

文章编号: 2095-4980(2023)03-0360-11

基于多模态遥感影像的边缘感知引导显著性检测

连远锋^{1a,1b}, 石旭^{1a}, 江澄²

(1. 中国石油大学(北京) a. 信息科学与工程学院; b. 石油数据挖掘北京市重点实验室, 北京 102249;
2. 北京空间机电研究所, 北京 100094)

摘要: 针对多模态遥感影像显著性检测鲁棒性差和检测精确度不佳等问题, 提出一种基于多模态边缘感知引导的显著性检测方法, 该方法主要由多模态遥感影像显著检测主干网络、跨模态特征共享模块和边缘感知引导网络构成。通过在特征提取主干网络中加入跨模态特征共享模块, 使得不同模态间特征通过共享交互实现协同增强, 并且抑制具有缺陷的特征信息。基于边缘感知引导网络, 通过边缘图监督模块来检测边缘特征的有效性, 从而生成准确边界。在 3 种显著目标检测遥感图像数据集上进行实验, 平均的 F_β 、平均绝对误差(MAE)、 S_m 分数分别为 0.917 6, 0.009 5 和 0.919 9。实验结果表明, 提出的多模态边缘感知引导网络(MEGNet)适用于在多模态场景中显著性检测。

关键词: 多模态遥感图像; 显著性检测; 边缘感知引导网络; 双线性特征融合

中图分类号: TP751

文献标志码: A

doi: 10.11805/TKYDA2022216

Edge aware guidance saliency detection based on multi-modal remote sensing image

LIAN Yuanfeng^{1a,1b}, SHI Xu^{1a}, JIANG Cheng²

(1a. Department of Computer Science and Technology; 1b. Beijing Key Laboratory of Petroleum Data Mining, China University of Petroleum, Beijing 102249, China; 2. Beijing Institute of Space Mechanics & Electricity, Beijing 100094, China)

Abstract: To address the problems of poor robustness and poor detection accuracy of multimodal remote sensing image saliency detection, this paper proposes a method based on the novel and efficient Multi-modal Edge aware Guidance Network(MEGNet), which mainly consists of a salient detection backbone network for multi-modal remote sensing images, a cross-modal feature sharing module and an edge aware guidance network. First of all, a Cross-modal Feature Sharing Module(CFSM) is used during feature extraction for remote sensing image pairs, which encourages different modalities to complement each other in the feature extraction process and suppresses the influence of defective feature data from different modalities. Secondly, based on the Edge Aware Guidance Network(EAGN), the effectiveness of edge features is detected through the edge map supervision module and the final salient detection map will have clear boundaries. Finally, experiments are carried out on three kinds of saliency objects detection remote sensing image datasets. The average F_β , Mean Absolute Error(MAE) and S_m scores are 0.917 6, 0.009 5 and 0.919 9, respectively. The experimental results show that the proposed MEGNet is suitable for saliency detection in multi-modal scenes.

Keywords: multi-modal remote sensing images; saliency detection; Edge Aware Guidance Network; bilinear fusion

显著性检测在智慧城市、军事、海洋资源勘探、环境监测、交通运输等领域得到了广泛应用, 也是计算机视觉中一项具有挑战性的任务^[1-3]。它模仿人类的视觉注意力, 突出显示图像中最为显著的物体。目前, 显著性

收稿日期: 2022-11-01; 修回日期: 2022-12-26

基金项目: 国家自然科学基金资助项目(61972353); 中国石油天然气集团有限公司-中国石油大学(北京)战略合作科技专项资助项目(2006A10401006)

检测方法主要分为传统机器学习方法和深度学习方法。传统基于手工特征的显著性检测方法又可分为自下而上的方法^[4-5]和自上而下的方法^[6-7]。LI等^[4]提出正则化随机游走排序的自底向上显著性检测方法(Regularized Random Walks Ranking, RRWR), 该方法同时利用了基于区域的特征和图像细节。YUAN等^[5]在此基础上构建一种显著性回归校正过程(Reversion Correction, RC), 有效地中和了边界相邻前景区域的负面影响。传统方法在复杂的场景中, 由于方法中手工设计特征具有局限性, 目标显著性检测的准确率降低。近年来, 随着深度学习的发展, 基于卷积神经网络(Convolutional Neural Network, CNN)的模型使用像素级标注的显著性映射进行端到端训练^[8-11], 使得显著性目标检测方法准确率进一步提升, LIU等^[12]在CNN中扩展池化操作的作用, 逐步细化高级语义特征以增强检测准确率(PoolNet)。但这些方法仍然存在一些问题, 首先显著性检测任务本质上是一个像素级的预测任务, 而大多数显著性检测模型忽略了对目标边界信息的约束^[9,12-14], 这会导致最终检测结果图具有模糊的边界细节, 同时降低了检测准确率。为此, HOU等^[8]在跳连接结构中引入短连接整合低层级的边缘特征(Deeply Supervised Salient object detection with short connections, DSS), EGN^[11]将局部边缘信息与全局位置信息进行整合, 而ZHANG等^[15]针对光学遥感图像构建端到端的密集注意流体网络(Dense Attention Fluid Network, DAFNet)。通过深入分析发现, 深度学习网络的编码器在特征提取阶段提取出具有不同尺度和接收域的多层次特征, 解码器在检测输出阶段通过上采样和反卷积操作逐步结合各个层次的特征。而在解码的同时, 浅层特征提供详细目标边界信息, 但其特征本身也包含大量的背景噪声。

近年来, 空间遥感技术有了长足的进步, 遥感数据的分辨力越来越高, 遥感探测的方式也越来越丰富。多模态遥感图像指针对同一场景通过多源传感器(可见光、红外等)获取的图像。通过对这些具有合作性、互补性的图像进行协同处理来取得比只使用单模态遥感图像更好的结果。如HOU等^[16]通过多种模态遥感图像的指数信息构建了一种星载建筑物热异常的检测提取算法。WU等^[17]则提出一种主动微调车辆检测模型, 该模型利用了多源遥感图像中相同的对象信息。可见光遥感图像具有高分辨力和图像细节, 显著性检测需要考虑克服目标复杂背景、云雾、太阳光等环境因素的影响。而红外遥感图像能够有效避免环境因素的干扰, 从而给显著性检测任务提供更多的目标信息, 但红外图像目标也存在边缘模糊和图像细节较少等问题。融合2种模态的显著检测技术能够促进模态间特征互补, 从而提升显著检测的准确度。现有的多模态显著性检测模型采用元素相加或连接等线性方式来融合多模态特征^[18-19]。这种策略适用于背景较为简单的图像, 图像中的显著目标与背景分界明显。对于高分辨力并且具有复杂背景的遥感图像, 线性策略不能完全捕获模态间的互补信息并进行更为准确的显著性检测。在真实复杂场景的遥感图像成像过程中, 由于可见光遥感图像和红外遥感图像各自的局限性, 会生成具有缺陷的模态图像。而大多数多模态融合显著性检测方法忽略了模态图像受环境因素影响而包含干扰信息的问题^[20-21], 从而导致了低准确度的检测结果。具有缺陷的模态图像需要在特征提取阶段进行筛选与校正, 否则会导致低质量的特征信息在网络中传播, 而后续基于特征融合解码的显著性检测网络输出结果与输入特征具有强相关性。具体来说, 如果显著性检测网络的输入特征具有缺陷性, 最终的检测结果有极大可能会含有噪声, 从而导致准确性降低。为此, 与大多数独立提取模态特征的方法不同^[22-23], 通过构建多层跨模态特征共享模块, 允许在编码过程中进行不同模态特征间的交互, 进而使图像特征更具鲁棒性。

