

基于细粒度深度质量感知的 RGB-D 显著性物体检测

张明月, 崔爽铤, 赵一汎, 卢搏, 王子铭, 陈程立诏

(中国石油大学(华东)青岛软件学院、计算机科学与技术学院, 山东 青岛 266580)

摘要: 基于融合的 RGB-D 显著性物体检测(RGB-D SOD)方法通常采用双流结构在 RGB 和深度图之间进行权衡,目的是使得 SOD 性能优于单独使用 RGB 图时的性能。然而,深度图因场景、拍摄设备等因素的影响而存在质量差异,低质量的深度图无法有效辅助 RGB 图像进行显著性检测,甚至可能产生负面影响。现有方法很少考虑深度通道的质量差异,导致低质量深度区域的干扰信息被不合理地融合,从而影响结果的准确性。为解决以上问题,提出一种基于细粒度深度质量感知的 RGB-D SOD 方法,该方法主要分为 2 个阶段,即细粒度深度质量感知和双线性路由融合。在第一阶段,通过细粒度地比较 RGB 和 D 分支的显著性结果与伪标签(PGT)的相似性差异,对深度通道进行质量感知;在第二阶段,基于不同深度区域的质量感知结果,设计一种新的双线性路由融合策略,路由方向由深度通道质量感知结果驱动,以实现比现有融合方案更有效的多模态互补结果。在 NJU2K、NLPR、SIP 和 STERE 数据集上,将所提方法与 9 种 SOTA 方法进行比较,结果表明,所提方法在评价指标上都优于当前最先进的模型。相比于性能次优的模型,该方法的 S-measure 指标平均提升 0.6%,F-measure 指标平均提升 0.7%,平均绝对误差(MAE)平均降低 0.4%。所提方法显著减少了低质量、无贡献甚至负影响的深度区域对结果的影响,实现了更优的多模态融合,提升了显著性检测的效果。

关键词: 显著性物体检测;深度图质量;细粒度质量感知;双线性路由融合;多模态互补

源代码链接: <https://github.com/zhangmoon1/Depth-Quality-Perception>

中图分类号: TP391

文献标志码: A

DOI: 10.19678/j.issn.1000-3428.0070592

RGB-D Salient Object Detection Based on Fine-Grained Depth Quality Perception

ZHANG Mingyue, CUI Shuangxin, ZHAO Yifan, LU Bo, WANG Ziming, CHEN Chenglizhao

(Qingdao Institute of Software, College of Computer Science and Technology,

China University of Petroleum (East China), Qingdao 266580, Shandong, China)

【Abstract】 Fusion-based RGB-D Salient Object Detection (RGB-D SOD) methods employ a dual-stream structure to balance RGB and depth maps, aiming to achieve superior performance in SOD compared with using RGB maps alone. However, owing to factors such as the scene and shooting equipment, depth maps vary in quality. Low-quality depth maps cannot effectively assist RGB images in salient detection and may even have a negative impact. Existing methods rarely consider the quality differences in depth channels, leading to the unreasonable fusion of interference information from low-quality depth regions, which affects the accuracy of the results. To address these issues, this paper proposes a fine-grained depth-quality-aware RGB-D SOD method, which is divided into two stages: fine-grained depth quality awareness and bilinear routing fusion. In the first stage, the depth channel quality is perceived by comparing the similarity differences between the salient results of the RGB and D branches and Pseudo Ground Truth (PGT) in a fine-grained manner. In the second stage, a new bilinear routing fusion strategy is designed based on the quality perception results for different depth regions. The routing direction is driven by the results of depth channel quality perception, achieving more effective multimodal complementary results than existing fusion schemes. The proposed method is compared to nine State-Of-The-Art (SOTA) methods on the NJU2K, NLPR, SIP, and STERE datasets. The results show that the proposed method outperforms current leading models in all evaluation metrics. Compared with the suboptimal models, the S-measure and F-measure metrics of the proposed method are improved by an average of 0.6% and 0.7%, respectively, and the Mean Absolute Error (MAE) is reduced by an average of 0.4%. The proposed method significantly reduces the impact of low-quality, non-contributing, or negatively impacting depth regions on the results, achieving better multimodal fusion and enhancing the effectiveness of salient detection.

【Key words】 Salient Object Detection (SOD); depth map quality; fine-grained quality perception; bilinear routing fusion; multi-modal complementarity

基金项目: 国家自然科学基金(62172246);山东省自然科学基金优秀青年科学基金项目(ZR2024YQ071);山东省高校青年创新科技支持计划(2021KJ062)。

作者简介: 张明月,女,硕士研究生,主研方向为显著性物体检测、计算机视觉、深度学习;崔爽铤、赵一汎、卢搏、王子铭,硕士研究生;陈程立诏(通信作者),教授。

收稿日期: 2024-11-07

修回日期: 2024-12-30

E-mail: cclz123@163.com

0 引言

RGB-D 显著性物体检测 (RGB-D SOD)^[1] 旨在利用 RGB 图像及其对应的深度图, 模拟人类视觉注意机制, 以快速定位和分割图像中的显著物体。该技术广泛应用于目标识别^[2]、图像检索^[3]、视频分析^[4]、视觉跟踪^[5]、图像质量评估^[6] 等下游任务。RGB-D 显著性检测方法通常采用双流结构: 一方面从 RGB 图像中提取颜色、纹理和语义信息, 以识别和区分不同的物体; 另一方面将深度信息作为额外的图像通道, 提供丰富的空间和距离信息, 从而使算法能够更准确地理解图像中物体的空间关系。通过结合 RGB 和深度信息, RGB-D 的检测性能显著优于仅使用 RGB 图像的方案。

近年来, 受益于深度卷积神经网络 (CNN) 强大的表示学习能力, RGB-D 显著性检测取得了显著进展。现有方法^[7-9] 通常采用预训练的深度模型, 学习 RGB 和深度通道的层次化特征表示。随后, 这些显著特征输入到融合层, 以实现 RGB 和深度通道的有效结合, 优化两者的互补作用。然而, 实现 RGB 与深度通道之间的最佳融合仍是当前的核心挑战, 特别是在处理低质量深度图像时, 大多数融合方案的表现仍不尽如人意, 导致性能下降。

