

卢浩博,李林,邢勇,等.双门控跨模态注意力显著目标检测方法[J].辽宁工程技术大学学报(自然科学版),2026,45(4):502-512.
DOI:10.11956/j.issn.1008-0562.20260063

Lu Haobo,Li Lin,Xing Yong,et al.ECCLNet: a bi-gated cross-modal attention salient target detection approach[J].Journal of Liaoning Technical University (Natural Science),2026,45(4):502-512.DOI:10.11956/j.issn.1008-0562.20260063

双门控跨模态注意力显著目标检测方法

卢浩博¹, 李 林¹, 邢 勇¹, 胡云飞², 于保才^{2*}

(1. 国能宝日希勒能源有限公司, 内蒙古 呼伦贝尔 021500; 2. 辽宁工程技术大学 电子与信息工程学院, 辽宁 葫芦岛 125105)

摘 要: 针对显著目标检测 (salient object detection, SOD) 中多模态特征融合存在的特征维度固定、计算复杂度高、类别不平衡突出等问题, 以及因上下文建模不足、细节特征丢失、跨模态耦合僵化导致的性能瓶颈, 提出一种基于双门控跨模态注意力的显著目标检测模型 ECCLNet。该模型通过跨模态增强模块 (cross-modal enhancement module, CEM) 实现 RGB 与深度/热红外模态特征的自适应交互。设计跨层特征细化模块 (cross-level refinement module, CLRM), 结合可变形卷积与动态加权策略, 实现多尺度特征的精确对齐与融合。针对类别不平衡问题, 构建基于平方根平衡策略的加权二值交叉熵 (weighted binary cross-entropy, WBCE) 损失函数, 以稳定训练过程并提升边界预测精度。在 NLPR、SIP、VT 系列 RGB-D、RGB-T 显著目标检测基准数据集上开展对比实验, 结果表明: ECCLNet 在检测精度与效率均优于对比算法, 且计算开销显著降低, 可为复杂场景下的多模态视觉感知、智能检测提供技术参考。

关键词: 显著目标检测; 跨模态融合; 双门控注意力; 多尺度特征细化; RGB-D/RGB-T

中图分类号: TP391

文献标志码: A

文章编号: 1008-0562(2026)04-0502-11

ECCLNet: a bi-gated cross-modal attention salient target detection approach

Lu Haobo¹, Li Lin¹, Xing Yong¹, Hu Yunfei², Yu Baocai^{2*}

(1. Guoneng BaoRixile Energy Company Limited, Inner Mongolia Autonomous Region, Hulunbuir 021500, China; 2. College of Faculty of Electronic and Information Engineering, Liaoning Technical University, Huludao 125105, China)

Abstract: In order to solve the problems of fixed feature dimensions, high computational complexity, and serious category imbalance in multimodal feature fusion in the current Salient Object Detection (SOD), and the deficiencies of the existing methods in terms of insufficient context modelling, loss of detailed features, and cross-modal coupling stiffness lead to the limited detection performance, this paper proposes a novel bimodal gated cross-modal attention salient target detection method, ECCLNet, to improve detection accuracy and efficiency. The Cross-modal Enhancement Module (CEM) is adopted to achieve adaptive interaction between RGB and depth/thermal infrared modal features, and the Cross-Level Refinement Module (CLRM) is designed to combine deformable convolution and dynamic weighting strategies to achieve accurate alignment and detection of multi-scale features. Meanwhile, for the problem of category imbalance, a Weighted Binary Cross-Entropy (WBCE) loss function based on the square root balancing strategy is proposed to stabilise the training and improve the boundary prediction accuracy. Comparative experiments are carried out on standard RGB-D and RGB-T salient object detection data-sets of NLPR, SIP and VT series. The results indicate that the proposed

收稿日期: 2026-02-02 修回日期: 2026-05-02 接受日期: 2026-05-09 责任编辑: 焦丽

基金项目: 辽宁省联合基金项目 (20240346); 辽宁省教育厅基金项目 (24-094); 辽宁工程技术大学鄂尔多斯研究院校地科技合作培育项目 (YJY-XD-2024-B-014)

作者简介: 卢浩博 (1991-), 男, 内蒙古 呼伦贝尔人, 硕士, 高级工程师, 主要从事机电管理、工程管理和计算机视觉方面的研究。

E-mail: 20006680@chenenergy.com.cn

通信作者: 于保才 (1993-), 男, 辽宁 凌源人, 博士, 讲师, 主要从事计算机视觉与模式识别、信息系统与智能决策方面的研究。

E-mail: lntuybc@163.com

ECCLNet outperforms all competing algorithms in both detection accuracy and inference efficiency, and can offer technical support for studies related to multi-modal visual perception and intelligent object detection under complex scenarios.

Keywords: salient object detection; cross-modal fusion; bi-gated attention; multi-scale feature refinement; RGB-D/RGB-T

0 引言

显著目标检测是计算机视觉领域的重要基础任务,在自动驾驶、视频监控和图像检索等领域应用广泛。随着多模态感知技术的发展,RGB-D与RGB-T数据融合方案逐步成熟,有效提升了显著目标检测在复杂环境下的适配能力。该方法能够充分挖掘不同模态间的互补信息,进而提升检测的准确性和鲁棒性。

随着深度学习的发展,基于卷积神经网络(CNN)的显著目标检测方法逐渐成为研究主流。相关研究表明,通过多尺度特征提取、上下文建模和端到端优化策略,能显著提升检测精度^[1-5]。但在低对比度、复杂遮挡或弱纹理场景下,单一RGB模态仍存在信息不足的问题。为此,学者引入深度(Depth)或热红外(Thermal)等辅助模态,推动RGB-D与RGB-T显著目标检测技术的发展。

在RGB-D显著目标检测领域,文献[6]构建系统性的基准数据集,验证深度信息对显著性建模的有效性。文献[7]~文献[10]围绕跨模态特征融合展开,通过多分支编码器、注意力机制和协同解码策略,实现RGB与深度特征的互补增强。文献[11]提出层级交互、边缘引导和模态感知解码等方法,有效提升目标边界定位能力和结构一致性。然而,在模态退化或噪声干扰情况下,现有显著目标检测方法仍容易引入冗余信息,影响算法整体性能。

相比RGB-D任务,RGB-T显著目标检测更关注低照度、夜间或复杂光照环境下的鲁棒性。文献[12]通过构建大规模RGB-T数据集和设计跨模态协同网络结构,验证了热红外模态在弱光条件下的优势。部分方法引入图模型或多尺度融合机制以增强模态互补性,但静态融合策略在复杂场景中适应性不足,难以兼顾细节保持与全局一致性。

近年来,多模态SOD取得显著进展^[13-17],但仍存在若干挑战:部分方法未能充分挖掘模态间互补性,静态融合(如拼接或统一加权)在噪声或模态退化情况下难以适应;语义层间特征对齐不足,易导致空间细节丢失;类别不平衡使模型偏向背景像