针对上述问题, 本文提出一种多模态边缘感知引导网络(MEGNet), 该网络通过在特征提取阶段中加入跨模态特征共享模块(CFSM), 使得不同模态间特征通过共享交互协同增强表示, 并且抑制具有缺陷的特征信息。同时, 设计多模态特征融合模块(Multi-modal feature Merging Block, MMB)来解决线性融合策略应用于复杂背景的多模态遥感影像时效果不佳的问题。该模块通过引入通道注意力与双线性融合模块来捕获更多的模态间互补信息。通道注意力机制能够从语义层面学习特征之间的关联, 筛选出与显著目标更为相关的通道特征。而双线性融合模块通过矩阵外积的计算方式, 在每一维度都进行不同模态特征间的交互, 可在后续的显著性检测中提供更为有效的特征信息。此外, 本文的模型还加入边缘感知引导网络(EAGN), 其中的边缘图监督模块用来检测边缘特征的有效性。经过有效性检验的边缘特征将集成到多尺度特征中, 最后解码器聚合多组特征, 生成边界准确且结构完整的显著性检测结果图。

1 数据与方法

1.1 实验数据

1.1.1 自建数据集

自建数据集的图片数据来自Landsat八号卫星, 将这些图像中包含船舶、岛屿、油罐等显著对象的区域裁剪成512×512像素的图像切片, 其分辨力为0.5~2.0 m/px, 并在像素级对图片中的显著对象进行手动标注。数据集

一共包含 800 张可见光遥感图像和 800 张红外遥感图像，其中 600 张可见光遥感影像和 600 张红外遥感图像构成训练集，而剩余的 200 张可见光遥感图像和 200 张红外遥感图像构成测试集。对于标签图像，由 3 名具有相关经验的工作者来进行显著对象标注，并且只保留半数以上工作者在标签图像中标注出的显著对象。与其他数据集相比，数据集图片检测目标更小、遥感场景更复杂，图 1 中展示了数据集中的部分样例图片。

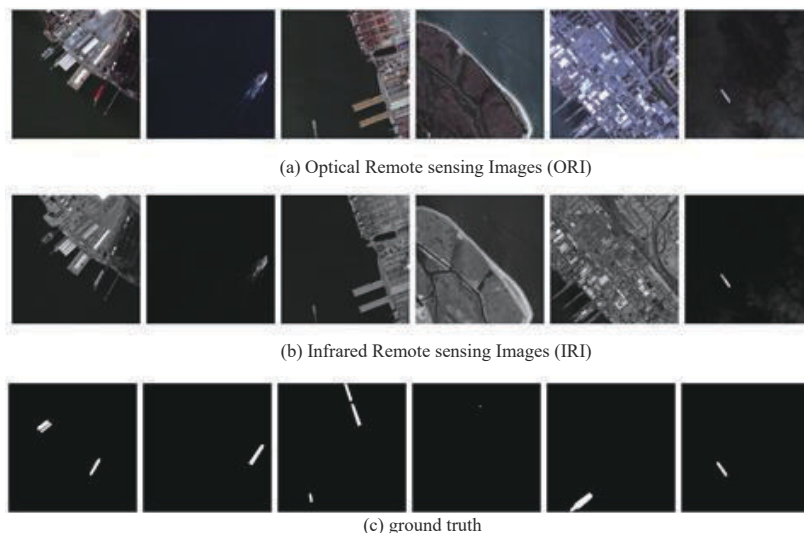


Fig.1 Our dataset
图 1 自建数据集

1.1.2 ORSSD 数据集

ORSSD 数据集^[9]是从谷歌地球和其他数据集中收集了 800 个可见光遥感影像，并且在像素级对图像中的显著目标(岛屿、船舶、车辆等)进行手动标记。由于此数据集不含有红外遥感影像，为此对 MEGNet 网络结构进行修改，将其中的跨模态特征共享(Cross-modal Feature Sharing Module, CFSSM)模块删除，网络结构由原来的双模态分支输入改为单模态分支输入，并针对此数据集进行了单独的训练。

1.1.3 EORSSD 数据集

EORSSD 数据集^[15]是基于 ORSSD 数据集构建，具有更多的显著目标类型与图像数量。除了原始的 800 张可见光遥感影像，该数据集还包括来自谷歌地球的 1 200 张可见光遥感影像。同样此数据集不包含红外遥感影像，因此采用在 1.1.2 节介绍的相同策略进行此数据集的实验。

1.2 基于多模态边缘感知引导网络

如图 2 所示，本文提出的多模态遥感影像显著性检测网络 MEGNet 由多模态编码器和解码器组成。在编码阶段借鉴了 LIAO 等^[24]的研究，首先对输入的多模态遥感图像分别进行特征提取，并在此基础上引入跨模态特征共享模块(CFSM)来进行模态间协作增强学习。与前人相比，针对于分辨率更大的遥感影像，编码阶段提取多尺度特征将有助于后续显著性检测。然后基于多模态特征融合模块(MMB)与边缘感知引导网络(EAGN)，对输入的多尺度多模态特征($InfF_i, OptF_i$)提取目标边缘特征并应用于后续的解码阶段。多模态联合特征在解码阶段通过反卷积层和上采样层来提升维度并与多尺度聚合特征 M_j 相结合，而边缘特征 F_{edge} 的加入提升了最终输出特征的边缘表达能力，进而生成高准确率的显著性检测结果图。

1.2.1 跨模态特征共享模块

由于目标的可见光/红外单模态信息存在局限性，这会降低最终显著性检测结果的准确率。为此提出一种具有鲁棒性的多模态特征共享模块来抑制有缺陷的模态信息表达并提升跨模态融合特征的特征表达能力^[24]，相比于下采样卷积操作，CFSM 针对遥感图像具有的复杂背景特性使用最大池化层来提取主要特征，过滤冗余特征。如图 3 所示，CFSM 可分为特征交叉协作过程与特征共享增强过程。

