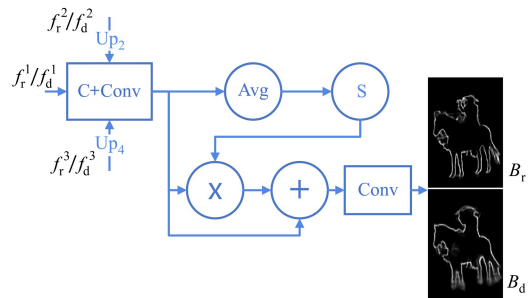


图3 BGM模型的结构图

首先,将前3层的RGB特征与深度特征进行融合,生成2个融合特征。具体操作如下:

$$F_{cat}^r = \text{Conv}(C(f_r^1, \text{Up}_2(f_r^2), \text{Up}_4(f_r^3))) \quad (1)$$

$$F_{cat}^d = \text{Conv}(C(f_d^1, \text{Up}_2(f_d^2), \text{Up}_4(f_d^3))) \quad (2)$$

式中: $\text{Conv}(\cdot)$ 指核大小为 3×3 的卷积操作; $C(\cdot)$ 表示连接运算符; $\text{Up}_2(\cdot)$ 和 $\text{Up}_4(\cdot)$ 分别表示 $2 \times$ 双线性和 $4 \times$ 双线性上采样操作。 F_{cat}^r 和 F_{cat}^d 分别表示融合RGB特征和深度特征。

然后,考虑到合成特征存在噪声干扰,本文采用平均池化和Sigmoid激活函数学习加权因子。最后,运用元素乘法和元素加法运算,将学习到的加权因子对集成特征进行加权。因此,可以同时增强边

界细节信息和抑制背景信息。具体操作为

$$W_o = S(\text{Avg}(F_{\text{cat}}^o)) \quad (3)$$

$$B_o = \text{Conv}(W_o \otimes F_{\text{cat}}^o \oplus F_{\text{cat}}^o) \quad (4)$$

式中: $o \in \{r, d\}$, W_o 表示学习到的权值因子; B_o 表示从 RGB 和深度特征获得的边界特征。 $S(\cdot)$ 表示 Sigmoid 激活函数; $\text{Avg}(\cdot)$ 为自适应平均池化操作; $\otimes$ 和 $\oplus$ 分别表示元素的乘积和元素求和。

2.3 边界驱动特征选择模型

如上所述,可将跨模态融合分为早期融合、中期融合及后期融合。本文利用中间融合策略,提出了一种边界驱动的跨模态融合特征选择模块。在跨模态融合过程中,首先使用 2 种类型的 RGB 和深度特征的边界信息引导网络学习目标的边界信息。其次,考虑到骨干网提取分层特征中存在大量冗余信息,本文引入了通道注意模型和空间注意模型,对通道维度和空间维度的重要特征信息进行关注。本文提出 BFSM 的结构如图 4a) 所示,主要包括 3 个分支:RGB 分支、深度分支和融合分支。每个分支由边界信息和一个特征选择模块(FSM)组成,目的是在保留边界信息的同时关注重要的目标信息。

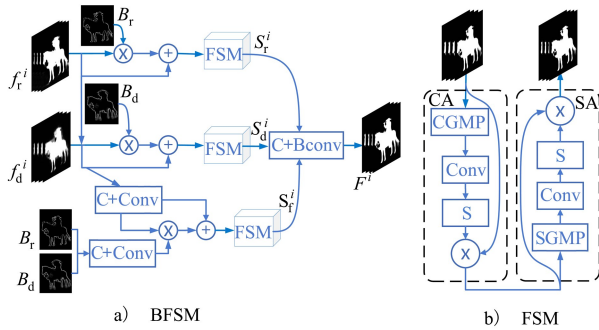


图 4 边界驱动特征选择模块和特征选择模块结构图

本文提出的 BFSM 由边界驱动和特征选择两部分组成。在 RGB 分支中加入 RGB 边界信息的引导,在深度分支中加入深度边界信息的引导,在融合分支中加入 RGB 和深度边界信息的联合引导。因此,跨模态融合特征不仅包含有模态特定的边界信息,而且包含模态融合的边界信息。其实现过程如(5)~(7)式所示。

$$B_r^i = B_r \otimes f_r^i \oplus f_r^i \quad (5)$$

$$B_d^i = B_d \otimes f_d^i \oplus f_d^i \quad (6)$$

$$B_f^i = \text{Conv}(C(B_r, B_d)) \otimes \text{Conv}(C(f_r^i, f_d^i)) \oplus \text{Conv}(C(f_r^i, f_d^i)) \quad (7)$$