素,降低前景区域的检测精度。此外,基于Transformer的模型^[18](如SwinNet、TriTransNet)虽具备全局上下文建模能力,但在保持空间一致性与边界细节^[19-21]方面存在不足。

针对上述问题,提出一种轻量化自适应的多模态显著目标检测框架ECCLNet。设计两种互补的门控机制,调节跨模态和跨尺度维度的特征交互;引入跨模态增强模块(CEM),通过模态级门自适应地过滤噪声模态特定信息;设计跨层细化模块(CLRM),配备尺度级门控机制,动态平衡高层语义特征与低层空间特征;构建基于平方根平衡的加权二分类交叉熵损失函数,缓解类别不平衡问题。

1 ECCLNet模型总体结构

ECCLNet模型由双分支编码器、跨模态增强模块(CEM)和跨层特征细化模块(CLRM)组成,其整体结构见图1。该模型能够提升跨模态特征融合能力并优化多层次特征表征,实现高精度的RGB-D显著目标检测。

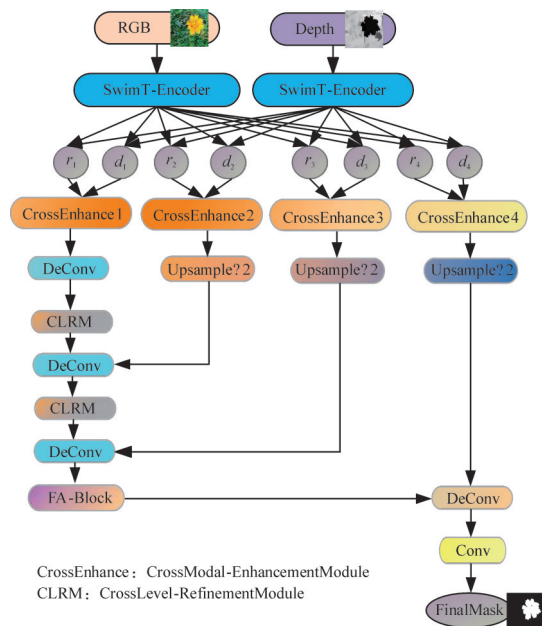


图1 ECCLNet的结构

Fig.1 Structure of the ECCLNet

ECCLNet 的核心设计采用双门控机制,即 CEM 中的跨模态门和 CLRM 中的跨层级门,共同调节特征交互,见图 2。ECCLNet 可视为一种双门控 (bi-gated) 结构,包含两个相互协同的门控机制。第一个门控作用于跨模态增强模块 (CEM),属于模态级门控,用于自适应抑制噪声模态信息(如退化的深度或热红外数据),并强化 RGB 与辅助模态之间的互补信息。第二个门控嵌入跨层细化模块 (CLRM),属于层级门控,用于在高层语义特征与低层空间细节之间动态分配权重,从而兼顾全局语义一致性与局部边界细节。两个门控分别从模态维度与层级维度协同作用,前者提升特征质量,后者优化特征表达,从而在复杂场景下实现鲁棒性更强的显著目标检测。

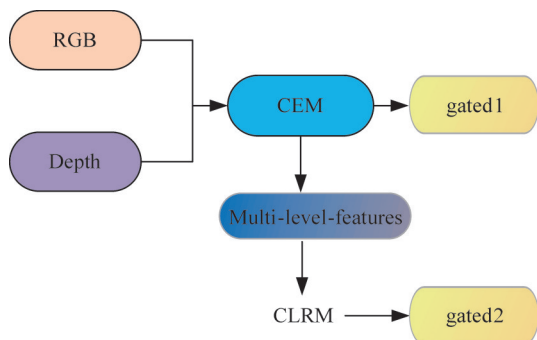


图 2 ECCLNet 中双门机制的示意

Fig.2 Illustration of the bi-gated mechanism in ECCLNet

1.1 交叉模态增强模块

CEM 模块位于编码器之后,用于实现动态跨模态特征交互,其核心思想是通过双向门控注意力机制抑制模态噪声并突出互补特征,提升特征融合质量和网络鲁棒性。

图 3 中 CEM 模块的工作流程如下。

(1) 逐元素相乘实现模态特征交互

设编码器输出 RGB 特征 $F_r \in \mathbf{R}^{H \times W \times C}$ 与深度特征 $F_d \in \mathbf{R}^{H \times W \times C}$, H 、 W 、 C 分别表示特征图的高度、宽度和通道数。对两组特征执行逐元素乘法运算,生成初始交互融合特征,表示为

$$F_{\text{INT}} = F_r \odot F_d, \quad (1)$$

式中, $\odot$ 表示逐元素乘法运算。

(2) 非线性变换与注意力权重生成

为进一步优化跨模态交互特征,将 F_{INT} 送入卷积层,并使用 GELU 激活函数进行非线性变换,在增强模型非线性的同时保留细粒度特征信息。卷积和激活后的特征表达式为

$$F_{\text{conv1}} = \text{GELU}(\text{Conv}_{3 \times 3}(F_{\text{INT}})). \quad (2)$$

经学习得到的注意力权重图定义为

$$M_{\text{att}} = \sigma(\text{Conv}_{3 \times 3}(F_{\text{conv1}})), \quad (3)$$

式 (2)、式 (3) 中: $\text{Conv}_{3 \times 3}(\cdot)$ 为卷积核尺寸为 3×3 的卷积操作; $\text{GELU}(\cdot)$ 为高斯误差线性单元激活函数; $\sigma(\cdot)$ 为 Sigmoid 激活函数; $M_{\text{att}} \in \mathbf{R}^{H \times W \times C}$, 取值区间为 $[0, 1]$ 。

(3) 特征重标定与增强

利用生成的注意力掩码对原始 RGB 特征进行重标定,通过再次逐元素相乘实现特征增强。经过注意力机制增强后的 RGB 特征表示为

$$F_{\text{enh}} = F_r \odot M_{\text{att}}, \quad (4)$$

模块的最终输出特征表示为

$$F_{\text{out}} = F_{\text{enh}} + F_d. \quad (5)$$

该操作能够突出显著区域,抑制无关特征。式 (5) 引入增强特征与原始特征的残差连接,有助于梯度传播,并防止特征增强过程中的信息损失。

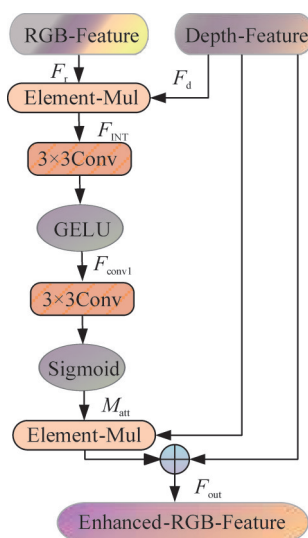


图 3 CEM 结构

Fig.3 Structure of cross-modal enhancement module

1.2 跨层细化模块

在显著目标检测中,不同层级特征融合对于准确定位目标和保持边界细节至关重要。CLRM 通过高层语义引导低层特征细化,实现多尺度特征的一致性与结构对齐。

CLRM 结构如图 4 所示,其遵循分层细化 (hierarchical refinement) 流程开展特征处理。高层特征对齐 (high-level feature alignment) 通过卷积调整通道维度,并采用双线性上采样 ($\times 2$) 使高层特征与对应低层特征在空间尺度上匹配。