深度图像的质量受到深度传感器/相机和场景的影响而存在差异^[10]。高质量的深度图像通常具有边界清晰、信息完整且一致性强的特点, 能够提供可靠的深度信息, 能够生成性能较高的深度显著图。然而, 低质量的深度图像常伴有模糊和噪点问题, 导致显著性目标难以被清晰地识别和提取, 从而降低深度显著图的准确性。此外, 同一张深度图像中不同区域的质量也可能存在显著差异。低质量的深度区域不仅无法有效促进 RGB 图像的显著性检测, 甚至可能引入噪声, 降低检测准确性, 增加局部信息干扰, 并加剧融合过程的难度。为减少低质量、无贡献甚至具有负面影响的深度区域对显著性结果的干扰, 本文提出了一种细粒度深度质量感知方法。

目前, 主流的融合方式可以分为早期融合^[11-12]、中期融合^[13] 和后期融合^[14-15]。尽管这些方法在一定程度上提升了性能, 但它们在融合过程中通常未考虑通道质量的差异。当面对不同质量的深度图像时, 采用单一的融合策略难以实现最佳效果, 往往只能达到次优性能。为此, 本文基于深度质量感知结果, 考虑 RGB 和深度信息之间的区别, 提出一种双线性路由融合策略, 以实现最优的多模态互补状态。

综上, 为解决低质量深度图对 RGB-D 显著性检

测的影响, 并实现多模态的有效融合, 本文提出一种基于细粒度深度质量感知的 RGB-D 显著性物体检测方法。该方法分为 2 个阶段: 在第一阶段, 通过细粒度比较 RGB 和深度分支的显著性结果与伪标签 (PGT) 的相似性差异, 对深度通道进行质量评估; 在第二阶段, 根据深度区域的质量感知结果, 设计一种新的双线性路由融合策略, 其中路由方向由深度区域质量感知结果驱动, 从而实现有效的多模态互补。值得注意的是, 本文方法具有较强的普适性和鲁棒性, 可以与任意 SOTA 显著性检测模型结合使用。

1 相关工作

RGB-D 显著性物体检测作为一种多模态学习任务, 其核心目标是高效融合来自 RGB 和深度图的显著性线索。早期的传统方法主要依赖于从 RGB 和深度图中提取手工设计的特征。文献[16] 通过从深度通道中提取多种中层手工特征来增强显著性融合的鲁棒性。文献[17] 则以超像素为单位, 计算 RGB 和深度图的局部和全局对比度以及边界对比度。这些方法尽管能够有效结合颜色、纹理和深度信息, 但是由于其有限的表示能力, 制作的特征往往难以提供足够信息来支持准确的显著性物体检测, 特别是在具有多个物体的复杂场景中。

随着深度学习的迅速发展, 研究重心逐渐转向自动特征学习方法, 以克服手工特征设计的局限性^[18]。现有的 RGB-D 显著性物体检测方法大多采用双流结构^[19-21], 通过将深度图作为额外输入来丰富 RGB 流的上下文和空间信息, 从而增强模型对物体位置和关系的理解。然而, 由于深度图质量的差异, 其在补充 RGB 数据时有效性往往不稳定。为此, 一些研究通过优化深度通道以提升检测性能。例如, 文献[10] 基于高质量深度区域的特性, 设计了多层次深度特征以评估深度质量。但这类方法往往手工成本高、耗时且难以应对复杂场景。随后, 基于深度学习的优化方法逐渐兴起, 研究者开始使用神经网络优化深度质量。文献[22] 设计了一个深度增强模块, 旨在改善深度信息并提高 RGB 和深度特征的兼容性。文献[14] 提出了一种基于 RGB 的对比度损失网络, 用于增强深度特征质量。文献[23] 则引入了一种名为干扰度的可解释机制, 用以评估深度图质量并重新加权单模态特征的贡献。

尽管这些方法在一定程度上能有效增强深度信息, 但在处理低质量深度图像时仍表现不佳。因此, 本文提出一种新的深度质量感知方法, 通过细粒度地感知深度图区域的质量, 减少低质量深度信息对

检测结果的误导,实现多模态之间更高效的融合,从而提升显著性检测的准确性。

2 本文方法设计与实现

2.1 整体结构

本文方法主要分为 2 个阶段,即细粒度深度质量感知和双线性路由融合,重点解决传统 RGB-D 显著性检测中对深度质量差异缺乏感知的问题。在细粒度深度质量感知阶段,本文通过比较 RGB 和深度显著性结果与伪标签的区域级相似性差异,对深度图的质量进行精细化评估,以有效识别高质量与低质量区域,从而减少低质量深度区域对后续处理的干扰。在双线性路由融合阶段,设计一种基于深度质量分数的动态融合策略,通过质量驱动的选择机制,自适应结合 RGB-D 和 RGB 的显著性区域,

有效优化多模态互补效果。本文方法显著提升了低质量深度信息处理的鲁棒性,并在多模态显著性检测任务中实现了性能的全面提升。

2.2 细粒度深度质量感知

深度信息作为 RGB 图颜色和纹理的补充,为模型检测提供了边缘和空间信息,在 RGB-D 显著性物体检测中发挥着重要作用。然而,受到深度设备和周围环境的影响,同一张深度图像的不同区域质量也存在差异。高质量的区域能够轻松地将显著物体与背景分离开来,例如图 1 中深度图的飞机机翼区域,其轮廓清晰,能够获得良好的区域显著性结果。然而,如图 1 深度图的飞机头部,低质量深度区域则呈现出与周围环境模糊或粘连的现象,这种情况使得难以准确区分前景物体,导致定位错误、识别不完整或未能识别出显著物体的问题。