首先，对由主干网络提取的多模态特征对(INF_i, OPT_i)使用卷积模块与最大池化层分别处理 2 种模态的特征。其中，可见光分支的输出特征分别是 F_{opt} 和 F'_{opt} ，红外分支的输出特征分别是 F_{inf} 和 F'_{inf} 。然后，通过模态间特征的矩阵乘法实现交叉协作，定义为：

$$M_{inf} = (F'_{inf})^T F_{opt} \quad (1)$$

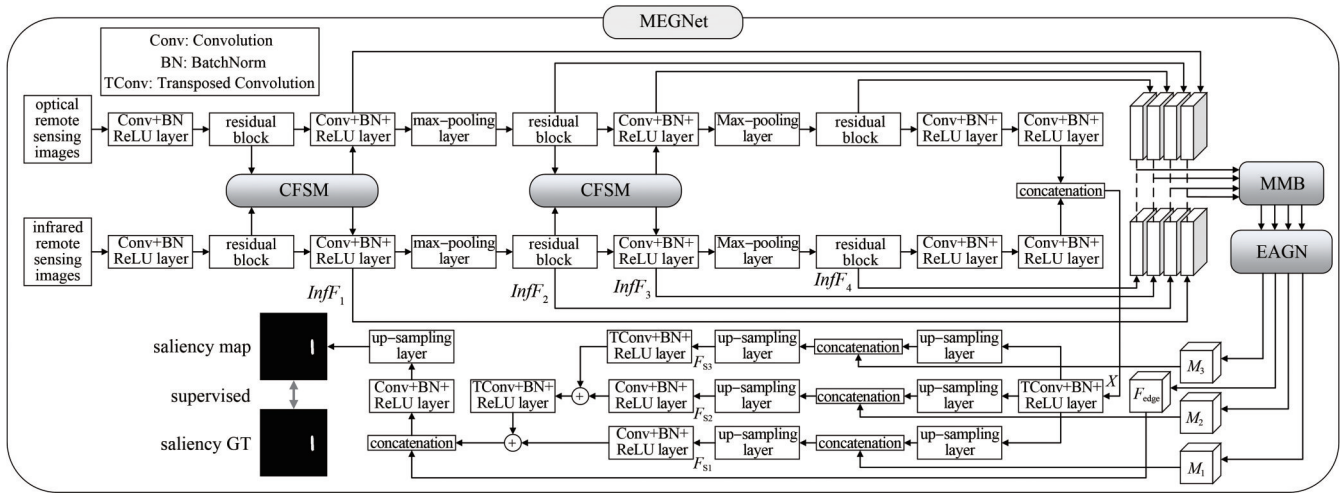


Fig.2 Architecture of MEGNet
图2 多模态边缘感知引导显著性网络(MEGNet)

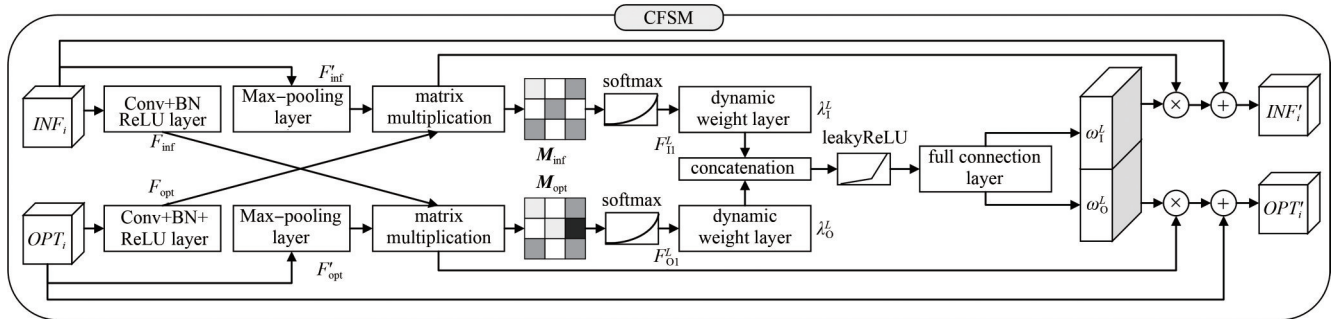


Fig.3 Architecture of CFMS
图3 跨模态特征共享模块(CFMS)

$$M_{opt} = (F'_{opt})^T F'_{inf} \quad (2)$$

式中： $(\cdot)^T$ 为转置变换； M_{inf} 为红外模态交叉协作矩阵； M_{opt} 为可见光模态交叉协作矩阵。在特征交叉融合过程中，CFMS能够实现不同模态间的特征相互增强。但是，由于单模态图像的信息局限性会使得参与交叉融合过程的特征包含冗余或者干扰信息，因此，这里使用动态权重层来重新加权多模态特征响应，其计算特征响应的表达式：

$$\lambda_c^l = \frac{1}{H \times W} \sum_{a=1}^H \sum_{b=1}^W F_c^l(a, b) \quad (3)$$

式中： (a, b) 为目标二维坐标； λ_c^l 为第 c 个通道的特征响应重要性； F_c^l 为多通道特征。根据特征响应重要性使用动态权重层自动聚合多模态特征 F'_{i1} 和 F'_{o1} ，其过程定义为：

$$\omega_i^l, \omega_o^l = FC(\text{LReLU}(C(\lambda_i^l, \lambda_o^l))) \quad (4)$$

式中：LReLU为LeakyReLU激活函数； C 为通道拼接函数；FC为全连接层； ω_i^l 和 ω_o^l 分别为每个通道相对模态的重要性。然后根据计算出的通道重要性对输入的特征图进行重新加权计算：

$$INF'_i = INF_i + (\omega_i^l \otimes M_{inf}) \quad (5)$$

$$OPT'_i = OPT_i + (\omega_o^l \otimes M_{opt}) \quad (6)$$

式中： \otimes 为元素相乘； INF'_i 和 OPT'_i 分别为红外分支与可见光分支的跨模态交互输出结果。CFMS作为编码阶段的模态间交互模块可以减轻单模态数据局限性所带来的负面影响。

1.2.2 多模态特征融合模块

遥感影像目标检测通常具有背景复杂、显著目标尺度小等特点，而红外模态图像与可见光模态图像在遇到

环境干扰时会产生较大的模态图像差距。为此设计一种多模态特征融合模块(MMB)来提取跨模态互补信息^[25], 促进模态间特征的融合, 相比于前人的研究, 红外图像特征与可见光图像特征在经过通道拼接操作后引入通道注意力来自适应地调节特征通道权重, 以此强化高效的特征通道。如图 4 所示, MMB 在线性融合策略的基础上, 加入双线性多模态融合操作与通道注意力机制。首先针对输入双模态特征对 $(InfF_i, OptF_i)$ 使用双线性融合(矩阵外积)的方式进行融合, 表示为:

$$F_i^{bf}(a, b) = OptF_i(a, b) \circ InfF_i(a, b)^T \quad (7)$$

式中 \circ 为矩阵外积。矩阵外积操作可以提取单个模态的局部特征与另一个模态的全部特征进行融合, 从而作为跨模态特征的互补信息。对 2 种模态的拼接特征, 使用通道注意力机制来筛选出具有语义信息的通道特征。