式中, B_r^i, B_d^i 和 B_f^i 分别表示第 i 层 RGB 特征、深度特征和边界保留信息的融合特征。

在经过 3 种分支学习到保留边界信息的特征后,本文提出通过 3 种 FSM 来选择重要特征。FSM 具体结构如图 4b) 所示,它由通道注意(CA)和借鉴卷积块注意模型(CBAM)的空间注意(SA)组成。首先,将 B_r^i, B_d^i 和 B_f^i 输入到 CA 中得到通道注意图,该注意图引导输入特征在信道维度上选择重要的特征区域,并对其进行增强。然后,将通道注意模型选择的特征发送给 SA,在空间维度上进行新一轮特征选择,进一步强化特征。具体操作为

$$C_m^i = S(\text{Conv}(\text{CGMP}(B_m^i))) \otimes B_m^i \quad (8)$$

$$S_m^i = S(\text{Conv}(\text{SGMP}(C_m^i))) \otimes C_m^i \quad (9)$$

式中: $m \in \{r, d, f\}$, $\text{CGMP}(\cdot)$ 是沿通道维度的全局最大池化操作; $\text{SGMP}(\cdot)$ 是沿着空间维度的全局最大池化操作。最后,将 3 分支双选后的特征整合为一个特征,完成跨模态融合。操作流程为

$$F^i = \text{Bconv}(C(S_r^i, S_d^i, S_f^i)) \quad (10)$$

式中含有 1 个内核大小为 3×3 的卷积层,1 个批处理归一化层和 1 个 ReLU 激活函数层。 F^i 表示第 i 层 RGB 特征与深度特征的融合特征。

2.4 边界驱动跨层融合模型

为了更好地利用跨模态融合特征 $\{F^i, i \in \{1, 2, 3, 4\}\}$ 和 2 种边界信息 $\{B_o, o \in \{r, d\}\}$,本文提出了一种简单有效的 BCFM 模块。如图 1 所示,通过将该模块嵌入到自顶向下的信息融合流中,预测显著性图包含准确的对象和清晰的边界。提出 BCFM 模块如图 5 所示。

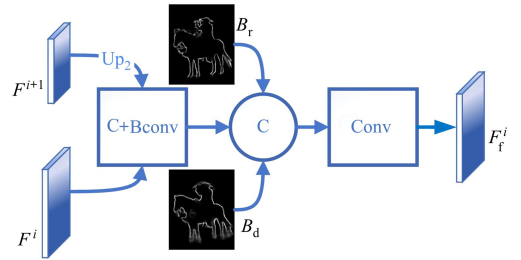


图 5 BCFM 模块框图

首先,对前一层的跨模态融合特征 F^{i+1} 进行上采样,并通过级联操作将其余当前层的跨模态融合特征 F^i 合并。考虑到上采样操作会丢失目标的边界信息,通过拼接操作将 2 类边界信息合并到跨层融合特征中,恢复丢失的边界信息。最后,通过 3×3 卷积层对通道维数进行降维,得到保留边界信

息的跨层融合特征 F_f^i 。

2.5 损失函数

1) 边界损失。本文借助 2 个边界特征 B_f 和 B_d ,通过卷积层和上采样层得到 2 个预测的边界映射 M_b^r 和 M_b^d 。将经典的 Canny 边缘检测算法应用于显著性映射真值图,得到边界映射的地面真值 (GT)。边界损失定义为

$$L_B = L_{bce}(M_b^r, G_b) + L_{bce}(M_b^d, G_b) \quad (11)$$

式中: G_b 表示边界映射地面真值; $L_{bce}(\cdot)$ 表示二值交叉熵损失函数,其表达式如 (12) 式所示。

$$L_{bce}(X, Y) = -\frac{1}{N} \sum_{i=1}^N (y_i \lg x_i + (1 - y_i) \lg (1 - x_i)) \quad (12)$$