(1) 上采样操作

上采样后的高层特征表示为

$$F'_h = \text{Upsample}(\text{DConv}(\text{Conv}_{1 \times 1}(F_h))), \quad (6)$$

式中: F_h 为高层特征图; $\text{Conv}_{1 \times 1}(\cdot)$ 为 1×1 卷积运算, 用于调整通道维度; $\text{Upsample}(\cdot)$ 表示双线性上采样操作; $\text{DConv}(\cdot)$ 为可变卷积运算。

(2) 门控注意力机制融合

为自适应调控高低层特征的贡献比例, 采用门控注意力机制生成融合权重。融合后的中间特征为

$$F_g = \text{GELU}(\text{Conv}_{3 \times 3}(F_L \oplus F'_h)), \quad (7)$$

式中: F_L 表示低层特征图; $\oplus$ 表示特征拼接操作。

(3) 注意力权重生成

卷积操作后通过 Sigmoid 激活函数生成注意力掩码, 即

$$M_g = \sigma(\text{Conv}_{1 \times 1}(F_g)), \quad (8)$$

式中, $M_g \in \mathbf{R}^{H \times W \times C}$, 取值范围为 $[0, 1]$ 。

(4) 权重分配与特征分解

将计算得到的掩码拆分为 2 组独立权重分量分别作用于 F_L 和 F'_h , 完成特征自适应加权。

$$\begin{cases} F'_L = W_{\text{Low}} \otimes F_L \\ F''_h = W_{\text{2AH}} \otimes F'_h \end{cases}, \quad (9)$$

式中: W_{Low} 和 W_{2AH} 分别为分配给低层特征、对齐后高层特征的权重; F'_L 表示加权后的低层特征; F''_h 表示加权后的高层特征。

(5) 残差连接与输出

为了保持原始 RGB 特征的完整性, 采用残差连接, 将细化后的特征与原始低层特征进行融合, 得到最终增强输出特征为

$$F_{\text{enh}} = F_{\text{ref}} + F_L, \quad (10)$$

式中: F_{ref} 为经过门控注意力机制细化后的特征, $F_{\text{ref}} = F'_L + F''_h$ 。

为进一步提升高层语义特征与低层空间细节特征之间的对齐能力, 在 CLRM 中引入可变形卷积 (deformable convolution)。具体而言, 在对高层特征进行通道调整与上采样后、跨层融合之前, 引入 3×3 可变形卷积对高层特征进行空间自适应建模。

该可变形卷积通过额外的卷积层学习偏移量 (offset), 使卷积采样位置能够根据目标结构进行动态调整, 从而突破传统固定感受野的限制。相比标准卷积, 其在处理目标形变、多尺度结构以及复杂背景时具有更强的适应能力。

该设计可有效缓解跨层特征在空间位置上的不对齐问题, 并在边界细节恢复与结构一致性建模方面取得更优表现。

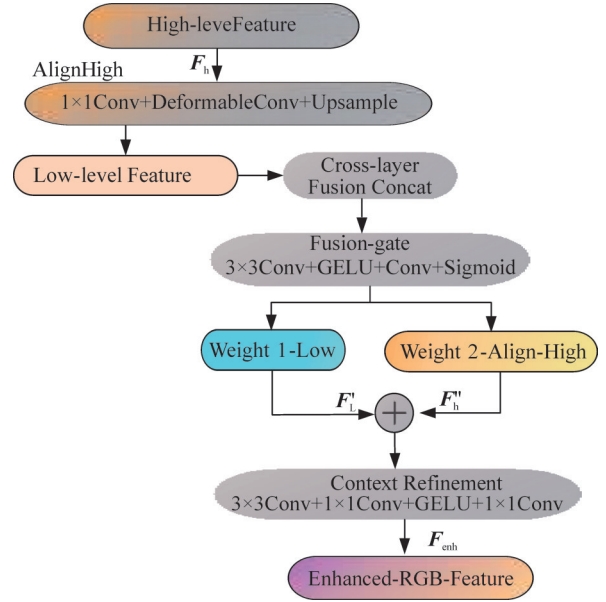


图4 CLRM结构

Fig.4 Structure of cross-level refinement module

1.3 加权二元交叉熵损失函数

为缓解显著性目标检测中的类别不平衡问题, 引入动态加权的二元交叉熵 (BCE) 损失函数。设预测显著性图为 S , 对应的真实标注掩码为 T , 则损失函数定义为

$$L_{\text{WBCE}} = \text{WBCE}(S, T, W_m), \quad (11)$$

式中: $S \in \mathbf{R}^{H \times W}$ 为模型预测的显著性图; $T \in \{0, 1\}^{H \times W}$ 为真实标注; $W_m \in \mathbf{R}^{H \times W}$ 为基于前景与背景像素分布动态计算得到的权重矩阵; $\text{WBCE}(\cdot)$ 表示带 Logits 的二元交叉熵损失函数。

权重矩阵表达式为

$$W_m = \alpha P_f + \lambda N, \quad (12)$$

式中, $P_f \in \{0, 1\}^{H \times W}$, $N \in \{0, 1\}^{H \times W}$, 分别为前景 (显著目标) 的二值掩码和背景像素的二值掩码; α 和 λ 分别表示前景和背景的权重系数。

$$\begin{cases} \alpha = \frac{\sum N}{\sum P_f + \sum N + \varepsilon} \\ \lambda = \sqrt{\frac{\sum P_f}{\sum P_f + \sum N + \varepsilon}} \end{cases}, \quad (13)$$

其中: $\sum P_f$ 和 $\sum N$ 分别为前景像素总数和背景像素总数; ε 为极小常数。

在 λ 中引入平方根因子, 用于确保较小的显著区域不会被大量的背景像素所掩盖。该加权策略能够动态调整不同像素类别的损失贡献, 提升模型对小型及复杂显著目标的检测鲁棒性。

2 实验研究

2.1 数据集与评价指标

为全面评估 ECCLNet 模型的性能,在 6 个具有代表性的 RGB-D 与 RGB-T 显著目标检测数据集上进行实验,涵盖多种复杂场景,以验证模型的鲁棒性与泛化能力。

在 RGB-D 显著目标检测任务中,选取 5 个公开数据集: DUT-RGBD、NLPR、NJU2K、SIP 与 LFSD。其中, DUT-RGBD 包含 1 200 张由 Lytro 相机采集的图像,场景结构复杂; NLPR 包含 1 000 张图像,目标尺度多样; NJU2K 包含 2 003 张立体图像,深度信息丰富; SIP 以人像显著性为主,包含 1 000 张高分辨率图像; LFSD 包含 100 张 Lytro 采集图像,是高难度的小样本数据集。在 RGB-T 显著目标检测任务中,选用 3 个经典数据集: VT821、VT1000 与 VT5000。VT821 包含多种真实场景; VT1000 聚焦复杂遮挡与背景纹理; VT5000 为大规模数据集,验证模型在大样本条件下的鲁棒性。