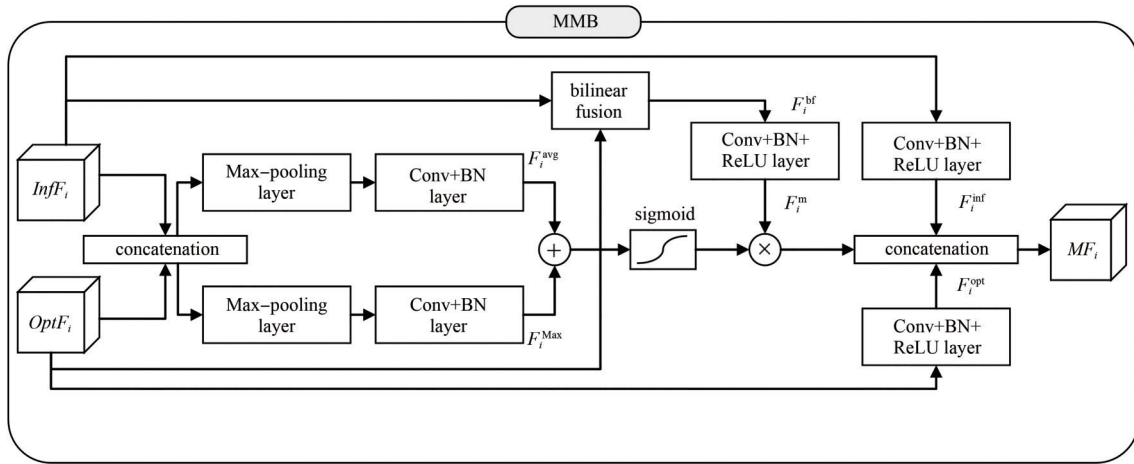


Fig.4 Architecture of the MMB
图 4 MMB 结构

虽然单一模态图像提取的特征具有局限性, 但是单一模态特征也能够为显著性检测提供有效信息^[26], 而 MMB 的输入特征都经过了 CFMS 模块的筛选。为此在 MMB 的特征输出阶段通过拼接层将模态间特征结合输出:

$$MF_i = C\left(\left(F_i^m \otimes \text{Sigmoid}\left(F_i^{avg} + F_i^{max}\right)\right), F_i^{inf}, F_i^{opt}\right) \quad (8)$$

式中: MF_i 为输出的多模态融合特征; Sigmoid 为激活函数; F_i^m 为矩阵外积结果 F_i^{bf} 经过卷积模块输出的结果。MMB 模块增强了多模态特征之间的交互作用, 同时保留有效单模态的特征信息, 可用于后续的边缘感知监督模块与显著性检测结果输出。

1.2.3 边缘感知引导网络

本文所提出的编码器—解码器主干结构虽然能够生成显著性检测结果图, 但其生成的结果图内部边界模糊, 且含有一定的噪声。为此, 受到提取边缘特征辅助显著性检测的相关文献^[11,27]与特征金字塔网络(Feature Pyramid Network, FPN)结构的启发, 构建一种输出多尺度特征与边缘特征的边缘感知引导网络(EAGN)辅助 MEGNet 生成更为精确的显著性检测结果图。FPN 结构能够自下而上地增强特征层次, 保障特征信息全面性。而边缘特征引入使得最终输出的显著性检测结果图边界清晰。如图 5 所示, 该网络可细分为多尺度特征学习模块与边缘感知模块。首先, 多尺度特征学习模块根据主干网络提取的多模态特征 $(InfF_i, OptF_i)$ 输出多尺度聚合特征 M_j , 边缘感知模块使用这些聚合特征生成用于监督训练的边缘图。同时, 边缘感知模块输出边缘特征用于解码阶段与多尺度聚合特征和编码器的输出特征进行融合, 进而生成更为准确的显著性检测图。

多尺度特征学习模块用来对融合的 4 组多模态融合特征 MF_i 做进一步处理。每组特征首先通过具有不同参数的空洞卷积层进行处理, 这里采用的空洞卷积层具有 3 种参数, 卷积核为 1×1 , 空洞率为 0, 卷积核为 3×3 , 空洞率为 3 和卷积核为 3×3 , 空洞率为 5。每组特征通过空洞卷积层后使用拼接层在通道维度对特征进行堆叠。空洞卷积层的使用使得多尺度学习模块在不增加内核大小的情况下捕获更为丰富的上下文数据。此外, 多尺度学习模块采用浅层与深层特征相结合的结构是因为浅层特征有助于捕获图像细节, 而深层特征有助于捕获图像语义信息。

现有针对红外遥感影像与可见光遥感影像的多模态显著性检测方法较少关注显著目标的边界信息, 这会导致输出的显著性检测结果图目标边界模糊。为此, 本文引入关注边界信息的边缘感知模块, 用于克服模糊边界的问题, 生成具有清晰边界的显著性检测结果图。首先针对不同尺度的特征使用上采样层变换到相同的维度,

并使用拼接层进行特征聚合，结果称为 F_t 。然后通过通道注意机制，对 F_t 进行进一步处理，过程如下：

$$F_{\text{edge}} = \text{Up} \left\{ \text{Conv}_t \left(\text{Conv}_s \left(\text{MaxP} \left(\text{Conv}_r(F_t) \right) \right) \otimes \text{Conv}_r(F_t) + \text{Conv}_r(F_t) \right) \right\} \quad (9)$$

式中：MaxP为最大池化层； Conv_r 代表卷积模块(卷积层+批归一化层+ReLU层)； Conv_s 代表卷积模块(卷积层+批归一化层+Sigmoid层)；Up为上采样层； F_{edge} 代表输出的边缘特征。同时，边缘特征将通过上采样层输出边缘检测图，并与边缘真实图通过损失函数进行监督训练。借助边缘特征与多尺度聚合特征，生成最终显著性检测结果图的过程如下：

$$F_{S_i} = \text{Up} \left(C \left(\text{Up} \left(\text{Conv}_t(X) \right), M_i \right) \right) \quad (10)$$

式中： M_i 为多尺度聚合特征； Conv_t 为反卷积模块(反卷积层+批归一化层+ReLU层)； X 为编码器输出特征。最后借助于边缘特征 F_{edge} ，对 F_{S_i} 进一步处理，生成最终的显著性检测图 S_m ，即：

$$S_m = \text{Up} \left\{ \text{Conv}_r \left[C \left(F_{\text{edge}}, \left(\text{Conv}_r(F_{S_1}) + \text{Conv}_r \left(\text{Conv}_r(F_{S_2}) + \text{Conv}_r(F_{S_3}) \right) \right) \right) \right] \right\} \quad (11)$$

最终的显著性检测结果图与显著性检测真实图也将通过损失函数进行监督训练。在多尺度学习模块与边缘感知模块的共同作用下，本文方法的输出特征会带有图像细节信息、语义信息和边缘信息，进而输出高准确度的显著性检测结果图。