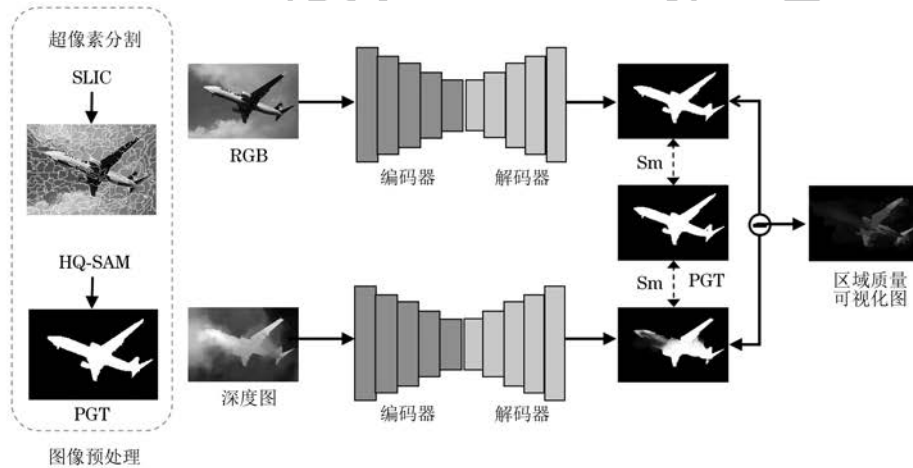


图 1 细粒度质量感知方法框架

Fig. 1 Framework of fine-grained quality perception method

为了解决图像中不同区域的质量差异问题,本文提出一种基于细粒度深度质量感知的检测方法。其中,细粒度深度质量感知是指以较小的空间单元为单位,对深度图像的质量进行精确的区域级评估。这种方法通过对每个超像素区域进行分析,能够有效区分不同区域的深度质量差异。如图 1 所示,细粒度深度质量感知包含 RGB 和深度 2 个分支:首先进行图像预处理,利用 HQ-SAM 高精度分割模型^[24]生成显著物体的 PGT,并采用 SLIC^[25]算法对图像进行超像素分割;接着,使用 RGB 显著性模型(可以是任意 SOTA 模型),分别将 RGB 图和深度图作为输入,经编码器和解码器,生成对应的显著性图,与 PGT 计算相似性差异,对深度区域进行质量感知。

为实现细粒度感知,采用 SLIC 算法将图像划分为多个紧凑且相似的超像素区域。通过区域内像素的颜色和空间信息计算来确定超像素中心,并在

迭代过程中不断优化。超像素中心作为深度质量感知的基准点,用于衡量每个像素与其所属超像素区域的特征相似性,从而实现高效、准确的区域划分与深度质量分析。最终,计算每个像素与其周围超像素中心之间的欧氏距离,将像素分配给最近的超像素中心。重复以上操作,直到超像素中心的位移足够小,具体公式如下:

$$\begin{cases} C_i = (l_i, a_i, b_i, x_i, y_i) \\ D_i = \sqrt{\left(\frac{d_c}{N_c}\right)^2 + \left(\frac{d_s}{N_s}\right)^2} \\ C' = \frac{1}{N_i} \sum_{p \in S_i} P \end{cases} \quad (1)$$

式中: l_i, a_i, b_i 表示超像素中心的颜色信息; x_i 和 y_i 表示超像素中心的位置; d_c 和 d_s 分别表示颜色距离和空间距离; N_c 和 N_s 是类内的最大颜色距离和最大空间距离; C' 是超像素的更新中心; N_i 是分

配给超像素中心 C_i 的像素数; S_i 是与 C_i 相关的像素集; p 表示 S_i 中的一个像素。SLIC 提供了高质量的分割结果,并具有较强的计算速度,同时在一定程度上保留了图像的边缘信息。通过上述公式,将图像划分为多个区域,记为 $sp_j, j \in [1, M]$,并评估深度图像各区域的质量。本文利用 RGB 显著性检测工具 X ,分别生成 RGB 显著性图 $I_{RGB, sal}$ 和深度显著性图 $I_{D, sal}$ 。其中, X 可以是任意 SOTA 模型,如 EDN^[26] 和 BBRF^[27]。 $I_{RGB, sal}$ 和 $I_{D, sal}$ 计算如下:

$$\begin{cases} I_{RGB, sal} = X(RGB) \\ I_{D, sal} = X(Depth) \end{cases} \quad (2)$$

为避免 RGB 和深度图同时被检测错误的情况,本文没有直接计算 $I_{RGB, sal}$ 和 $I_{D, sal}$ 的相似性,而是

通过计算与 PGT 的相似性的差异,以提高评估的准确性。PGT 是通过随机选取多个显著图的像素值 $I(x, y)$,在 RGB 图片上标记前景(Foreground)或背景(Background),利用高精度分割模型 HQ-SAM 分割前景物体而生成的,公式如下:

$$\begin{cases} I(x, y) \text{ 标记为 Foreground, } I(x, y) \geq 254 \\ I(x, y) \text{ 标记为 Background, } I(x, y) \leq 0 \end{cases} \quad (3)$$

式中:像素值 $I(x, y) \geq 254$ 时标记为前景;像素值 $I(x, y) \leq 0$ 时标记为背景; $I(x, y)$ 取其他值时则丢弃并重新选取。

图 2 中展示了高质量和低质量深度显著性图及 PGT 的可视化结果。

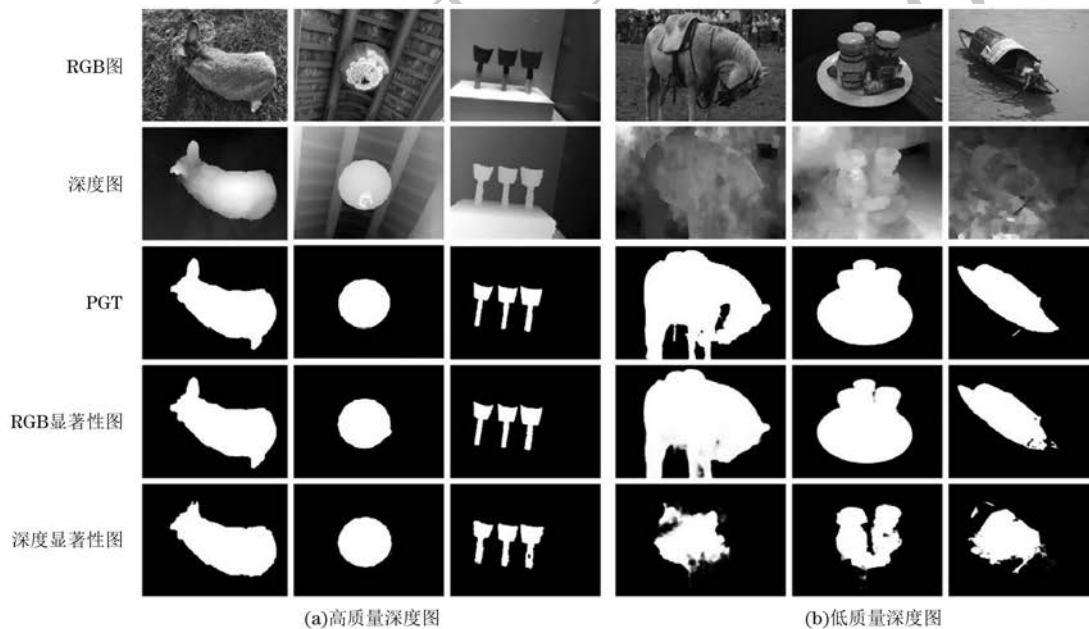


图 2 高质量和低质量深度显著性图及 HQ-SAM 模型分割的 PGT

Fig. 2 High-quality and low-quality deep saliency maps and PGT segmented by the HQ-SAM model