2.2 评价指标

采用显著目标检测(SOD)领域中常用的 4 类指标对模型进行评价。

平均绝对误差 E_{MAE} 用于衡量预测显著图与真实标签之间的平均绝对差异,即

$$E_{MAE} = \frac{1}{WH} \sum_{i=1}^W \sum_{j=1}^H [S(i,j) - G(i,j)], \quad (14)$$

式中: $S(i,j)$ 表示预测显著图在坐标 (i,j) 处的像素值; $G(i,j)$ 表示真实标签图在对应位置的像素值。

F_β 测度是衡量精确率 P 与召回率 R 的常用指标,定义为

$$F_\beta = \frac{1 + \beta^2 PR}{\beta^2 P + R}, \quad (15)$$

式中, β 为权重因子,当 $\beta < 1$ 时,表示对精确率的重视程度高于召回率。本文遵循通用设置,取 β 为 0.3,侧重精确率的量化评估。

结构相似度 S_m 融合区域感知和目标感知的结构相似性,表达式为

$$S_m = aS_o + (1-a)S_r, \quad (16)$$

式中: S_o 和 S_r 分别为目标感知和区域感知的结构相似性; $a \in [0,1]$, 用于调节两项相似度的权重。

增强对齐指标 E_m 用于衡量预测显著图与真实标签之间的空间相关性。本文引入增强对齐矩阵 M , 实现分布一致性量化,即

$$E_m = \frac{1}{WH} \sum_{i=1}^W \sum_{j=1}^H M(i,j). \quad (17)$$

2.3 与其他方法的对比

为验证 ECCLNet 的优越性,与多种 RGB-D 与 RGB-T 显著性检测方法进行对比,包括 TriTransNet、SwinNet、CPNet 等代表性模型。所有对比实验均在相同数据集与评测协议(FMC 标准)下进行,确保公平性与可复现性。可视化对比结果见图 5。

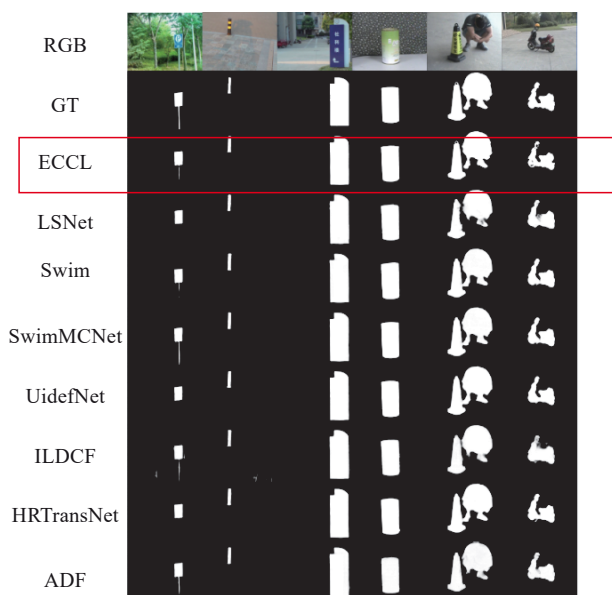


图 5 不同方法在典型样本上的显著性检测结果

Fig.5 Visual comparison of significance detection results of different methods on typical samples

由图 5 可知, ECCLNet 在捕捉细粒度细节与保持显著目标完整性方面表现突出。与地面真实标签 (GT) 相比, ECCLNet 能够更准确地刻画目标轮廓, 显著减少边界模糊和背景误检现象。在包含复杂形状目标 (如锥形物体或人物) 的场景中, ECCLNet 能有效保留边缘结构, 而 JLDCF 和 ADF 等方法则存在边缘破损或区域混淆等情况。对于细长或低对比度结构 (如路标、树枝等), ECCLNet 能较好地恢复其空间连续性, 避免在局部细节处出现的断裂、遗漏或伪预测现象。上述可视化结果表明, ECCLNet 在多尺度结构建模与

跨模态融合方面具有更优的稳定性与泛化能力。在复杂光照条件和多样化背景环境下, ECCLNet 稳定性较好, 能有效抑制噪声预测, 凸显其跨模态增强与特征细化机制的优势。

综上, ECCLNet 不仅在定量评价指标上具备竞争力, 在视觉质量和结构一致性方面同样具有明显优势。

为全面评估所提出 ECCLNet 的性能, 在 GB-D 数据集上进行定量对比实验, 包括 RGB-D 和 RGB-T 两类任务, 并结合 PR 曲线与 F-measure 曲线对模型性能进行深入分析。ECCLNet 与主流方法在多数数据集上的定量对比结果见表 1。

表 1 多数据集上算法性能比较
Tab.1 Performance comparison of algorithms on multiple datasets on multiple datasets

算法	DUT 数据集指标				NLPR 数据集指标				NJU2K 数据集指标			
	S_m	F_β	E_m	E_{MAE}	S_m	F_β	E_m	E_{MAE}	S_m	F_β	E_m	E_{MAE}
C2DFNet	0.930	0.934	0.958	0.025	0.934	0.899	0.955	0.021	0.908	0.899	0.919	0.038
CAVER	0.913	0.920	0.947	0.039	0.922	0.923	0.954	0.023	0.927	0.932	0.956	0.028
DCMNet	0.921	0.923	0.947	0.035	0.940	0.924	0.970	0.024	0.915	0.899	0.920	0.036
CPNet	0.951	0.955	0.971	0.019	0.940	0.924	0.970	0.016	0.935	0.933	0.935	0.025
TriTransNet	0.928	0.924	0.952	0.032	0.921	0.891	0.955	0.025	0.916	0.903	0.912	0.035
EBFSP	0.858	0.842	0.890	0.067	0.924	0.897	0.952	0.026	0.903	0.894	0.907	0.039
HAINet	0.910	0.906	0.938	0.038	0.931	0.897	0.957	0.024	0.912	0.900	0.922	0.038
SwinNet	0.945	0.941	0.967	0.023	0.927	0.908	0.959	0.018	0.931	0.919	0.929	0.028
本文	0.952	0.960	0.978	0.018	0.936	0.927	0.971	0.017	0.932	0.935	0.961	0.023

算法	SIP 数据集指标				LSFD 数据集指标			
	S_m	F_β	E_m	E_{MAE}	S_m	F_β	E_m	E_{MAE}
C2DFNet	0.872	0.867	0.915	0.054	0.863	0.863	0.883	0.065
CAVER	0.895	0.909	0.935	0.043	0.863	0.865	0.907	0.074
DCMNet	0.885	0.883	0.926	0.047	0.877	0.867	0.906	0.064
CPNet	0.907	0.917	0.941	0.035	0.892	0.890	0.919	0.049
TriTransNet	0.886	0.892	0.924	0.043	0.866	0.868	0.889	0.066
EBFSP	0.885	0.869	0.917	0.049	0.861	0.862	0.893	0.070
HAINet	0.880	0.875	0.919	0.053	0.854	0.852	0.877	0.079
SwinNet	0.911	0.912	0.943	0.035	0.878	0.868	0.899	0.060
本文	0.919	0.936	0.952	0.030	0.896	0.888	0.926	0.048