式中, $x_i \in X, y_i \in Y; N$ 为像素总数。

2) 显著损失。本文将最后一个边界驱动的跨层融合模块生成的特征 F_f^l 输入到卷积层和上采样层中,以获得最终的显著性图 S_p 。同理,本文使用二元交叉熵损失函数 (BCE) 反映预测的显著性图 S_p 和显著性图地面真值 G_s 之间的差异。显著性损失定义为

$$L_S = L_{bce}(S_p, G_s) \quad (13)$$

因此,网络训练过程中总损失包括边界损失和显著性损失两部分。

$$L_{total} = \lambda_1 L_B + \lambda_2 L_S \quad (14)$$

式中, λ_1 和 λ_2 为权值系数,在实验中其值均取为 1.0。

3 实验结果分析

3.1 数据集

本文采用 5 种标准数据集验证所提方法的有效性,包括 NJUD^[4]、NLPR^[20]、LFSD^[21]、STERE^[22]、和 SIP^[23]。NJUD 由 1 985 幅来自互联网、3D 电影和富士 W3 立体相机拍摄的图像组成。NLPR 包含 1 000 张微软 Kinect 收集的室内和室外图像。LFSD 由 Lytro 光场相机收集的 100 幅光场图像组成,包括 60 幅室内场景和 40 幅室外场景。STERE 包含 1 000 张全部来自互联网的立体图像。SIP 由 929 张具有不同场景和各种挑战性因素的高分辨率图像组成。为了与其他方法进行公平的比较,根据以前的工作,使用来自 NJUD 的 1 485 个样本和来自 NLPR 的 700 个样本作为训练数据集。将训练集和其他数据集中剩余的数据作为测试集。

采用 S 量测 (S_α)、E 量测 (E_ξ)、F 量测 (F_β) 和 M 4 个评价指标,将本文方法与其他方法进行比较。S 量测用于计算区域感知与对象感知之间的结构相似度;E 量测用于获取图像层统计信息和局部像素匹配信息;F 量测用来计算召回率;M 用于计算预测显著性图与 GT 之间的平均绝对值误差。其中 S_α, E_ξ 和 F_β 的值越大越好, M 的值越小越好。

3.2 实现细节

本文方法采用 Pytorch 框架,并在单个 GeForce RTX 3080 GPU 上进行了实验。主干网络采用 PVTv2 或 ResNet50,去掉最后一个分类层,在 ImageNet 对预训练模型进行初始化。在训练和测试阶段,将输入图像的分辨率调整为 352×352。为了防止出现过拟合,本文采用随机翻转、旋转和裁剪边界等变换对所有数据进行增强。采用 Adam 对模型进行优化,初始学习率定为 0.000 1,每 50 个 epoch 为学习率除以 10。在实验中,批处理大小设置为 6,共训练了 150 个 epoch。

3.3 对比实验

将本文提出的 BC²FNet 在 5 种标准数据集上与 8 种最先进的 RGB-D 显著性检测算法进行比较,包括 CoNet^[4]、HdFNet^[3]、DsA2F^[20]、BiANet^[21]、DFM^[22]、SsLSOD^[23]、C2DFNet^[24] 和 MvSalNet^[25]。为了公平比较,本文采用相应文献提供的显著性图和显著性计算评价结果。

1) 客观评价。表 1 显示了本文提出的 BC²FNet 与 8 种最先进的方法在 5 种数据集上的 4 种评估标准对比。可以观察到,基于 PVTv2 骨干网的方法在 NJUD、NLPR、STERE 和 SIP 数据集上的性能优于所有方法。在 LFSD 数据集上,本文方法的 E_ξ 和 M 值略低于 DSA2F 方法。对于 5 种标准数据集,本文方法在 S_α 和 F_β 2 种指标上的性能提高了 1%。对于基于 ResNet50 的方法,可以看到在 4 种评估指标下的性能大多处于前 3 个级别,与最优结果相差很小。

HdFNet 方法提出一种混合增强损失函数来优化显著目标的边界,本文方法与此方法相比在 NJUD、NLPR、LFSD、STERE 和 SIP 数据集上分别提升了 2.1%, 1.3%, 2.9%, 1.9% 和 1.5%。CoNet 方法首先从 RGB 模态的低层特征中获取边界信息,然后将边界信息融入 RGB 模态和深度模态预测显著图。本文方法与 CoNet 方法相比在 NJUD、NLPR、LFSD、STERE 和 SIP 数据集上分别提升了 3.3%, 3.1%,