可以看出,无论深度图质量高低,高精度分割模型 HQ-SAM 均能生成高质量的 PGT。此外,低质量深度区域的显著性结果与 PGT 差异较大,而高质量区域的显著性结果与 PGT 更加一致,为深度质量感知提供了可靠的参考依据。为了评估显著图与 PGT 之间的相似性,使用 S-measure (S_m, S_{sm})^[28] 指标,它通过结合全局和局部信息,比较显著性结果与标签的结构相似性和区域相似性,以评估定位显著目标的准确性以及图像结构的完整性和准确性,具体公式表达如下:

$$S_{sm} = \alpha \cdot S_{so} + (1 - \alpha) \cdot S_{sr} \quad (4)$$

式中: $\alpha = 0.5$ 用于平衡对象感知(S_o, S_{so})和区域感知(S_r, S_{sr})之间的结构相似性。 S_m 值越接近 1,表示显著性图的检测效果越好。通过计算 RGB 显著图和深度显著图与伪标签之间的 S_m 差异,得到深

度通道不同区域的质量分数(S_d, S_{sd}),用于评估深度信息的贡献大小,具体公式如下:

$$S_{sd} = \{S_{sm}(I_{RGB, sal}, P_{PGT}) - S_{sm}(I_{D, sal}, P_{PGT})\} \quad (5)$$

通过上述公式,对深度通道进行细粒度质量感知, S_d 为 RGB 显著性图和深度显著性图与 PGT 的相似性差异。较大的 S_d 值,说明深度区域与对应的 RGB 显著区域的相似性差异较大,深度质量较低,可能对显著性检测产生干扰;较小的 S_d 值,则表明深度信息在分离前景和背景时具有更大的贡献。因此, S_{sd} 可以动态反映深度信息对显著性检测的贡献大小,该方法无须额外的学习过程即可完成有效评估,为后续融合提供了指导分数。

2.3 双线性路由融合

通过细粒度深度质量感知,得到深度通道各区

域的质量分数。在这些质量评分的指导下,引入一种基于质量感知的融合方案,采用双线性路由策略来实现通道融合。该策略的目的是尽量减少来自低贡献或错误的深度信息的干扰,从而优化模态的互补性。如图 3 所示,双线性路由融合包括 2 个分支:一个分支是利用 SOTA RGB 显著物体模型 X 生成 RGB 显著性结果;另一个是通过 SOTA RGB-D 显著物体模型 Y 生成 RGB-D 显著性结果。在深度通道质量分数的指导下,动态地结合 RGB-D 和 RGB 的显著性区域。模型 X 和 Y 可以是任意的 SOTA 显著性检测模型,如 RGB 显著性检测模型 EDN^[26] 和 BBRF^[27]、RGB-D 显著性检测模型 CAVER^[29] 和 PICRNet^[30]。

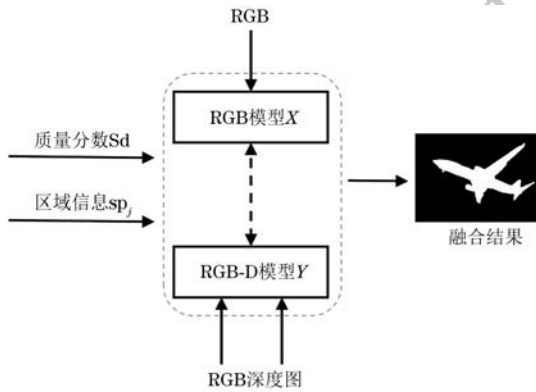


图 3 双线性路由融合策略示意图

Fig.3 Schematic diagram of bilinear routing fusion strategy

本文双线性路由融合策略根据深度通道的质量感知结果,直接选择 RGB 或 RGB-D 显著性结果,以减少低质量深度信息的影响。具体来说,当深度区域表现出低质量时,RGB-D 融合会引入噪声等干扰信息,不如直接丢弃深度信息,选择 RGB 显著性结果(RGB, sal);相反,当深度通道表现出高质量时,选择 RGB-D 显著性结果(RGBD, sal),因为高质量的深度信息为 RGB 样本提供了有价值的空间信息,可以改善 RGB 显著性图。通过利用深度通道质量分数来指导双线性路由融合,生成最终显著性结果(F, sal),公式如下:

$$I_{F, sal} = \omega \cdot I_{RGB, sal} + (1 - \omega) \cdot I_{RGBD, sal} \quad (6)$$

式中:参数 $\omega \in \{0, 1\}$ 由式(5)的深度区域质量分数 S_d 决定,设置阈值以区分深度质量高低。经消融实验(详见下文表 5),将关键阈值设为 0.3,具体 ω 取值如下:

$$\omega = \begin{cases} 1, & S_d \leq 0.3 \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

本文方法通过评估深度通道的质量,并且基于深度区域质量分数实行双线性路由融合,无须进行

额外的学习过程。这一过程显著减少了深度通道中低质量或负影响区域对结果的影响,从而确保了 RGB 和深度通道之间更高效的融合。

3 实验结果及分析

3.1 数据集

本文在 NJU2K^[31]、NLPR^[32]、SIP^[4] 和 STERE^[33] 这 4 个公开基准数据集上评估方法性能:NJU2K 数据集^[31] 包含从互联网、3D 电影和立体照片中收集的多样化图像,深度信息质量差异较大,适用于测试方法在复杂场景中的适应性;NLPR 数据集^[33] 涵盖丰富的室内外场景,提供多种光照和纹理条件,可验证方法在不同光照下深度区域质量感知的鲁棒性;SIP 数据集^[4] 聚焦于高分辨率图像和不同人体姿势的现实场景,评估模型处理复杂细节和动态特征的能力;STERE 数据集^[33] 通过双目立体图像展现了高深度精度场景,适用于评估方法对深度信息精确性的依赖。通过对多样化数据集的测试,有效验证本文方法应对深度质量差异的有效性和适应性。