由表 1 可知, ECCLNet 在绝大多数评价指标中相较于对比算法具有优势, E_{MAE} 最小、 S_m 与 F_β 最大, 验证了跨模态增强与跨层细化模块的有效性。ECCLNet 在基准数据集 DUT-RGBD、NLPR、NJU2K、SIP 和 LFSD 上均取得了具有竞争力的检测效果。在 DUT-RGBD 数据集上, ECCLNet 各项评价指标均达到最优。相较于次优算法 CPNet, E_{MAE} 降低 6.3%。在 NLPR 数据集上, ECCLNet 的 E_m 为 0.971, E_{MAE} 为 0.017, 优于 CPNet 和 DCMNet, 在保持高结构相似性的同时有效降低预测误差。在 NJU2K 数据集上, 相比次优算法 CAVER, ECCLNet 在增强对齐测度上提升 0.5%, 验证了跨模态特征融合策略的有效性。在 SIP 数据集上,

ECCLNet 各项指标均为最优, 相比 CPNet 方法, F_β 提升 2.1%。在 LFSD 数据集上, ECCLNet 的 E_{MAE} 最小, 相比于次优算法 CPNet, 降低 2.0%, 表明 ECCLNet 在小样本、复杂背景场景中具有良好的泛化能力和鲁棒性。

综合定量评估结果可证明, ECCLNet 通过跨模态增强模块和跨层级细化模块构建的特征融合策略, 能够有效提升 RGB-D 显著目标检测的精度和鲁棒性。

为从阈值和召回率的角度评估模型性能, 绘制不同数据集上的 P - R 曲线, 见图 6。ECCLNet 在中高召回率区间 ($R > 0.7$) 保持高精确率, 在复杂数据集 (如 NJU2K 与 SIP) 上性能优势更为突出。

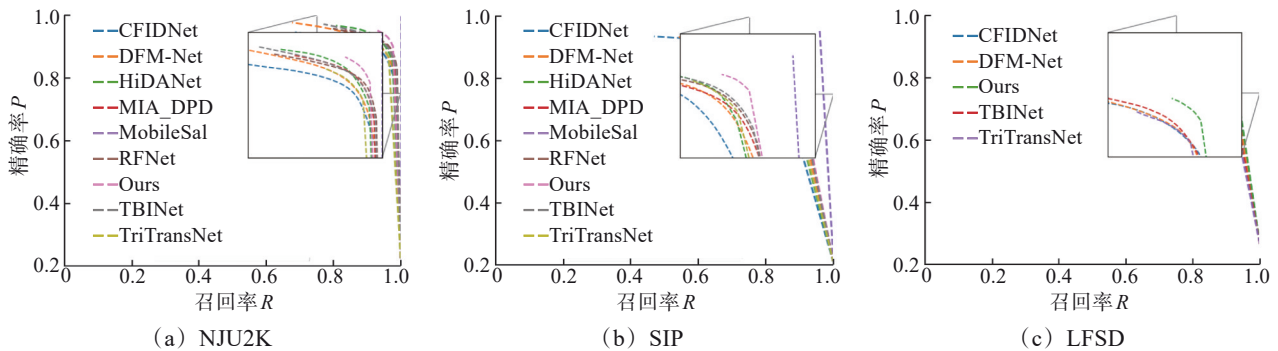


图 6 4 个 RGB-D 数据集上不同模型的 P - R 曲线

Fig.6 Precision-Recall (P - R) curves of different methods on four RGB-D datasets

为进一步验证 ECCLNet 在红外多模态场景下的泛化能力, 在 3 个数据集 (VT821、VT5000、VT1000) 上进行 P - R 评估, 结果见图 7。

在 3 个数据集上, ECCLNet 的 P - R 特性更为平稳。在 VT821 数据集的中高召回率区间内, 该模型维持较高精度, 其 P - R 特性优于 MTMR、PoolNet 等方法, 能够在中等规模 RGB-T 场景中有效抑制背景干扰, 保持目标区域的完整性。在大规模样本数据集 VT5000 上, 该模型于中低召

回区间呈现显著的精确率优势, 检测性能优于 PDNet、R3Net 等方法, 表明其在大样本条件下具备良好的稳定性与泛化能力; 在 VT1000 数据集上, ECCLNet 召回率在低至中召回率区间性能优势突出, 其 P - R 曲线优于 SGDL、M3S-NIR 和 CPD 等方法。尽管召回率较大时, 部分方法精度略高, 但在不同召回率区间, ECCLNet 均保持了较为均衡的性能, 验证了其在中小规模 RGB-T 数据集上的鲁棒性。

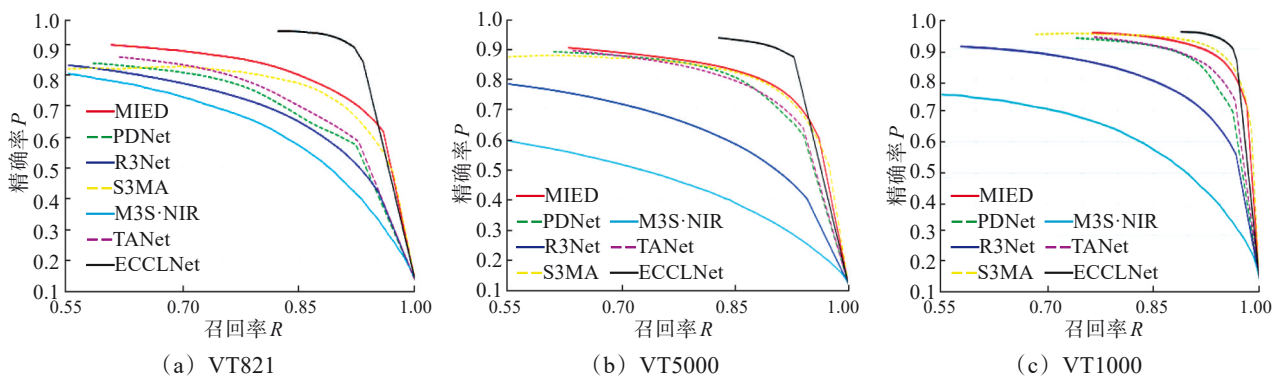


图 7 不同模型在 RGB-T 数据集上的 P - R 曲线

Fig.7 P - R curves of different methods on RGB-T datasets

为分析模型对不同阈值的敏感性,在上述 3 个 RGB-T 数据集上对比 F_{β} 值,结果见图 8。在 VT821、VT1000 和 VT5000 数据集上,ECCLNet 均表现出良好的阈值鲁棒性。在 VT821 数据集上,ECCLNet 在较宽阈值范围内维持了较高的 F_{β} 值,其整体性能优于 PoolNet、MTMR 和 CPD 等方法,对显著性阈值变化具有更优异的适应能力。在