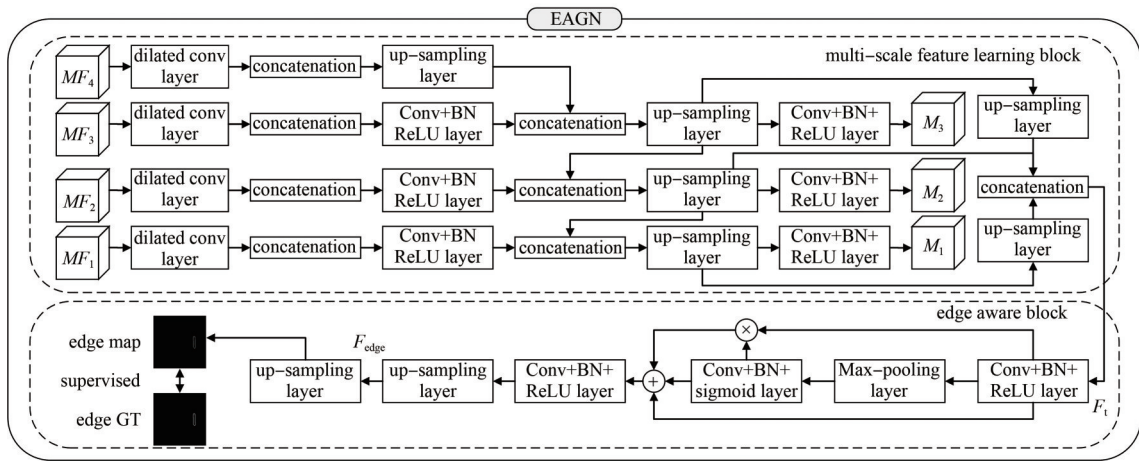


Fig.5 Architecture of EAGN

图5 EAGN结构

1.3 损失函数

为输出高准确度的显著性检测结果图，这里采用交叉熵损失函数实现边缘图监督与显著性检测图的监督训练。边缘图 E_m 是由边缘特征 F_{edge} 通过上采样操作生成的，交叉熵的损失函数分别定义如下：

$$L_{\text{edge}}(E_m) = - \sum_{i=1}^{W \times H} \left(G_{e_1}(i) \log(E_m(i)) \right) + G_{e_0}(i) \log(1 - E_m(i)) \quad (12)$$

$$L_{\text{object}}(S_m) = - \sum_{j=1}^{W \times H} \left(G_{s_1}(j) \log(S_m(j)) \right) + G_{s_0}(j) \log(1 - S_m(j)) \quad (13)$$

式中： G_{e_1} 与 G_{e_0} 分别为边缘真实图值为1和为0的像素； G_{s_1} 与 G_{s_0} 分别为显著性检测真实图值为1和为0的像素。 S_m 定义为：

$$S_m = \alpha S_o + (1 - \alpha) S_r \quad (14)$$

式中 S_o 和 S_r 分别为预测图像与真实图像之间的对象感知和区域感知结构相似性。本文设置 α 为0.5^[28]。

MEGNet的总体损失函数为：

$$L_{\text{MES}} = \theta_1 L_{\text{edge}}(E_m) + \theta_2 L_{\text{object}}(S_m) \quad (15)$$

式中 θ_1 和 θ_2 分别代表 2 项损失函数的系数。这里, θ_1 的值为 0.35, θ_2 的值为 0.65。

2 结果与分析

2.1 训练环境配置

本文采用 2 块 NVIDIA GeForce GTX1080Ti GPU 和 1 颗 16 核 CPU 在内存为 32 GB 的服务器平台进行模型训练。模型训练使用 Pytorch 深度学习框架, 动量设置为 0.9, 权重衰减设置为 0.000 5。模型的迭代次数为 1 000 个批次, 批处理大小为 2。前 750 个批次学习率设置为 0.001, 后 250 个批次学习率设置为 0.000 1。采用 Adam 优化器^[29]用于优化模型, 参数设置如下: $\beta_1=0.9, \beta_2=0.999, \varepsilon=1 \times 10^{-8}$ 。

2.2 评价指标

本文使用 P-R 曲线、MAE 分数、 F_β 和 S_m 四种指标来评价不同方法的性能。显著性检测可以通过 0~255 的整数将其阈值划分为一些二进制显著性掩码, 再与真实值进行比较, 从而获得精确度和查全率。以精确率为纵轴, 召回率为横轴绘制 P-R 曲线。 F_β 是一个整体性能指标, 计算过程为:

$$F_\beta = \frac{(1 + \beta^2) \text{Precision} \times \text{Recall}}{\beta^2 \text{Precision} + \text{Recall}} \quad (16)$$

式中 β^2 为 0.3^[28]。MAE 分数定义为:

$$MAE = \frac{1}{n} \sum_{i=1}^n |G(x)_i - y_i| \quad (17)$$

式中: n 为图片的像素数量; $G(x)_i$ 和 y_i 分别为显著性检测图像和真实图像在像素 i 处的像素值。

2.3 结果分析

为了验证本文所提出的网络 MEGNet 的有效性, 将其与 PoolNet^[12]、DAFNet-V^[15]、DAFNet-R^[15]、DSS^[8]、EGNet^[11]、RCRR^[5]和 RRWR^[4]等方法进行对比。上述方法均基于自建数据集、ORSSD 数据集和 EORSSD 数据集使用默认参数进行了相应的再训练。

表 1 不同显著性检测方法在 3 种数据集上的各项评价指标实验结果

Table 1 Experimental results of evaluation indicators for different salient detection methods on three datasets

evaluation indicators	ORSSD			EORSSD			our dataset		
	$F_\beta(\uparrow)$	MAE(\downarrow)	$S_m(\uparrow)$	$F_\beta(\uparrow)$	MAE(\downarrow)	$S_m(\uparrow)$	$F_\beta(\uparrow)$	MAE(\downarrow)	$S_m(\uparrow)$
DAFNet-V ^[15]	0.917 4	0.012 5	0.919 1	0.892 2	0.006 0	0.916 7	0.868 1	0.015 3	0.912 1
DAFNet-R ^[15]	0.923 5	0.010 6	0.918 8	0.906 0	0.005 3	0.918 5	0.885 5	0.011 3	0.916 4
PoolNet ^[12]	0.791 1	0.035 8	0.840 3	0.781 2	0.020 9	0.821 8	0.753 6	0.037 9	0.800 8
DSS ^[8]	0.783 8	0.036 3	0.826 2	0.715 8	0.018 6	0.787 4	0.646 7	0.039 7	0.737 9
EGNet ^[11]	0.843 8	0.021 6	0.872 1	0.806 0	0.010 9	0.860 2	0.757 8	0.026 8	0.849 3
RCRR ^[5]	0.594 4	0.127 7	0.684 9	0.449 5	0.164 4	0.601 3	0.368 7	0.187 9	0.512 8
RRWR ^[4]	0.595 0	0.132 4	0.683 5	0.449 5	0.167 7	0.599 7	0.370 5	0.192 3	0.507 6
proposed	0.935 5	0.011 7	0.924 1	0.913 2	0.005 7	0.918 9	0.909 1	0.009 1	0.925 5

图 6 给出了 MEGNet 与其他方法在 3 种不同遥感图像数据集上的 P-R 曲线。当召回率接近于 1 时, MEGNet 将产生更高的准确性, 这说明假阳性较低, 同时 P-R 曲线展示出的优势也说明 MEGNet 输出的显著性检测图与真实图像更为接近。表 1 列出了不同方法在 3 种数据集下各项评价指标的具体数值, MEGNet 展现出了具有良好的检测效果。可以看出, 本文所提出的方法在自建数据集上的 3 项评价指标都要优于其他方法, 而在 ORSSD 数据集与 EORSSD 数据集上也具有较好的结果。同时, 本文提出的方法在包含多模态图像的数据集上的各项评价指标与其他方法相比有较大的数值提升, 这说明多模态的特征提取策略与跨模态特征共享模块在多模态显著性检测中是有效的。在其他方法中, DAFNet-R 的 3 项评价指标分数较优, 该方法在自建数据集上的 F_β 数值为 0.885 5, MAE 分数为 0.011 3, S_m 数值为 0.916 4。而 MEGNet 在 3 项指标上都取得了较大的提升, 其中在 F_β 上提升了 2.3%, MAE 分数提升了 0.2%, 在 S_m 上提升了 1.1%。