为了与其他 RGB-D SOD 方法进行公平比较,本文遵循与 CAVER^[29]、PICRNet^[30] 等 SOTA 方法相同的训练和测试数据分割方案。训练集包含 2 185 张图像,其中包括来自 NJU2K 的 1 485 对图像和来自 NLPR 训练集的 700 对图像。NJU2K 和 NLPR 中其余的图像以及 SIP 和 STERE 的全部图像用于测试。

3.2 评价指标

为充分评估不同方法的性能差异,使用 Sm ^[28]、 $F\text{-measure}(F_\beta)$ ^[34] 和平均绝对误差(MAE)^[35] 3 个常用的度量标准对模型进行定量评价。 Sm ^[28] 关注显著图和 GT 之间的区域感知和物体感知的结构相似性,这有助于综合考虑显著性目标的定位、形状以及显著性物体的边界等因素。 F_β 是一种在区域级别上平衡精确率和召回率的度量指标。MAE 用于衡量预测显著性映射和真实映射之间每个像素值的平均绝对误差。其中, Sm 和 F_β 越高越好,MAE 则越低越好。

3.3 实验设置

实验和测试都在一台配备 NVIDIA GeForce RTX 3090 GPU(Windows 10 系统,16 GB 内存,24 GB 显存)的工作站上进行。本文方法用 Python 3.8 和 PyTorch 1.9 实现。显著性检测模型的参数在 ImageNet 上进行预训练,而其他参数和训练策略采用与 PICRNet^[30] 相同的设置。

3.4 实验结果分析

为验证本文方法的有效性,选择近年来 9 种具有代表性的 SOTA RGB-D 显著性物体检测方法作为对比,包括 DCF^[19]、SPNet^[20]、RD3D^[21]、CCFENet^[36]、MobileSal^[37]、CIR-Net^[38]、LSNet^[20]、CAVER^[29]和 PICRNet^[30],这些方法覆盖了多种模型架构和融合策略,能全面反映当前领域的技术水平。为了进行公平比较,使用原文提供的显著性结果,或严格按照 SOTA 方法的公开代码和参数设置,获取显著图并进行评估。其中,本文方法分别以 BBRF^[27]和 PICRNet^[30]作为基线 RGB SOD 和 RGB-D SOD 模型,按照第 2.2 节方法得出的细粒度深度质量分数,指导双线性融合生成显著性结果,通过比较来验证本文方法的普适性和鲁棒性。

各方法在 NJU2K、NLPR、SIP 和 STERE 这 4 个基准数据集上的定量评估结果如表 1 所示,最优结果加粗标注。可以看出,本文方法优于其他方法,相比于性能次优的模型,在 NJU2K、NLPR、SIP 和 STERE 数据集上,本文方法的 Sm 平均提升 0.6%, F_{β} 平均提升 0.7%,MAE 平均降低 0.4%,不同数据集上的性能提升存在差异。经分析,NJU2K 数据集深度质量差异显著,本文方法能够有效区分高、低质量区域,从而显著提升检测性能;SIP 数据集由于高分辨率图像质量较高,对深度感知的依赖性较弱,因此性能提升相对有限。本文方法取得较好性能的原因在于对深度图像进行了有效的细粒度质量感知,减少了低质量、无贡献甚至负影响的深度区域对显著性结果的影响。

表 1 不同方法的定量对比结果

Table 1 Quantitative comparison results of different methods

方法	NJU2K			NLPR			SIP			STERE		
	Sm $\uparrow$	F_{β} $\uparrow$	MAE $\downarrow$	Sm $\uparrow$	F_{β} $\uparrow$	MAE $\downarrow$	Sm $\uparrow$	F_{β} $\uparrow$	MAE $\downarrow$	Sm $\uparrow$	F_{β} $\uparrow$	MAE $\downarrow$
DCF	0.912	0.905	0.039	0.922	0.899	0.024	0.874	0.886	0.052	0.906	0.894	0.037
SPNet	0.920	0.899	0.038	0.927	0.904	0.021	0.894	0.893	0.043	0.907	0.886	0.037
RD3D	0.916	0.901	0.036	0.930	0.892	0.022	0.885	0.874	0.048	0.911	0.885	0.038
CCFENet	0.916	0.913	0.032	0.927	0.908	0.021	0.882	0.890	0.047	0.907	0.897	0.035
MobileSal	0.922	0.915	0.029	0.920	0.877	0.025	0.873	0.864	0.053	0.903	0.865	0.041
CIR-Net	0.925	0.907	0.035	0.933	0.885	0.023	0.888	0.875	0.052	0.896	0.902	0.039
LSNet	0.911	0.900	0.038	0.918	0.885	0.024	0.886	0.884	0.048	0.871	0.854	0.054
CAVER	0.923	0.920	0.028	0.930	0.903	0.023	0.895	0.897	0.043	0.914	0.892	0.034
PICRNet	0.926	0.919	0.029	0.934	0.910	0.021	0.899	0.900	0.040	0.916	0.904	0.033
本文方法	0.929	0.924	0.027	0.936	0.912	0.018	0.909	0.911	0.034	0.924	0.912	0.028