VT5000 数据集上,ECCLNet 在整个阈值区间内 F_{β} 变化相对平稳,且在主要阈值范围内优于 PDNet、R3Net 和 SwinNet,进一步验证了其在大规模复杂场景中的稳定检测性能。对于 VT1000 数据集,ECCLNet 在大多数阈值下均取得较高的 F_{β} 值,优于 S2MA 和 M3S-NIR 等方法,相较于对比算法具有更强的阈值鲁棒性。

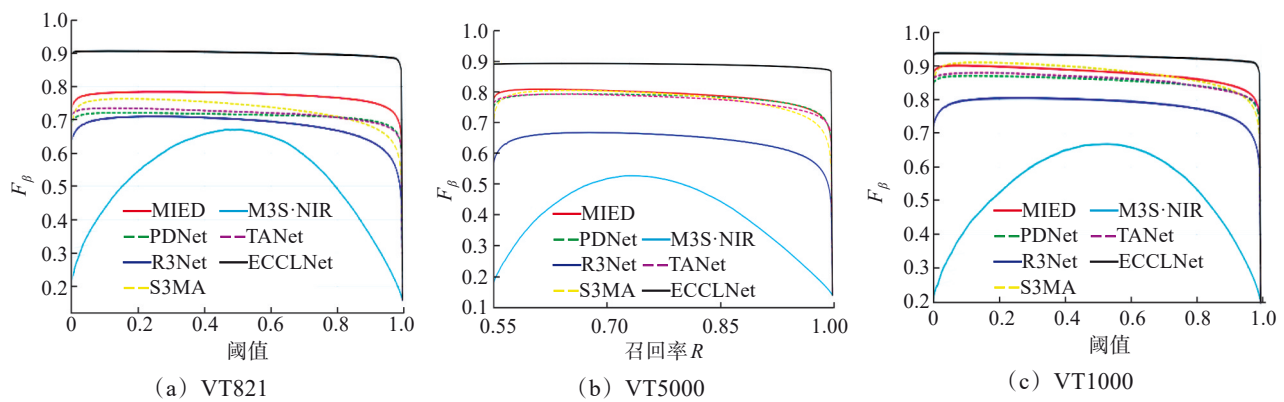


图8 不同方法在 RGB-T 数据集上的 F_{β}

Fig.8 F_{β} of different methods on RGB-T datasets

为进一步验证本文算法的有效性,选取 3 种基于 Transformer 的算法进行对比,各种算法的 E_{MAE} 见表 2。基于 Transformer 类算法采用全局上下文建模,在多个数据集上表现出色,但由于其采用粗

粒度的特征表示,在处理细粒度边界保持问题时能力不足。ECCLNet 利用双门机制提升特征质量,并细化多级表示,在多个基准测试中实现了更准确的边界定位和更具竞争力的性能表现。

表 2 基于 Transformer 算法与本文算法的 E_{MAE} 对比
Tab.2 E_{MAE} comparison based on Transformer algorithm and this algorithm

算法	Backbone	E_{MAE}			
		NLPR	NJU2K	SIP	DUT
SwinNet	Transformer	0.025	0.028	0.035	0.024
CMFormer	Transformer	0.023	0.026	0.033	0.022
CATNet	Transformer	0.024	0.027	0.034	0.023
ECCLNet	CNN	0.021	0.023	0.030	0.018

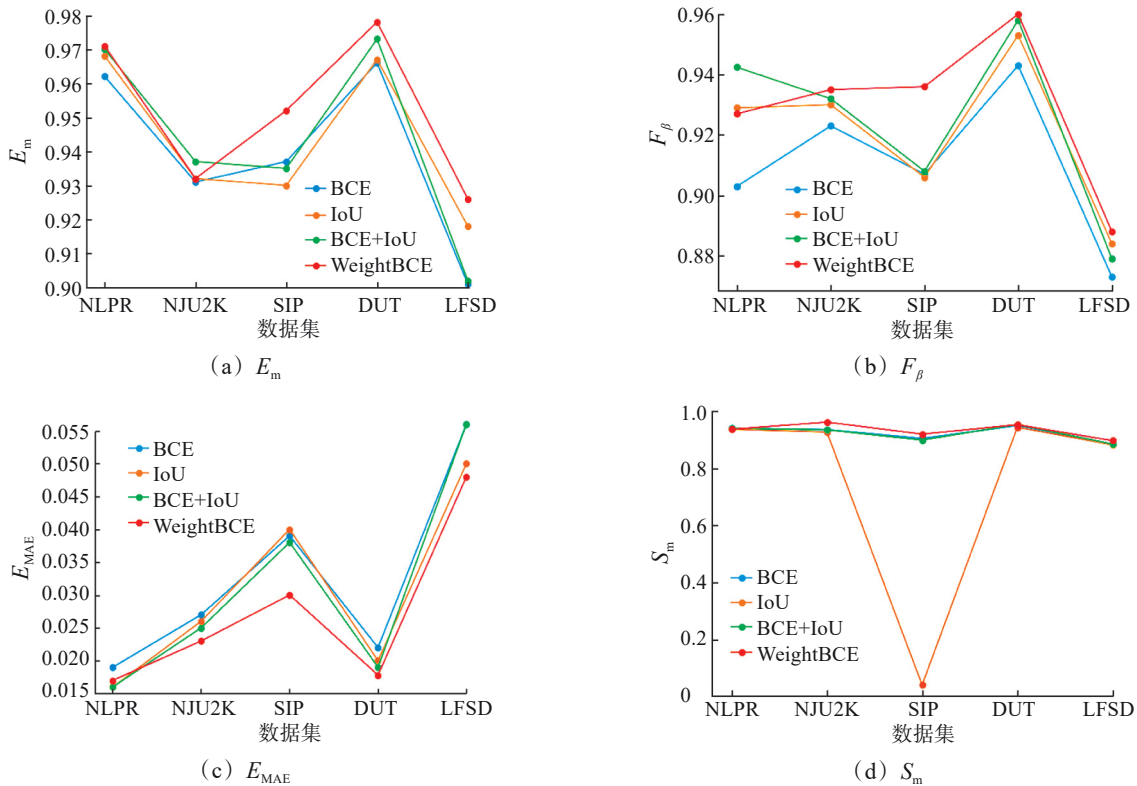
2.4 消融实验

为验证各模块贡献,设计 4 组消融实验,包含去除跨层细化模块 (No-CLRM) 与跨模态增强模块 (No-CEM),以及不同损失函数组合 (BCE、IoU、BCE+IoU、Dice Loss、Focal Loss、Weighted BCE)。实验结果表明,去除 CLRM 或 CEM 均导致显著性能下降,尤其在 DUT 与 SIP 数据集上, E_{MAE} 上升明显。Weighted BCE 损失函数表现最佳,在所有数据集上 E_{MAE} 最小, F_{β} 最大。