图 7 和图 8 给出了具有不同图像属性的图像显著性检测结果, 如小目标、大目标、多目标和中心偏差等。图 7 第 3 行的测试图像和图 8 的第 3 行测试图像都具有复杂的背景, 其他方法输出的检测结果包含较多的虚警目标或未能检测出显著目标, 而 MEGNet 输出的检测结果不仅包含显著目标, 也在一定程度上减少了虚警目标。图 7 第 6 行的测试图像和图 8 第 4 行的测试图像中的显著目标较大, 其他方法输出的检测结果包含模糊的边界, 而

MEGNet 由于边缘感知引导网络的加入，输出的检测结果具有清晰的边界。从图中可看出，MEGNet 在具有各种属性的图像中都表现良好，并且比大多数其他方法都更接近真实图像。

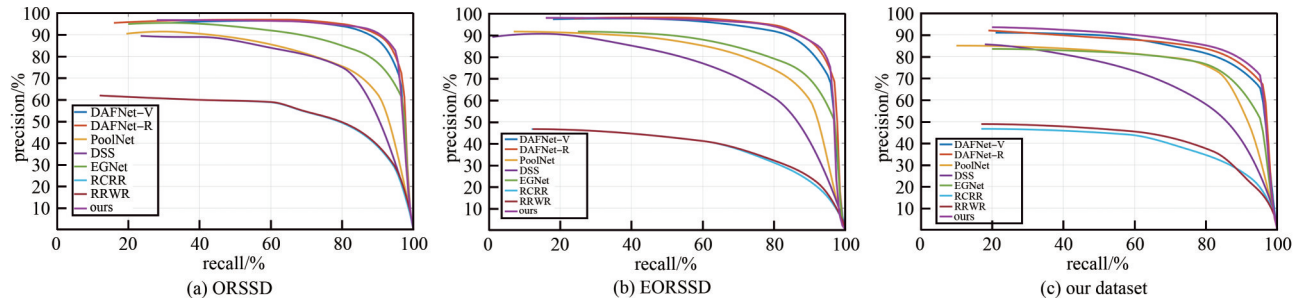


Fig.6 P-R curves of different salient detection methods on three datasets
图 6 不同显著性检测方法在 3 种数据集上的 P-R 曲线实验结果

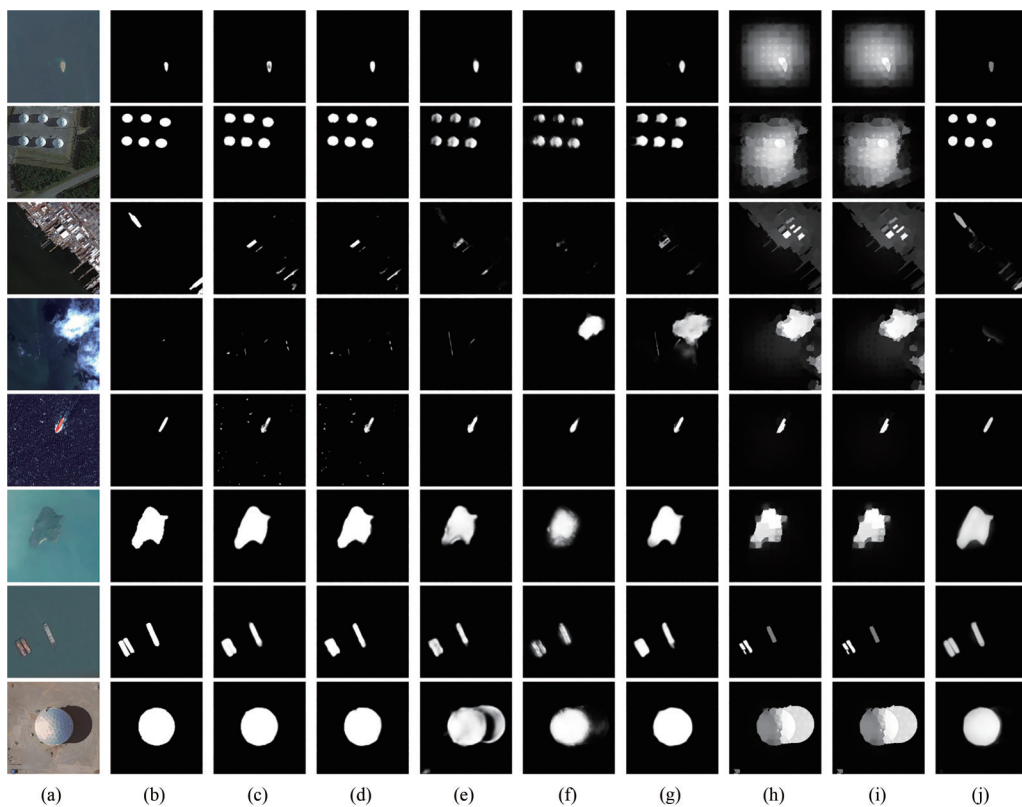


Fig.7 Visualization results of different saliency detection methods in optical remote sensing images (a) input optical remote sensing images; (b) salient detection ground truth images; (c) DAFNet-V; (d) DAFNet-R; (e) PoolNet; (f) DSS; (g) EGNet; (h) RCRR; (i) RRWR; (j) our method
图 7 在可见光遥感影像中不同显著性检测方法的可视化结果 (a) 输入可见光遥感图像; (b) 显著性检测真实图像; (c) DAFNet-V; (d) DAFNet-R; (e) PoolNet; (f) DSS; (g) EGNet; (h) RCRR; (i) RRWR; (j) 本文方法

2.4 消融实验

为了验证边缘感知引导网络的有效性，此处对 MEGNet 进行了模型分析与消融实验。EAGN 根据输入的多尺度模态融合特征 MF_i ，输出多尺度聚合特征 M_j 与边缘特征 F_{edge} ，针对 2 处在 MEGNet 的特征添加分别进行消融实验，不添加多尺度聚合特征、不添加边缘特征和 2 种特征都不添加的模型分别称之为 $Net_M^{without}$ 、 $Net_{edge}^{without}$ 与 $Net_{all}^{without}$ 。

如表 2 和图 9 所示，本文模型在评价指标和可视化结果上都明显优于上述 3 种模型，多尺度聚合特征的加入减少了背景划分为显著目标的错误，而边缘特征指导输出更加准确的结果图。同时，前 2 种添加部分特征的模型生成的结果优于最后一种 2 种特征都不添加的模型，说明边缘特征信息与多尺度聚合特征信息在 MEGNet 中的重要性。