2.4%,1.6%和 3.7%。与这 2 种典型方法的比较结果证实了本文方法能够有效地保留显著目标的区域和边界。总之,与 8 种方法的比较充分证实了本文提出的 BC²FNet 的优越性。

2) 主观评价。图 6 显示了本文提出的 BC²Net

与 6 种最先进方法在复杂场景图像下的视觉对比,可以看出,所提出方法预测的显著性图更接近于 GT。同时,本文的显著性图与其他方法相比包含更多边界清晰的显著性目标。证明了本文所提 BC²FNet 网络的有效性。

表 1 本文方法与 8 个最先进的方法使用 4 种评估指标进行定量比较

数据集		CoNet	HdFNet	DsA2F	BiANet	DFM	SsLSOD	C ² DFNet	MvSalNet	本文*	本文+
NJUD	$S_{\alpha} \uparrow$	0.894	0.908	0.904	0.922	0.907	0.909	0.908	0.912	0.921	0.936
	$F_{\beta} \uparrow$	0.895	0.910	0.907	0.924	0.913	0.910	0.909	0.914	0.919	0.936
	$E_{\xi} \uparrow$	0.934	0.944	0.938	0.952	0.946	0.942	0.942	0.949	0.950	0.964
	$M \downarrow$	0.046	0.038	0.040	0.034	0.043	0.038	0.039	0.036	0.035	0.027
NLPR	$S_{\alpha} \uparrow$	0.908	0.926	0.918	0.929	0.923	0.923	0.928	0.930	0.930	0.943
	$F_{\beta} \uparrow$	0.887	0.917	0.907	0.918	0.905	0.910	0.914	0.921	0.925	0.937
	$E_{\xi} \uparrow$	0.945	0.963	0.952	0.963	0.956	0.960	0.959	0.964	0.963	0.972
	$M \downarrow$	0.031	0.023	0.024	0.023	0.026	0.025	0.022	0.022	0.023	0.018
LFSD	$S_{\alpha} \uparrow$	0.862	0.854	0.883	0.814	0.865		0.864	0.858	0.880	0.892
	$F_{\beta} \uparrow$	0.859	0.862	0.889	0.787	0.864		0.867	0.861	0.883	0.894
	$E_{\xi} \uparrow$	0.907	0.896	0.924	0.847	0.903		0.903	0.904	0.910	0.922
	$M \downarrow$	0.071	0.077	0.055	0.095	0.072		0.066	0.072	0.065	0.057
STERE	$S_{\alpha} \uparrow$	0.905	0.900	0.897	0.908	0.908	0.905	0.902	0.913	0.912	0.927
	$F_{\beta} \uparrow$	0.900	0.910	0.899	0.904	0.902	0.899	0.896	0.909	0.911	0.924
	$E_{\xi} \uparrow$	0.945	0.931	0.940	0.942	0.946	0.941	0.942	0.947	0.946	0.957
	$M \downarrow$	0.036	0.041	0.038	0.039	0.039	0.038	0.037	0.035	0.036	0.031
SIP	$S_{\alpha} \uparrow$	0.858	0.886	0.862	0.889	0.885	0.888	0.872		0.890	0.904
	$F_{\beta} \uparrow$	0.867	0.894	0.875	0.893	0.889	0.893	0.877		0.901	0.918
	$E_{\xi} \uparrow$	0.913	0.930	0.912	0.928	0.926	0.928	0.916		0.929	0.942
	$M \downarrow$	0.063	0.047	0.057	0.047	0.049	0.046	0.053		0.048	0.041

注:“ $\uparrow$ ”表示越大越好,“ $\downarrow$ ”表示越小越好;前 3 种优秀结果分别用红色、蓝色和绿色标记;上标“*”表示使用 Res-Net50 作为骨干网络,上标“+”表示使用 PVTv2 作为骨干网络。