图 4 中展示了 SOTA 方法与本文方法的可视化显著性结果,在多样化场景下,本文方法均表现出

更优的显著性检测效果。对于第 1 行的单一物体场景,尽管大部分 SOTA 方法均能检测出显著物体,

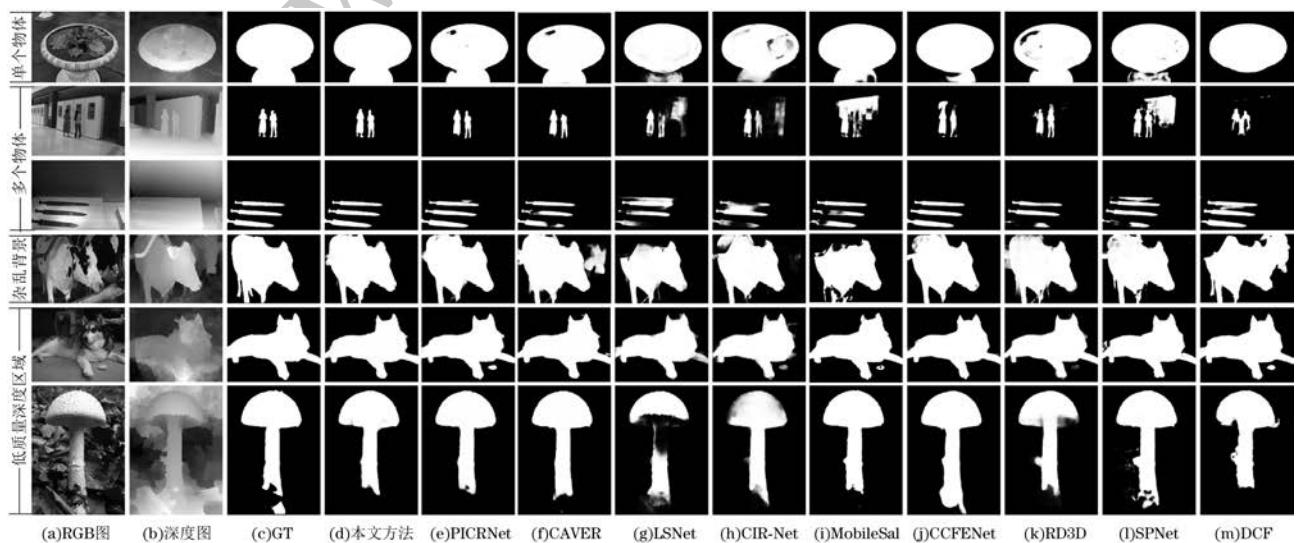


图 4 不同方法的显著性可视化结果

Fig.4 Visualization results of significance for different methods

但本文方法在边缘处理上表现更优,检测结果更为锐利且细节丰富,体现了对结构信息的精准捕捉;在第 2 行、第 3 行的多个物体场景中,部分方法难以检测较小或被遮挡的目标,而本文方法通过细粒度深度质量感知,有效融合高质量区域信息,实现了所有显著物体的完整检测;对于第 4 行的杂乱背景场景,部分方法因背景干扰而出现误检或漏检,而本文方法通过有效区分深度区域质量,减少背景干扰,准确定位显著物体,并保持了边缘清晰度;在第 5 行、第 6 行的低质量深度图场景中,部分方法表现出边缘模糊问题,而本文方法通过双线性融合机制,自适应抑制低质量深度信息的影响,确保了检测结果的完整性和精确性,进一步验证了本文方法的普适性和鲁棒性。

表 2 细粒度深度质量感知有效性的消融分析结果

Table 2 Ablation analysis results of the effectiveness of fine-grained depth quality perception

方法	NJU2K		NLPR		SIP		STERE	
	$F_{\beta} \uparrow$	MAE $\downarrow$	$F_{\beta} \uparrow$	MAE $\downarrow$	$F_{\beta} \uparrow$	MAE $\downarrow$	$F_{\beta} \uparrow$	MAE $\downarrow$
空间注意力	0.915	0.042	0.907	0.035	0.885	0.047	0.898	0.042
通道注意力	0.913	0.045	0.905	0.031	0.880	0.049	0.893	0.043
混合注意力	0.920	0.037	0.910	0.022	0.894	0.045	0.913	0.038
细粒度深度质量感知	0.924	0.027	0.912	0.018	0.911	0.034	0.912	0.028

3.5.2 双线性路由融合的有效性

为了将 RGB 和高质量的深度图像有效融合,本文提出双线性路由融合策略。为验证所提策略的有效性,将本文策略与 2 种直接融合方式(加法融合、乘法融合)进行比较,实验结果如表 3 所示。在 4 个基准数据集上,相比加法融合和乘法融合策略,本文

3.5 消融分析

3.5.1 细粒度深度质量感知的有效性

为了验证所提细粒度深度质量感知方法的有效性,将本文方法与 3 种深度图增强方法(空间注意力、通道注意力和混合注意力)进行比较,实验结果如表 2 所示。其中,在 4 个基准数据集上,相比空间注意力、通道注意力和混合注意力,本文方法的 F_{β} 指标分别平均提高 1.3%、1.7%和 0.6%,MAE 指标分别平均下降 1.4%、1.5%和 0.8%。实验结果表明,通过单一或者混合注意力机制对深度图质量进行增强,不如本文在第 2.2 节提出的方法,通过细粒度地挖掘高质量深度区域,可以有效减少低质量干扰信息的影响。

方法的 F_{β} 指标分别平均提高 2.7%和 2.5%,MAE 指标分别平均下降 1.9%和 1.7%,说明简单的线性加法、乘法在结合 2 种显著图时没有考虑相应深度通道的贡献,效果不佳,本文方法可以有效减轻低贡献、无贡献或负面区域的影响,实现更优的多模态融合结果。

表 3 双线性路由融合策略有效性的消融分析结果

Table 3 Ablation analysis results of the effectiveness of bilinear routing fusion strategy

融合策略	NJU2K		NLPR		SIP		STERE	
	$F_{\beta} \uparrow$	MAE $\downarrow$	$F_{\beta} \uparrow$	MAE $\downarrow$	$F_{\beta} \uparrow$	MAE $\downarrow$	$F_{\beta} \uparrow$	MAE $\downarrow$
加法融合	0.902	0.046	0.901	0.033	0.875	0.057	0.870	0.049
乘法融合	0.903	0.045	0.902	0.031	0.878	0.055	0.875	0.043
双线性路由融合	0.924	0.027	0.912	0.018	0.911	0.034	0.912	0.028

3.5.3 不同的显著性检测工具

双线性路由融合生成的最终显著性结果是由 RGB 和 RGB-D 显著性检测结果组合确定的,因此,本文比较了几种不同的 RGB 和 RGB-D 显著性检测工具,其中,EDN^[26]和 BBRF^[27]是 RGB 显著性检测工具(只测试 RGB 通道图像),CAVER^[29]和 PICRNet^[30]是 RGB-D 显著性检测工具,结果如表 4 所示。可以看出,这些工具的随机组合在性能上都优于基线模型,这证明了本文方法的泛化性和鲁棒性,即可以在不依赖特定显

著性检测工具的情况下实现性能改进。

3.5.4 阈值的消融分析

在式(7)中,需要一个关键阈值来区分深度质量的高低。在融合阶段,质量感知结果将根据此阈值进行选择融合,该关键阈值直接影响融合深度信息的量。经过统计分析,本文设置了 0、0.3 和 0.5 这 3 个固定阈值进行消融分析,实验结果如表 5 所示。综合来说,当阈值为 0.3 时融合效果最佳,因此,本文将此关键阈值设为 0.3。