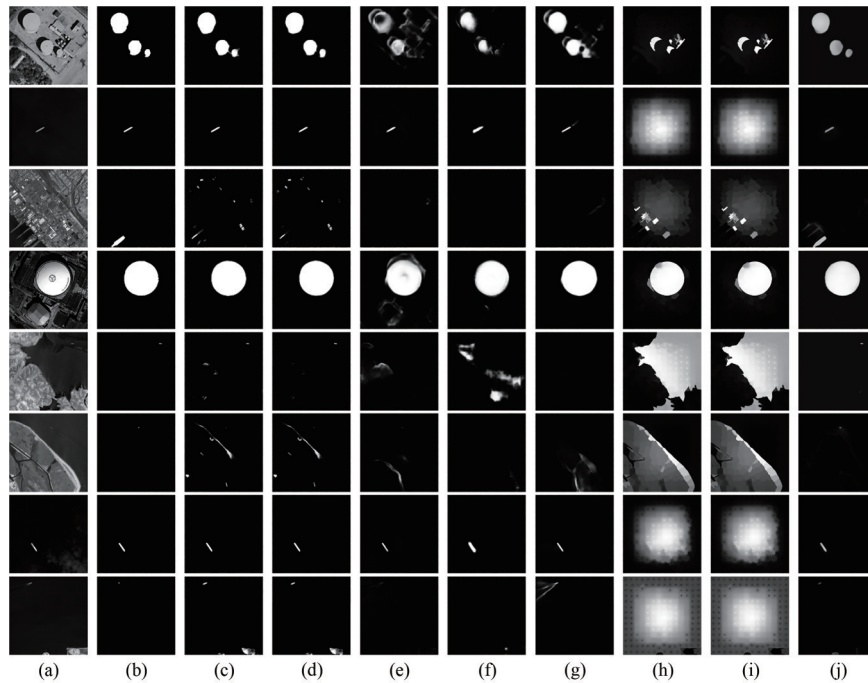


Fig.8 Visualization results of different saliency detection methods in infrared remote sensing images (a) input infrared remote sensing images; (b) salient detection ground truth images; (c) DAFNet-V; (d) DAFNet-R; (e) PoolNet; (f) DSS; (g) EGNet; (h) RCRR; (i) RRWR; (j) our method
图8 在红外遥感影像中不同显著性检测方法的可视化结果 (a) 输入红外遥感图像; (b) 显著性检测真实图像; (c) DAFNet-V; (d) DAFNet-R; (e) PoolNet; (f) DSS; (g) EGNet; (h) RCRR; (i) RRWR; (j) 本文方法

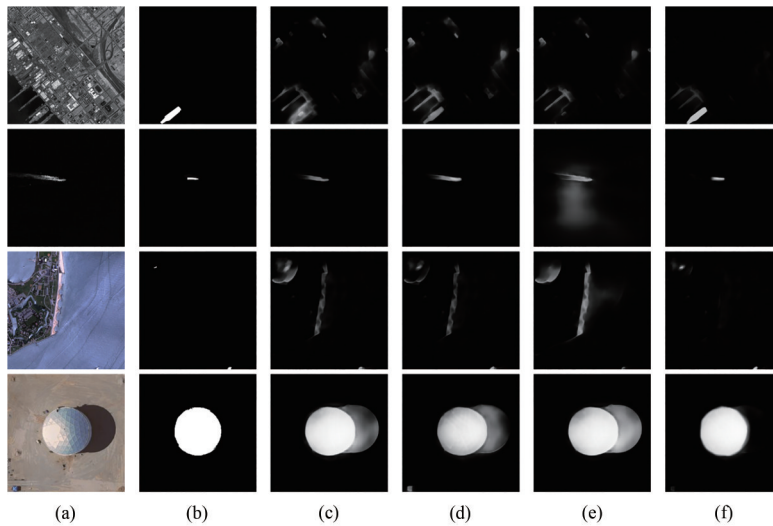


Fig.9 Comparison of visualization results of MEGNet ablation experiment (a) input images;(b) salient detection ground truth images;
(c) $Net_{edge}^{without}$ (d) $Net_M^{without}$; (e) $Net_{all}^{without}$; (f) MEGNet
图9 MEGNet消融实验的可视化结果对比 (a) 输入图像; (b) 显著性检测真实图像; (c) $Net_{edge}^{without}$; (d) $Net_M^{without}$; (e) $Net_{all}^{without}$; (f) MEGNet

表2 不同模型在自建数据集上的显著性检测实验中各项评价指标平均实验结果

Table2 Average experimental results of each evaluation metric in the salient detection experiment of different models on the self-made dataset

evaluation indicators	$F_{\beta}(\uparrow)$	MAE(\downarrow)	$S_m(\uparrow)$
$Net_M^{without}$	0.888 6	0.010 3	0.903 0
$Net_{edge}^{without}$	0.857 7	0.011 7	0.891 4
$Net_{all}^{without}$	0.831 8	0.012 1	0.875 9
MEGNet	0.909 1	0.009 1	0.925 5

3 结论

针对多模态遥感影像显著性检测精确度不佳、鲁棒性差等问题，提出了一种面向多模态遥感影像的显著性

检测方法 MEGNet, 该方法通过跨模态特征共享模块 CFMSM、多模态特征融合模块 MMB 以及边缘感知引导网络 EAGN 实现高准确度的显著性检测。具体来说, MEGNet 的主干网络首先对输入的双模态遥感影像对进行特征提取与编码, 以此获得的多尺度双模态特征输入至 CFMSM 模块与 MMB 模块进行进一步处理。CFMSM 模块能够抑制具有缺陷的特征信息进一步在网络中传播, 而 MMB 的特征融合策略相比于线性融合策略会捕获更多的模态间互补信息。其次, EAGN 模块针对 MMB 输出的多模态融合特征进一步提取边缘特征, 并与边缘图真实图像进行监督训练, 确保加入至主干解码网络的边缘特征的正确性。最后的主干解码网络融合多尺度聚合特征与边缘特征使用上采样层输出最终显著性检测结果图。实验结果表明, 基于多模态遥感影像的显著性检测方法 MEGNet 在可视化分析和评价指标分析中都取得了具有竞争力的结果; 同时, 针对边缘感知引导网络的消融实验证明, 多尺度聚合特征与边缘特征的加入在 MEGNet 显著性检测实验中具有显著提升评价指标的作用。

虽然 MEGNet 具有较高的显著性检测效果, 但检测速率相比于其他方法并无优势。未来工作中将着重研究基于知识蒸馏的网络模型优化, 特别是多模态特征融合模块中矩阵外积的优化。

参考文献:

