

基于多图神经网络协同学习的显著性物体检测方法

刘冰^① 王甜甜^① 高丽娜^① 徐明珠*^② 付平^①

^①(哈尔滨工业大学电子与信息工程学院 哈尔滨 150000)

^②(山东大学软件学院 济南 250101)

摘要: 目前基于深度卷积神经网络的显著性物体检测方法难以在非欧氏空间不规则结构数据中应用,在复杂视觉场景中易造成显著物体边缘及结构等高频信息损失,影响检测性能。为此,该文面向显著性物体检测任务提出一种端到端的多图神经网络协同学习框架,实现显著性边缘特征与显著性区域特征协同学习的过程。在该学习框架中,该文构造了一种动态信息增强图卷积算子,通过增强不同图节点之间和同一图节点内不同通道之间的信息传递,捕获非欧氏空间全局上下文结构信息,完成显著性边缘信息与显著性区域信息的充分挖掘;进一步地,通过引入注意力感知融合模块,实现显著性边缘信息与显著性区域信息的互补融合,为两种信息挖掘过程提供互补线索。最后,通过显式编码显著性边缘信息,指导显著性区域的特征学习,从而更加精准地定位复杂场景下的显著性区域。在4个公开的基准测试数据集上的实验表明,所提方法优于目前主流的基于深度卷积神经网络的显著性物体检测方法,具有较强的鲁棒性和泛化能力。

关键词: 显著性物体检测; 图神经网络; 注意力感知融合; 协同学习

中图分类号: TN911.73; TP391.41

文献标识码: A

文章编号: 1009-5896(2023)07-2561-10

DOI: [10.11999/JEIT220706](https://doi.org/10.11999/JEIT220706)

Salient Object Detection Based on Multiple Graph Neural Networks Collaborative Learning

LIU Bing^① WANG Tiantian^① GAO Lina^① XU Mingzhu*^② FU Ping^①

^①(School of Electronics and Information Engineering, Harbin Institute of Technology, Harbin 150000, China)

^②(Software College, Shandong University, Jinan 250101, China)

Abstract: In complex visual scene, the performance of existing deep convolutional neural network based methods of salient object detection still suffer from the loss of high-frequency visual information and global structure information of the object, which can be attributed to the weakness of convolutional neural network in capability of learning from the data in non-Euclidean space. To solve these problems, an end-to-end multiple graph neural networks collaborative learning framework is proposed, which realizes the cooperative learning process of salient edge features and salient region features. In this learning framework, this paper constructs a dynamic message enhancement graph convolution operator, which captures non-Euclidean space global context structure information by enhancing message transfer between different graph nodes and between different channels within the same graph node. Further, by introducing an attention perception fusion module, the complementary fusion of salient edge information and salient region information is realized, providing complementary clues for the two information mining processes. Finally, by explicitly encoding the salient edge information to guide the feature learning of salient regions, salient regions in complex scenes can be located more accurately. The experiments on four open benchmark datasets show that the proposed method has strong robustness and generalization ability, which make it superior to the current mainstream deep convolutional neural network based salient object detection methods.

Key words: Salient object detection; Graph neural network; Attention perception fusion; Collaborative learning

收稿日期: 2022-05-31; 改回日期: 2022-12-05; 网络出版: 2022-12-22

*通信作者: 徐明珠 xumingzhu@sdu.edu.cn

基金项目: 国家自然科学基金(62171156)

Foundation Item: The National Natural Science Foundation of China (62171156)

1 引言

人类视觉系统通过对视觉场景内的局部信息进行分析整合,建立对整个视觉场景的理解,进而快速精确地定位到视觉场景中最吸引注意的区域。不同于传统高成本的划窗遍历式搜索^[1]来检测物体,显著性物体检测^[2]通过模拟人类视觉系统,高效定位图像中视觉上最显著的区域,并将其从背景中均匀地分离出来,使得后续视觉算法能够根据区域的显著性程度合理分配计算资源,从而降低处理成本,缓解海量图像信息与有限计算资源之间的矛盾,适用于资源受限的一些应用场合,如卫星^[3]、无人机^[4]等,同时作为预处理步骤,显著性物体检测能够辅助各种高级视觉任务进一步获取视觉场景中的高阶语义信息,从而建立起对视觉场景的局部到整体的理解,如目标识别^[5]、目标跟踪^[6]、图像/视频字幕^[7,8]等。

近年来,基于深度卷积神经网络的显著性物体检测方法得到了广泛的研究,其有效提高了检测的性能,使得显著性物体检测研究取得持续进展。Hou等人^[9]通过在整体嵌套边缘检测器架构内的跨层结构中引入短连接,充分利用了从全卷积神经网络中提取的多尺度特征。Zhang等人^[10]设计了一种对称的全卷积网络,在无损特征反射的指导下有效学习互补显著性特征,此外又提出了一种加权结构损失函数,进一步提高检测性能。张文明等人^[11]在深度卷积神经网络的基础上,采用跨层连接和多尺度融合方式来提高显著性检测精度。Ma等人^[12]提出了一种金字塔特征收缩网络,通过逐层收缩来聚合相邻特征节点,有效融合低层细节信息和高层语义信息。Liu等人^[13]提出了一种能够进行显著目标分割、边缘检测和骨架提取的多任务网络,提升了多个任务的执行效率。然而,目前基于深度卷积神经网络的显著性物体检测模型在处理高频视觉信息及捕获物体全局结构信息方面还不够理想,表现为获取的显著图中存在物体边缘模糊、检测完整性差以及背景被错误凸显等问题,进而影响后续高层视觉分析任务的性能。这种情况主要是由深度卷积神经网络的结构特性影响所致,一方面由于深度卷积神经网络学习方法是设计各向同性滤波器挖掘图像的平移不变性及可分解性,常采用下采样(卷积+池化)降低输入数据的维度、上采样(反卷积)恢复输入数据的原始大小。这种下采样+上采样操作造成物体边缘及结构等高频信息发生不可逆损失,因此会产生物体边缘模糊及检测完整性差等问题。许多学者通过采用边缘感知损失函数^[14]或者以一种端到端的方式学习融合显著性边缘和显著性物体区