表 4 显著性检测工具的消融分析结果
Table 4 Ablation analysis results of saliency detection tools

分类	工具	NJU2K		NLPR		SIP		STERE	
		$F_{\beta} \uparrow$	MAE $\downarrow$	$F_{\beta} \uparrow$	MAE $\downarrow$	$F_{\beta} \uparrow$	MAE $\downarrow$	$F_{\beta} \uparrow$	MAE $\downarrow$
RGB	EDN	0.881	0.041	0.878	0.034	0.871	0.047	0.885	0.039
	BBRF	0.905	0.036	0.888	0.028	0.895	0.033	0.891	0.033
RGB-D	CAVER	0.920	0.028	0.903	0.023	0.897	0.043	0.892	0.034
	PICRNet	0.919	0.029	0.910	0.021	0.900	0.040	0.904	0.036
融合结果	EDN+CAVER	0.922	0.028	0.905	0.022	0.899	0.041	0.895	0.033
	EDN+PICRNet	0.921	0.028	0.911	0.020	0.903	0.039	0.905	0.034
	BBRF+CAVER	0.924	0.027	0.910	0.022	0.905	0.036	0.903	0.030
	BBRF+PICRNet	0.924	0.027	0.912	0.018	0.911	0.034	0.912	0.028

表 5 关键阈值的消融分析结果
Table 5 Ablation analysis results of key thresholds

阈值	NJU2K		NLPR		SIP		STERE	
	$F_{\beta} \uparrow$	MAE $\downarrow$	$F_{\beta} \uparrow$	MAE $\downarrow$	$F_{\beta} \uparrow$	MAE $\downarrow$	$F_{\beta} \uparrow$	MAE $\downarrow$
0	0.921	0.029	0.909	0.019	0.913	0.034	0.909	0.032
0.3	0.924	0.027	0.912	0.018	0.911	0.034	0.912	0.028
0.5	0.920	0.028	0.911	0.018	0.908	0.035	0.906	0.034

3.6 时间复杂度

本文方法无须额外的训练过程,避免了引入过多的参数和计算复杂度。深度质量评估和显著性检测仅需计算一次 RGB 显著性结果,并且通过细粒度质量评估,仅高质量的深度区域参与融合,

可以明显减少计算开销。表 6 展示了 SOTA 模型与本文方法在时间、参数量、FLOPs 和性能方面的对比结果,可以看出,本文方法在提升性能的同时保持了较好的时间效率,实现了性能与效率的良好平衡。

表 6 不同方法的时间、参数量、FLOPs、性能的对比结果

Table 6 Comparison results of time, parameter quantity, FLOPs, and performance of different methods

方法	时间/s	参数量/ 10^6	FLOPs/ 10^9	NJU2K		NLPR	
				$F_{\beta} \uparrow$	MAE $\downarrow$	$F_{\beta} \uparrow$	MAE $\downarrow$
CAVER	0.148	93.8	31.6	0.920	0.028	0.903	0.023
PICRNet	0.069	111.9	27.1	0.919	0.029	0.910	0.021
本文方法	0.115	142.2	55.8	0.924	0.027	0.912	0.018

4 结束语

深度图的质量受场景和拍摄设备的影响而存在差异,低质量的深度图包含噪声等干扰信息,对模型性能产生不利影响。针对此问题,本文提出一种基于细粒度深度质量感知的 RGB-D 显著性物体检测方法。通过细粒度地比较 RGB 和 D 分支的显著性结果与伪标签的相似性差异,准确评估深度通道不同区域的质量,该评估结果指导一种双线性路由融合策略,其中路由方向由深度通道质量感知结果驱动,优化了 RGB 和深度通道之间的协同作用。在 NJU2K、NLPR、SIP 和 STERE 数据集上的实验结果表明,即使在具有挑战性的场景中,本文方法也能

实现准确的显著性目标检测,该方法显著降低了低质量、无贡献甚至负面影响深度区域的干扰,实现了更优的多模态互补状态,并获得了优于其他 SOTA 方法的检测结果。在未来,将探索更高效的深度质量感知算法,以减少计算开销并增强实时性能。同时,通过结合生成对抗网络,提升显著性检测模型对复杂场景的泛化能力。

参考文献

[1] SONG M K, SONG W F, YANG G W, et al. Improving RGB-D salient object detection via modality-aware decoder [J]. IEEE Transactions on Image Processing, 2022, 31: 6124-6138.

[2] 李振鲁, 黄威, 孙锴. 复杂环境下的轻量化道路目标识别算法研究 [J]. 计算机工程, 2024, 50(4): 219-227.