- [1] 罗群,刘俊. 基于光谱尺度空间与管道滤波的红外目标检测[J]. 太赫兹科学与电子信息学报, 2022,20(4):346–353. (LUO Qun,LIU Jun. Infrared target detection algorithm based on fast spectral scale space and dynamic pipeline filtering[J]. Journal of Terahertz Science and Electronic Information Technology, 2022,20(4):346–353.) doi:10.11805/TKYDA2020605.
- [2] 牛文渊,黄先锋,金洁,等. 小型文物摄影测量三维建模主体对象识别[J]. 遥感学报, 2021,25(12):2409–2420. (NIU Wenyuan, HUANG Xianfeng, JIN Jie, et al. Recognition method of the main object of three-dimensional photogrammetric modeling of cultural relics[J]. National Remote Sensing Bulletin, 2021,25(12):2409–2420.) doi:10.11834/jrs.20211185.
- [3] 于野,艾华,贺小军,等. A-FPN 算法及其在遥感图像船舶检测中的应用[J]. 遥感学报, 2020,24(2):107–115. (YU Ye, AI Hua, HE Xiaojun, et al. Attention-based feature pyramid networks for ship detection of optical remote sensing image[J]. National Remote Sensing Bulletin, 2020,24(2):107–115.) doi:10.11834/jrs.20208264.
- [4] LI Changyang, YUAN Yuchen, CAI Weidong, et al. Robust saliency detection via regularized random walks ranking[C]// IEEE Conference on Computer Vision and Pattern Recognition(CVPR). Boston, MA, USA: IEEE, 2015:2710–2717.
- [5] YUAN Yuchen, LI Changyang, KIM Jinman, et al. Reversion correction and regularized random walk ranking for saliency detection[J]. IEEE Transactions on Image Processing, 2017,27(3):1311–1322.
- [6] 刘亚宁,吴清,魏雪. 基于流行排序的前景背景显著性检测算法[J]. 科学技术与工程, 2018,18(18):74–81. (LIU Yaning, WU Qing, WEI Xue. Saliency detection combined foreground with background based on manifold ranking[J]. Science Technology and Engineering, 2018,18(18):74–81.)
- [7] 戴玉超,张静, PORIKLI F, 等. 深度残差网络的多光谱遥感图像显著目标检测[J]. 测绘学报, 2018,47(6):873–881. (DAI Yuchao, ZHANG Jing, PORIKLI F, et al. Salient object detection from multi-spectral remote sensing images with deep residual network[J]. Acta Geodaetica et Cartographica Sinica, 2018,47(6):873–881.) doi:10.11947/j.AGCS.2018.20170633.
- [8] HOU Qibin, CHENG Mingming, HU Xiaowei, et al. Deeply supervised salient object detection with short connections[C]// Proceedings of the IEEE conference on computer vision and pattern recognition. Hawaii, USA: IEEE, 2017:3203–3212.
- [9] LI Chongyi, CONG Runmin, HOU Junhui, et al. Nested network with two-stream pyramid for salient object detection in optical remote sensing images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2019,57(11):9156–9166.
- [10] 王福斌,刘贺飞,王蕊,等. 烧结断面火焰图像多核 Boosting 显著性检测[J]. 计算机辅助设计与图形学学报, 2021,33(9):1466–1474. (WANG Fubin, LIU Hefei, WANG Rui, et al. Multiple kernel boosting saliency detection of flame image of sintering section[J]. Journal of Computer-Aided Design & Computer Graphics, 2021,33(9):1466–1474.) doi:10.3724/SP.J.1089.2021.18686.
- [11] ZHAO Jiaying, LIU Jiangjiang, FAN Dengping, et al. EGNet: Edge Guidance Network for salient object detection[C]// IEEE/CVF International Conference on Computer Vision. Seoul, Korea(South): IEEE, 2019:8779–8788.
- [12] LIU Jiangjiang, HOU Qibin, CHENG Mingming, et al. A simple pooling-based design for real-time salient object detection[C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, CA, USA: IEEE, 2019:3917–3926.
- [13] 陈正,赵晓丽,张佳颖,等. 基于跨模态特征融合的 RGB-D 显著性目标检测[J]. 计算机辅助设计与图形学学报, 2021,33(11):1688–1697. (CHEN Zheng, ZHAO Xiaoli, ZHANG Jiaying, et al. RGB-D image saliency detection based on cross-model feature fusion[J]. Journal of Computer-Aided Design & Computer Graphics, 2021,33(11):1688–1697.) doi:10.3724/SP.J.1089.2021.18710.
- [14] LI Chongyi, CONG Runmin, GUO Chunle, et al. A parallel down-up fusion network for salient object detection in optical remote sensing images[J]. Neurocomputing, 2020(415):411–420.

- [15] ZHANG Qijian, CONG Runmin, LI Chongyi, et al. Dense attention fluid network for salient object detection in optical remote sensing images[J]. IEEE Transactions on Image Processing, 2020(30):1305–1317.
- [16] 侯舒维, 郭宝龙, 李晓博, 等. 城区建筑物的在轨热异常检测提取[J]. 太赫兹科学与电子信息学报, 2022, 20(5):498–505. (HOU Shuwei, GUO Baolong, LI Xiaobo, et al. On-orbit thermal anomaly detection and extraction of urban buildings[J]. Journal of Terahertz Science and Electronic Information Technology, 2022, 20(5):498–505.) doi:10.11805/TKYDA2021023.
- [17] WU Xin, LI Wei, HONG Danfeng, et al. Vehicle detection of multi-source remote sensing data using active fine-tuning network[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2020(167):39–53.
- [18] HAN Junwei, CHEN Hao, LIU Nian, 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.
- [19] CHEN Hao, LI Youfu. Progressively complementarity-aware fusion network for RGB-D salient object detection[C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE, 2018:3051–3060.
- [20] GAO Wei, LIAO Guibiao, MA Siwei, et al. Unified information fusion network for multi-modal RGB-D and RGB-T salient object detection[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 32(4):2091–2106.
- [21] TU Zhengzheng, LI Zhun, LI Chenglong, et al. Multi-interactive dual-decoder for RGB-thermal salient object detection[J]. IEEE Transactions on Image Processing, 2021(30):5678–5691.
- [22] LIANG Fangfang, DUAN Lijuan, MA Wei, et al. A deep multimodal feature learning network for RGB-D salient object detection[J]. Computers & Electrical Engineering, 2021(92):107006.
- [23] ZHANG Qiang, HUANG Nianchang, YAO Lin, et al. RGB-T salient object detection via fusing multi-level CNN features[J]. IEEE Transactions on Image Processing, 2019(29):3321–3335.
- [24] LIAO Guibiao, GAO Wei, LI Ge, 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.
- [25] HUANG Nianchang, YANG Yang, ZHANG Dingwen, et al. Employing bilinear fusion and saliency prior information for RGB-D salient object detection[J]. IEEE Transactions on Multimedia, 2021(24):1651–1664.
- [26] HUANG Nianchang, LIU Yi, ZHANG Qiang, et al. Joint cross-modal and unimodal features for RGB-D salient object detection[J]. IEEE Transactions on Multimedia, 2020(23):2428–2441.
- [27] LIU Zhengyi, TAN Yacheng, HE Qian, 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.
- [28] FAN Dengping, CHENG Mingming, LIU Yun, et al. Structure-measure: a new way to evaluate foreground maps[C]// IEEE International Conference on Computer Vision. Venice, Italy: IEEE, 2017:4548–4557.
- [29] 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.

作者简介:

连远锋(1977-), 男, 博士, 副教授, 主要研究方向为计算机视觉及遥感图像处理技术. email: lianyuanfeng@cup.edu.cn.

石旭(1998-), 男, 在读硕士研究生, 主要研究方向为遥感图像处理及显著性检测.

江澄(1985-), 男, 博士, 高级工程师, 主要研究方向为高光谱遥感建模与仿真、遥感数据预处理及定量应用.