为进一步验证算法采用 WBCE 损失函数的有

效性,在 DUT 数据集上扩展比较范围,纳入 Focal Loss 和 Dice Loss。实验表明,采用 WBCE 算法的 E_{MAE} 和 F_{β} 表现最佳。与 Focal Loss 相比, WBCE 避免了过度强调难样本,优化过程更稳定。与 Dice Loss 相比, WBCE 采用像素级平衡策略,能更好保留边界细节。

在 NLPR、NJU2K、SIP、DUT 及 LFSD 五个数据集上开展 4 组消融实验 (BCE、IoU、BCE+IoU 和 WeightBCE),分析各损失函数在不同数据集上的性能差异, E_m 、 S_m 、 F_{β} 和 E_{MAE} 的变化情况见图 9。

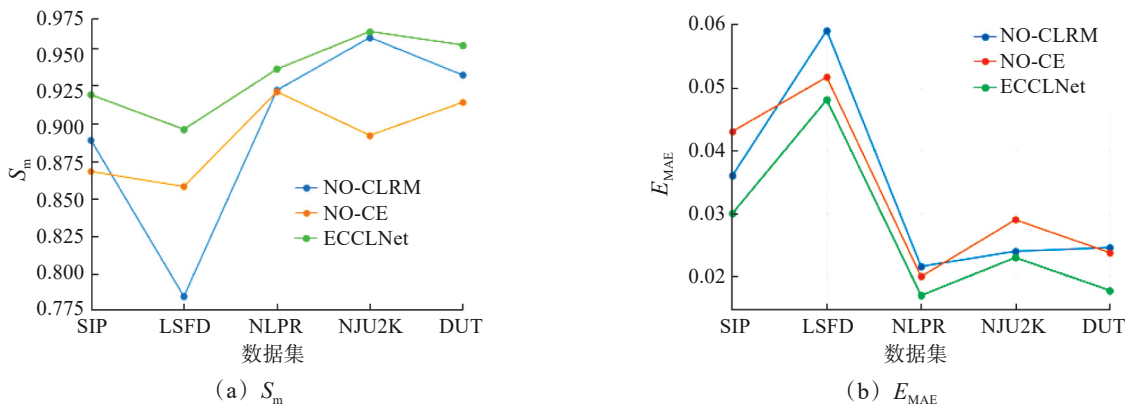
图9 各数据集上不同损失函数的 E_m 、 F_β 、 S_m 、 E_{MAE} 对比Fig.9 Comparison of E_m , S_m , F_β , E_{MAE} metrics of different loss functions on each dataset

为评估 ECCLNet 中各关键模块的贡献, 在 5 个基准数据集 (SIP、LFSD、NLPR、NJU2K、DUT) 上进行模块消融实验, 分别去除跨层细化模块 (cross-level refinement module, CLRM) 和跨增强模块 (cross-enhance module, CEM)。 E_m 、 S_m 、 F_β 和 E_{MAE} 的变化情况见图 10。

去除 CLRM 模块 (No-CLRM) 后, 模型性能在所有数据集上均出现明显下降。在 DUT 数据集上, ECCLNet 的 E_{MAE} 由 0.017 8 恶化至 0.246 0, 表明 CLRM 在多尺度特征精细化处理中起着关键作用; 在 NLPR 数据集上, No-CLRM 同样导致 E_{MAE} 升高至 0.021 6, 且 F_β 和 S_m 均明显下降。

去除 CEM (No-CEM) 后, 各项评价指标均有所退化。在 SIP 数据集上, E_{MAE} 由 0.030 增加至 0.043, F_β 则从 0.936 下降至 0.904。在 LFSD 和 NJU2K 数据集上的变化趋势类似, No-CEM 导致 E_{MAE} 升高、分割准确度下降, 说明 CEM 在特征增强及目标一致性方面发挥了重要作用。

集成 CLRM 和 CEM 后, ECCLNet 在所有数据集上均取得最优表现, 获得最低的 E_{MAE} 及最高的 F_β 、 E_m 和 S_m 值。例如, 在 NLPR 数据集上, ECCLNet 的 F_β 为 0.927, E_m 为 0.971, 显著优于 No-CLRM 和 No-CEM。上述结果充分体现了所提模块在提升分割精度和鲁棒性方面的有效性。



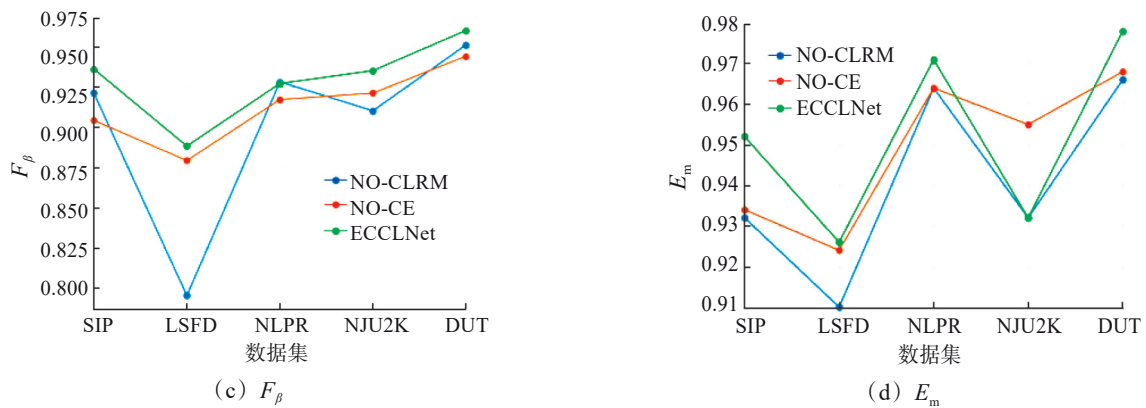


图 10 各数据集上移除CRLM、CEM模块后的模型性能对比

Fig.10 Comparison of CRLM,CEM mouldles of different loss functions on each dataset

为验证可变形卷积在CLRM中的有效性,从模型中移除可变形卷积进行消融试验,结果见表3。移除可变形卷积后,模型在所有数据集上的性能均持续下降, E_{MAE} 较大、 F_{β} 较小,说明固定采样网格无法自适应适配目标多样几何变化。

表 3 CLRM 中可变形卷积的消融研究
Tab.3 Ablation study on deformable convolution in CLRM

数据集	变异体	可变卷积状态	E_{MAE}	F_{β}	E_m
NLPR	ECCLNet	×	0.020	0.918	0.964
	ECCLNet	√	0.017	0.927	0.971
NJU2K	ECCLNe	×	0.025	0.921	0.955
	ECCLNet	√	0.023	0.935	0.961
SIP	ECCLNet	×	0.036	0.912	0.941
	ECCLNet	√	0.030	0.936	0.952