- LI Z L, HUANG W, SUN K. Research on lightweight road-target-recognition algorithm in complex environment [J]. Computer Engineering, 2024, 50(4): 219-227. (in Chinese)
- [3] CHENG M M, HOU Q B, ZHANG S H, et al. Intelligent visual media processing: when graphics meets vision [J]. Journal of Computer Science and Technology, 2017, 32(1): 110-121.
- [4] SONG M K, LI L F, WU D Q, et al. Rethinking object saliency ranking: a novel whole-flow processing paradigm [J]. IEEE Transactions on Image Processing, 2024, 33: 338-353.
- [5] LI L M, CHEN C, ZHANG X W. Mask-guided self-distillation for visual tracking [C]//Proceedings of the IEEE International Conference on Multimedia and Expo (ICME). Washington D. C., USA: IEEE Press, 2022: 1-6.
- [6] CHEN C, ZHAO H M, YANG H, et al. Full-reference screen content image quality assessment by fusing multilevel structure similarity [J]. ACM Transactions on Multimedia Computing, Communications, and Applications, 2021, 17(3): 1-21.
- [7] 孙福明, 胡锡航, 武景宇, 等. 跨模态交互融合与全局感知的 RGB-D 显著性目标检测 [J]. 软件学报, 2024, 35(4): 1899-1913.
SUN F M, HU X H, WU J Y, et al. RGB-D salient object detection based on cross-modal interactive fusion and global awareness [J]. Journal of Software, 2024, 35(4): 1899-1913. (in Chinese)
- [8] 于洋洋, 吴敦全, 陈程立诏. 基于稀疏对比自蒸馏的 RGB-D 显著性物体检测 [J]. 计算机工程, 2025, 51(10): 327-335.
YU Y Y, WU D Q, CHEN C. RGB-D saliency object detection based on sparse contrastive self-distillation [J]. Computer Engineering, 2025, 51(10): 327-335. (in Chinese)
- [9] 吴映霓, 闫河, 姜彬, 等. HDF+: 一种新的 RGB-D 显著性目标检测网络模型 [J]. 小型微型计算机系统, 2025, 46(7): 1645-1651.
WU Y N, YAN H, JIANG B, et al. HDF+: a new RGB-D salient object detection network [J]. Journal of Chinese Computer Systems, 2025, 46(7): 1645-1651. (in Chinese)
- [10] 宋梦柯, 郑元超, 陈程立诏. 基于深度质量感知和分层特征引导的 RGB-D 显著性检测 [J]. 计算机工程, 2023, 49(5): 255-261, 268.
SONG M K, ZHENG Y C, CHEN C. RGB-D saliency detection via depth quality perception and hierarchical feature guidance [J]. Computer Engineering, 2023, 49(5): 255-261, 268. (in Chinese)
- [11] LIU Z Y, SHI S, DUAN Q T, et al. Salient object detection for RGB-D image by single stream recurrent convolution neural network [J]. Neurocomputing, 2019, 363: 46-57.
- [12] REN J Q, GONG X J, LU Y, et al. Exploiting global priors for RGB-D saliency detection [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Washington D. C., USA: IEEE Press, 2015: 25-32.
- [13] CHEN H, LI Y F. Progressively complementarity-aware fusion network for RGB-D salient object detection [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2018: 3051-3060.
- [14] ZHAO J X, CAO Y, FAN D P, et al. Contrast prior and fluid pyramid integration for RGBD salient object detection [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Washington D. C., USA: IEEE Press, 2020: 3922-3931.
- [15] DING Y, LIU Z, HUANG M K, et al. Depth-aware saliency detection using convolutional neural networks [J]. Journal of Visual Communication and Image Representation, 2019, 61: 1-9.
- [16] SHIGEMATSU R, FENG D, YOU S D, et al. Learning RGB-D salient object detection using background enclosure, depth contrast, and top-down features [C]//Proceedings of the IEEE International Conference on Computer Vision Workshops (ICCVW). Washington D. C., USA: IEEE Press, 2018: 2749-2757.
- [17] QU L Q, HE S F, ZHANG J W, et al. RGBD salient object detection via deep fusion [J]. IEEE Transactions on Image Processing, 2017, 26(5): 2274-2285.
- [18] CHEN Q, FU K R, LIU Z, et al. EF-Net: a novel enhancement and fusion network for RGB-D saliency detection [J]. Pattern Recognition, 2021, 112: 107740.
- [19] JI W, LI J J, YU S, et al. Calibrated RGB-D salient object detection [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Washington D. C., USA: IEEE Press, 2021: 9466-9476.
- [20] ZHOU W J, ZHU Y, LEI J S, et al. LSNet: lightweight spatial boosting network for detecting salient objects in RGB-thermal images [J]. IEEE Transactions on Image Processing, 2023, 32: 1329-1340.
- [21] CHEN Q, LIU Z, ZHANG Y, et al. RGB-D salient object detection via 3D convolutional neural networks [J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2021, 35(2): 1063-1071.
- [22] FAN D P, ZHAI Y J, BORJI A, et al. BBS-Net: RGB-D salient object detection with a bifurcated backbone strategy network [EB/OL]. [2024-07-05]. <https://arxiv.org/abs/2007.02713v1>.
- [23] CHENG X L, ZHENG X, PEI J L, et al. Depth-induced gap-reducing network for RGB-D salient object detection: an interaction, guidance and refinement approach [J]. IEEE Transactions on Multimedia, 2023, 25: 4253-4266.
- [24] KE L, YE M Q, MARTIN D, et al. Segment anything in high quality [EB/OL]. [2024-07-05]. <https://arxiv.org/abs/2306.01567>.
- [25] ACHANTA R, SHAJI A, SMITH K, et al. SLIC superpixels compared to state-of-the-art superpixel methods [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(11): 2274-2282.
- [26] WU Y H, LIU Y, ZHANG L, et al. EDN: salient object detection via extremely-downsampled network [J]. IEEE Transactions on Image Processing, 2022, 31: 3125-3136.
- [27] MA M C, XIA C Q, XIE C X, et al. Boosting broader receptive fields for salient object detection [J]. IEEE Transactions on Image Processing, 2023, 32: 1026-1038.
- [28] 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.
- [29] PANG Y W, ZHAO X Q, ZHANG L H, et al. CAVER: cross-modal view-mixed transformer for bi-modal salient object detection [J]. IEEE Transactions on Image Processing, 2023, 32: 892-904.
- [30] CONG R M, LIU H Y, ZHANG C, et al. Point-aware interaction and CNN-induced refinement network for RGB-D salient object detection [EB/OL]. [2024-07-05]. <https://arxiv.org/abs/2308.08930>.
- [31] JU R, LIU Y, REN T W, et al. Depth-aware salient object detection using anisotropic center-surround difference [J]. Signal Processing: Image Communication, 2015, 38: 115-126.
- [32] PENG H W, LI B, XIONG W H, et al. RGBD salient object detection: a benchmark and algorithms [EB/OL]. [2024-07-

- 05]. https://link.springer.com/chapter/10.1007/978-3-319-10578-9_7.
- [33] NIU Y Z, GENG Y J, LI X Q, et al. Leveraging stereopsis for saliency analysis [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2012: 454-461.
- [34] ACHANTA R, HEMAMI S, ESTRADA F, et al. Frequency-tuned salient region detection [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2009: 1597-1604.
- [35] PERAZZI F, KRÄHENBÜHL P, PRITCH Y, et al. Saliency filters: contrast based filtering for salient region detection [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2012: 733-740.
- [36] LIAO G B, GAO W, LI G, et al. Cross-collaborative fusion-encoder network for robust RGB-thermal salient object detection [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 32(11): 7646-7661.
- [37] 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): 10261-10269.
- [38] CONG R M, LIN Q W, ZHANG C, et al. CIR-Net: cross-modality interaction and refinement for RGB-D salient object detection [J]. IEEE Transactions on Image Processing, 2022, 31: 6800-6815.

文字编辑 吴云芳
栏目编辑 宋 圆