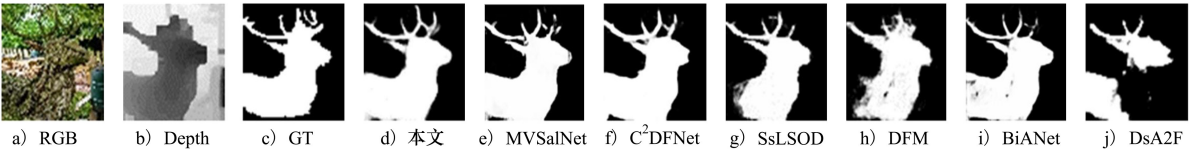


图 6 将本文提出的方法与 6 种最先进的方法进行视觉比较

3.4 消融实验

在本节中,为了证明所提出的 BGM、BFSM 和 BCFM 对整个网络的贡献,分别在 3 个模块上进行消融实验。

1) 边界生成模块的有效性。为了验证 BGM 的有效性,本文进行了不同策略的实验。①增加 B_r (only B_r):仅使用 RGB 模态的前 3 层边界特征 B_r 进行后续跨模态融合和跨层融合。②增加 B_d (only B_d):仅将前 3 层深度模态特征得到的边界特征 B_d

加入到跨模态融合和跨层融合中来预测显著性图。③w/o B_o :在跨模态和跨层融合过程中,不使用提取的 2 个边界特征来预测显著性图。根据表 2,可以观察到,无论是使用边界信息还是仅使用其中一种都会降低所提 BC²FNet 的性能。用上述 3 种方案得出的显著图如图 7e)~7g)所示,与图 7d)相比,花朵和玩偶的边界比较模糊,证实了本文提出的边界生成模块的有效性。

表 2 BGM、BFSM 和 BCFM 的消融实验

策略	NJUD				LFSD				SIP			
	$S_{\alpha} \uparrow$	$F_{\beta} \uparrow$	$E_{\xi} \uparrow$	$M \downarrow$	$S_{\alpha} \uparrow$	$F_{\beta} \uparrow$	$E_{\xi} \uparrow$	$M \downarrow$	$S_{\alpha} \uparrow$	$F_{\beta} \uparrow$	$E_{\xi} \uparrow$	$M \downarrow$
only B_r	0.932	0.934	0.960	0.030	0.882	0.882	0.911	0.065	0.897	0.912	0.934	0.046
only B_d	0.933	0.936	0.964	0.028	0.885	0.884	0.916	0.059	0.902	0.915	0.939	0.044
w/o B_o	0.929	0.931	0.959	0.031	0.879	0.879	0.911	0.067	0.897	0.906	0.935	0.046
w/o BFSM_B	0.932	0.935	0.960	0.030	0.891	0.892	0.924	0.058	0.901	0.916	0.938	0.044
w/o FSM	0.931	0.933	0.959	0.031	0.884	0.884	0.915	0.064	0.898	0.911	0.936	0.045
w/o BCFM_B	0.930	0.932	0.959	0.030	0.883	0.891	0.918	0.060	0.896	0.905	0.934	0.044
w/o BCFM	0.932	0.935	0.960	0.030	0.885	0.888	0.919	0.059	0.904	0.915	0.940	0.042
本文 ⁺	0.936	0.936	0.964	0.027	0.892	0.894	0.922	0.057	0.904	0.918	0.942	0.041

注:最优实验结果用粗体标出,上标“+”表示使用 PVTv2 作为骨干网络。

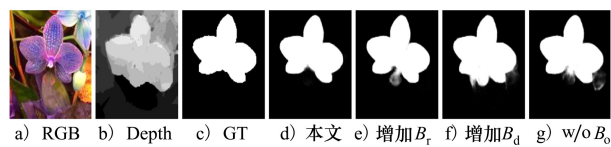


图 7 边界生成模型视觉比较消融研究

2) 边界驱动特征选择模块的有效性。在跨模态融合过程中,BFSM 用于保留目标边界信息,并选择重要特征。因此,本文对该模块的 2 个部分进行了 2 次实验,以验证其有效性。①w/o BFSM_B:去掉边界驱动特征选择模块中引入 2 种类型的边界信息。②w/o FSM:在边界驱动特征选择模块中,去掉负责特征选择功能的通道注意模块和空间注意模块。根据表 2 可以发现去除任何部分都会降低检测性能。从图 8 所示的视觉对比可以看出,去除边界信息之后得出的图 8e) 中花朵和玩偶的边界信息并不清楚,去除 2 个注意模块后得出的图 8f) 中花朵和玩偶都出现重要目标区域特征的缺失。