注:“×”“√”分别表示模型中未采用和采用可变形卷积模块。

2.5 模型复杂度与效率分析

为全面评估模型计算效率,选取参数量(Params)、浮点运算量(FLOPs)及推理速度(FPS)为对比指标,对比结果见表4。

ECCLNet的参数量略高于SwinNet和ADF,这表明ECCLNet在捕捉细粒度结构信息方面表达能力更强。ECCLNet的浮点运算量明显高于轻量级模型,如AFNet和LSNet,但远低于EGNet,这说明ECCLNet模型对复杂度与精度平衡良好。ECCLNet的推理速度低于ADF和EGNet,但其检测精度提升显著,该推理速度仍具竞争力。

表 4 模型复杂性和效率对比

Tab.4 Model complexity and efficiency analysis

模型	参数量/ 10^6	浮点运算量/ 10^9	推理速度/(帧· s^{-1})
ADF	175.291	135.857	135.857
DCNet	108.491	107.815	107.815
MIDD	52.428	87.774	87.774
SwinNet	198.784	124.724	124.724
LSNet	4.564	1.231	1.231
R3Net	60.240	56.160	56.160
EGNet	108.070	270.800	270.800
AFNet	37.100	38.400	38.400
ECCLNet	217.200	131.900	66.100

3 结论

本文提出一种双门控跨模态注意力框架ECCLNet,用于实现鲁棒且高效的多模态显著目标检测。通过引入跨模态增强模块(CEM)、跨层特征细化模块(CLRM)以及动态平衡的加权二值交叉熵损失函数,有效缓解了模态融合噪声、语义与细节错配以及类别严重不平衡等关键问题。

在6个RGB-D与RGB-T显著目标检测基准数据集上的实验结果表明,ECCLNet在保持边缘细节和实时推理速度方面均具有显著优势。同时,模块化设计提升了模型的可解释性与泛化能力,为伪装目标检测等相关视觉任务提供了可迁移的结构参考。

然而,ECCLNet的参数规模较大,在边缘计算与移动设备上的部署仍面临一定挑战;此外,尽管在深度与热红外模态下表现优异,但其在红外或极低照度环境中的适应性尚需进一步研究与验证。

未来工作将重点关注 ECCLNet 的轻量化设计与自适应模态选择机制,以提高其在实际复杂场景中的部署效率和通用性。

参考文献(References):

- [1] Kim J H, Lee J. Layered non-photorealistic rendering with anisotropic depth-of-field filtering[J]. *Multimedia Tools and Applications*, 2020, 79(1): 1291-1309.
- [2] Le T N, Sugimoto A. Video salient object detection using spatiotemporal deep features[J]. *IEEE Transactions on Image Processing*, 2018, 27(10): 5002-5015.
- [3] Wei S K, Liao L X, Li J, et al. Saliency inside: learning attentive CNNs for content-based image retrieval[J]. *IEEE Transactions on Image Processing*, 2019, 28(9): 4580-4593.
- [4] Yue H H, Guo J C, Yin X J, et al. Salient object detection in low-light images via functional optimization-inspired feature polishing [J]. *Knowledge-Based Systems*, 2022, 257: 109938.
- [5] Wang X H, Liu Z B, Liesaputra V, et al. Feature specific progressive improvement for salient object detection[J]. *Pattern Recognition*, 2024, 147: 110085.
- [6] Zhou W J, Guo Q L, Lei J S, et al. ECFFNet: effective and consistent feature fusion network for RGB-T salient object detection [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2022, 32(3): 1224-1235.
- [7] Fan D P, Zhai Y J, Borji A, et al. BBS-Net: RGB-D salient object detection with a bifurcated backbone strategy network[C]// *Computer Vision-ECCV 2020*. Cham: Springer, 2020:275-292.
- [8] Zhao X Q, Pang Y W, Zhang L H, et al. Suppress and balance: a simple gated network for salient object detection[C]// *Computer Vision-ECCV 2020*. Cham: Springer, 2020: 35-51.
- [9] Zhao J X, Liu J J, Fan D P, et al. EGNet: edge guidance network for salient object detection[C]//2019 IEEE/CVF International Conference on Computer Vision. October 27-November 2, 2019. Seoul, Korea. IEEE, 2019: 8778-8787.
- [10] Bao L X, Zhou X F, Lu X K, et al. Quality-aware selective fusion network for V-D-T salient object detection[J]. *IEEE Transactions on Image Processing*, 2024, 33: 3212-3226.
- [11] Liang B C, Luo H L. MEANet: an effective and lightweight solution for salient object detection in optical remote sensing images[J]. *Expert Systems with Applications*, 2024, 238: 1217785.
- [12] Hao C, Yu Z T, Liu X, et al. A simple yet effective network based on vision transformer for camouflaged object and salient object detection[J]. *IEEE Transactions on Image Processing*, 2025, 34: 608-622.
- [13] Yuan G J, Song J T, Li J J. IF-USOD: multimodal information fusion interactive feature enhancement architecture for underwater salient object detection[J]. *Information Fusion*, 2025, 117: 102806.
- [14] Yang Q N, Zheng J H, Chen J. Multilevel diverse feature aggregation network for salient object detection[J]. *Neurocomputing*, 2025, 628: 129648.
- [15] Xu M Y, Zhou Z P, Xu H B, et al. CP-net: contour-perturbed reconstruction network for self-supervised point cloud learning[J]. *IEEE Transactions on Multimedia*, 2024, 26: 8799-8810.
- [16] Fu K R, Fan D P, Ji G P, et al. JL-DCF: Joint learning and densely-cooperative fusion framework for RGB-D salient object detection [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 2020: 3049-3059.
- [17] Peng D G, Zhou W Y, Pan J Z, et al. MSEDNet: multi-scale fusion and edge-supervised network for RGB-T salient object detection [J]. *Neural Networks*, 2024, 171: 410-422.
- [18] Chen J, Zhang H Y, Gong M M, et al. Collaborative compensative transformer network for salient object detection[J]. *Pattern Recognition*, 2024, 154: 110600.
- [19] Liu Y, Li C X, Dong X H, et al. Seamless detection: unifying salient object detection and camouflaged object detection[J]. *Expert Systems with Applications*, 2025, 274: 126912.
- [20] 陈虹,李泓绪,金海波. 多尺度卷积与双注意力机制融合的入侵检测方法[J]. *辽宁工程技术大学学报(自然科学版)*,2024,43(1): 93-100.
- Chen Hong, Li Hongxu, Jinhaibo. Intrusion detection method based on multi-scale convolution and dual attention mechanism [J]. *Journal of Liaoning Technical University (Natural Science)*, 2024,43(1):93-100.
- [21] 陈占国,陈振军,薛晨霞,等. 基于单目相机的复杂场景深度估计网络[J]. *辽宁工程技术大学学报(自然科学版)*,2025,44(4):505-512.
- Chen Zhanguo, Chen Zhenjun, Xue Chenxia, et al. A complex scene depth estimation network based on monocular camera[J]. *Journal of Liaoning Technical University (Natural Science)*,2025, 44(4):505-512.