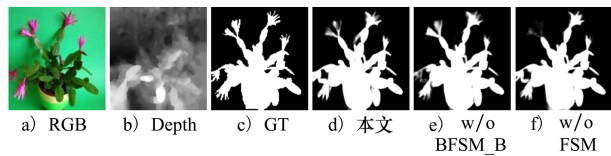


图 8 边界驱动特征选择模型视觉对比消融研究

3) 边界驱动跨层融合模块的有效性。在本文提出的 BC²FNet 中,BCFM 主要用于融合相邻层特征,并将该模块嵌入到自顶而下的信息流中预测显著性图。因此,本文通过以下实验来验证其有效性。①w/o BCFM_B:去掉边界驱动跨层融合模块中添加

的 2 种边界信息。②w/o BCFM:不使用 BCFM 进行自顶向下的预测显著性图,经过跨模态融合后的增强特征直接上采样成 GT 大小,并通过拼接操作对增强特征整合预测最终的显著性图。从表 2 和图 9 可以看出,去除任何部分都会降低性能,并且会出现对象不完整和边界不清晰问题。

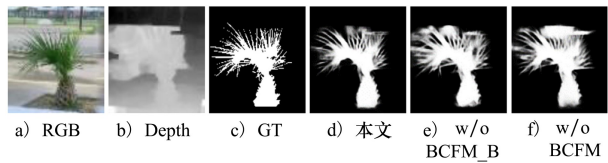


图 9 边界驱动跨层融合模型视觉对比消融研究

4 结 论

针对现有的许多方法预测显著性图边界不清晰问题,提出了一种边界驱动跨模态跨层融合网络的 RGB-D SOD 方法。首先,提出了一种边界生成模型,分别从 RGB 模态和深度模态的低层特征中提取 2 种类型的边界信息。其次,设计了边界驱动特征选择模块,将 2 种类型边界信息和 2 种类型注意机制引入跨模态融合中,使融合后的特征既能保留边界信息,又能保留重要特征。最后,提出了一种边界驱动跨层融合模块,将 2 种边界信息引入相邻层的融合过程。该模块嵌入自顶向下的信息流中,使预测的显著性图包含清晰边界的正确目标。在 5 种标准数据集上的主观和客观实验分析中证明了本文所提 BC²FNet 的优越性。

参考文献:

- [1] NIU Y, LONG G, LIU W, et al. Boundary-aware RGB-D salient object detection with cross-modal feature sampling[J]. IEEE Trans on Image Process, 2020, 29(5): 9496-9507
- [2] QIN X, ZHANG Z, HUANG C, et al. BASNet: boundary-aware salient object detection[C]//Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, 2019: 7471-7481
- [3] PANG Y, ZHANG L, ZHAO X, et al. Hierarchical dynamic filtering network for RGB-D salient object detection[C]//Proceedings of the Computer Vision, 2020: 235-252
- [4] JI W, LI J, ZHANG M, et al. Accurate RGB-D salient object detection via collaborative learning[C]//Proceedings of the Computer Vision, 2020: 52-69
- [5] 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 Process Letters, 2016, 23(8): 819-823
- [6] REN J, GONG X J, YU L, et al. Exploiting global priors for RGB-D saliency detection[C]//Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops, Boston, 2015: 25-32
- [7] LIANG F, DUAN L, MA W, et al. Stereoscopic saliency model using contrast and depth-guided-background prior[J]. Neurocomputing, 2018, 275(15): 2227-2238
- [8] CHEN S, FU Y. Progressively guided alternate refinement network for RGB-D salient object detection[C]//Proceedings of the Computer Vision, 2020: 520-538
- [9] LI G, LIU Z, YE L, et al. Cross-modal weighting network for RGB-D salient object detection[C]//Proceedings of the Computer Vision, 2020: 665-681
- [10] ZHANG C, CONG R, LIN Q, et al. Cross-modality discrepant interaction network for RGB-D salient object detection[C]//Proceedings of the 29th ACM International Conference on Multimedia, 2021: 2094-2102
- [11] ZHANG W, JIANG Y, FU K, et al. BTS-net: bi-directional transfer-and-selection network for RGB-D salient object detection [C]//Proceedings of the 2021 IEEE International Conference on Multimedia and Expo, Shenzhen, 2021: 1-6
- [12] LI G, LIU Z, CHEN M, et al. Hierarchical alternate interaction network for RGB-D salient object detection[J]. IEEE Trans on Image Process, 2021, 30(6): 3528-3542
- [13] WANG W, ZHAO S, SHEN J, et al. Salient object detection with pyramid attention and salient edges[C]//Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, 2019: 1448-1457
- [14] LIU J J, HOU Q, CHENG M M, et al. A simple pooling-based design for real-time salient object detection[C]//Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, 2019: 3912-3921
- [15] SONG D, DONG Y, LI X. Hierarchical edge refinement network for saliency detection[J]. IEEE Trans on Image Process, 2021, 30: 7567-7577
- [16] YANG Z, SOLTANIAN-ZADEH S, FARSIU S. BiconNet: an edge-preserved connectivity-based approach for salient object detection[J]. Pattern Recognition, 2022, 12(6): 121-131
- [17] ZHAI Y, FAN D P, YANG J, et al. Bifurcated backbone strategy for RGB-D salient object detection[J]. IEEE Trans on Image Process, 2021, 30(9): 8727-8742
- [18] LI C, CONG R, PIAO Y, et al. RGB-D salient object detection with cross-modality modulation and selection[C]//Proceedings of the Computer Vision, 2020: 225-241
- [19] JIN W D, XU J, HAN Q, et al. CDNet: complementary depth network for RGB-D salient object detection[J]. IEEE Trans on Image Process, 2021, 30(9): 3376-3390
- [20] SUN P, ZHANG W, WANG H, et al. Deep RGB-D saliency detection with depth-sensitive attention and automatic multi-modal fusion[C]//Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, 2021: 1407-1417
- [21] ZHANG Z, LIN Z, XU J, et al. Bilateral attention network for RGB-D salient object detection[J]. IEEE Trans on Image Process, 2021, 30(5): 1949-1961
- [22] ZHANG W, JI G P, WANG Z, et al. Depth quality-inspired feature manipulation for efficient RGB-D salient object detection

[C]//Proceedings of the 29th ACM International Conference on Multimedia, 2021: 731-740

[23] ZHAO X, PANG Y, ZHANG L, et al. Self-supervised representation learning for RGB-D salient object detection[C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2022

[24] ZHANG M, YAO S Y, HU B Q, et al. C2DFNet: criss-cross dynamic filter network for RGB-D salient object detection[J]. IEEE Trans on Multimedia, 2022, 25(3): 1-13

[25] ZHOU J Y, WANG L J, LU H C, et al. MVSaNet: multi-view augmentation for RGB-D salient object detection[C]//Proceedings of the Computer Vision, 2022: 270-287

RGB-D salient object detection based on BC² FNet network

WANG Feng^{1,2}, CHENG Yongmei³

(1.School of Physics and Electrical Engineering, Weinan Normal University, Weinan 714099, China;
2.Engineering Research Center of X-ray Imaging and Detection, University of Shaanxi Province, Weinan 714099, China;
3.School of Automation, Northwestern Polytechnical University, Xi'an 710072, China)

Abstract: In the face of complex scene images, the introduction of depth information can greatly improve the performance of salient object detection. However, up-sampling and down-sampling operations in neural networks maybe blur the boundaries of objects in the saliency map, thereby reducing the performance of salient object detection. Aiming at this problem, a boundary-driven cross-modal and cross-layer fusion network (BC²FNet) for RGB-D salient object detection is proposed in this paper, which preserves the boundary of the object by adding the guidance of boundary information to the cross-modal and cross-layer fusion, respectively. Firstly, a boundary generation module is designed to extract two kinds of boundary information from low-level features of RGB and depth modalities, respectively. Secondly, a boundary-driven feature selection module is designed, which is dedicated to simultaneously focusing on important feature information and preserving boundary details in the process of RGB and depth modality fusion. Finally, a boundary-driven cross-layer fusion module is proposed which simultaneously adds two kinds of boundary information in the process of up-sampling fusion on adjacent layers. By embedding this module into the top-down information fusion flow, the predicted saliency map can contain accurate objects and sharp boundaries. Simulation results on five standard RGB-D data sets show that the proposed model can achieve better performance.

Keywords: salient object detection; boundary-driven; cross-modal fusion; cross-layer fusion

引用格式:王峰, 程咏梅. 基于 BC²FNet 网络的 RGB-D 显著性目标检测[J]. 西北工业大学学报, 2024, 42(6): 1135-1143
WANG Feng, CHENG Yongmei. RGB-D salient object detection based on BC²FNet network[J]. Journal of Northwestern Polytechnical University, 2024, 42(6): 1135-1143 (in Chinese)