域^[15]来解决这个问题。Liu等人^[16]通过两个基于池化的模块来逐步细化高级语义特征,从而生成细节丰富的显著图。Qin等人^[17]设计了一种边界感知分割网络,通过预测和细化分割概率图以及使用一种混合损失函数,实现了较好的图像分割效果。Wei等人^[18]提出了一个由标签解耦过程和特征交互网络构成的标签解耦框架,缓解了边缘像素因分布不平衡而导致的难以预测的问题。Wu等人^[19]提出分解与补全网络,对显著目标的完整性进行建模,取得了较好的性能。然而,针对深度卷积神经网络结构特性所导致的输出显著图中物体边缘模糊以及检测完整性差等问题,以上方法只能在一定程度上予以缓解。

图神经网络^[20]通过定义各向异性滤波器从非欧氏空间不规则的图结构数据中汇聚信息,能够更好地学习、推理物体内部必然存在的隐含的复杂逻辑关系,弥补深度卷积神经网络的不足。因此,本文充分利用图神经网络在捕获全局上下文结构信息方面的优势,设计了一种新颖的多图神经网络协同学习方法,改善显著图中存在的物体边缘模糊、检测完整性差以及背景被错误凸显等问题。具体而言,本文主要贡献如下:

(1) 设计了一个动态信息增强图卷积(Dynamic Message Enhancement Graph Convolution, DMEGC)模块以充分挖掘图节点嵌入信息。该模块利用不同图节点之间隐式存在的多种语义关系和注意力机制,增强了不同图节点之间和同一图节点内部的信息传递,提升了对显著性信息的获取能力。

(2) 设计了一个注意力感知融合(Attention Perception Fusion, APF)模块对显著性边缘特征与显著性区域特征进行互补融合,为两种特征后续学习过程提供互补线索,有效提高网络对显著性信息的挖掘效率。

(3) 以DMEGC和APF为基础,设计了一种面向图模型的多图神经网络协同学习框架,通过边缘图神经网络与区域图神经网络有效学习、推理子区域之间存在的复杂逻辑关系,从而充分挖掘显著性边缘信息与显著性区域信息。与此同时,利用显著性边缘特征与显著性区域特征双方提供的互补线索,完成边缘图节点与区域图节点的信息交互,从而实现对显著性区域的高效精确定位,抑制复杂背景的干扰,提升显著性物体检测任务的准确性和鲁棒性,弥补深度卷积神经网络的关系学习能力不足的问题。在DUTS-TE^[21]、ECSSD^[22]、PASCAL-S^[23]和DUT-OMRON^[24]公开基准数据集上的实验表明,本文方法在主流评估指标上优于目前先进的方法,实现了更好的检测效果。

2 所提方法整体框架

为了解决深度卷积神经网络结构特性所导致的信息挖掘不充分以及关系学习能力不足等问题, 本文提出了一种端到端的基于多图神经网络协同学习的显著性物体检测方法, 该方法整体框架如图1所示。首先, 采用ResNet50作为骨干网络分别从其低层和高层初步提取边缘特征和区域特征; 进一步地, 通过一致差异聚合模块(Consistent Difference Aggregation Module, CDAM)^[25]利用高阶语义信息对边缘特征进行指导优化, 从而获取初始显著性边缘特征, 与此同时, 通过融合高层多尺度区域特征来获取鲁棒的初始显著性区域特征; 对获取的初始显著性边缘特征与初始显著性区域特征分别经过图投影操作^[26], 使他们变换到非欧氏空间, 以获得相应的图结构数据, 即边缘图与区域图; 然后将构建好的边缘图与区域图输入到多图神经网络协同学习框架中进行处理, 在各自采用本文构造的动态信息图卷积层充分挖掘图节点嵌入信息的同时, 通过本文设计的注意力感知融合模块进行有效融合, 引导各自的学习过程更倾向于关注有价值的信息, 从而高效捕获丰富的上下文结构信息; 对经过4层动态信息增强图卷积层信息挖掘处理之后的边缘图与区域图分别经过图重投影操作^[26], 使其恢复到原始的欧氏空间, 通过边缘支持图卷积(Edge Supportive Graph Convolution, ESGC)算子^[27]显式编码显著性边缘信息, 进一步指导显著性区域的特征学习,

从而更加准确地定位复杂场景下的显著性区域; 将得到充分有效增强和优化的显著性区域特征经过全连接层处理之后, 生成最终的显著性预测结果。

3 基于多图神经网络协同学习的显著性物体检测方法

3.1 获取初始显著性特征

3.1.1 获取初始显著性边缘特征

考虑到网络的高层特征主要包含抽象的语义信息, 而低层特征主要包含更多的细节信息, 为捕获显著性边缘特征, 本文利用高层的高阶语义信息以一致差异聚合的方式跨层指导边缘特征的学习, 从而捕获包含部分全局语义信息和局部边缘信息的初始显著性边缘特征。

为了进一步更加准确地提取初始显著性边缘特征, 本文对其进行显式建模, 在训练整个网络时采用边缘真值图对一致差异聚合模块输出的显著性边缘特征进行监督学习, 本文所采用的损失函数为二元交叉熵损失函数, 在这里定义为

$$L^e(\mathbf{E}_{\text{initial}}, \mathbf{G}^e) = -\frac{1}{N} \sum_{i=1}^N \{g_i^e \ln(e_i) + (1 - g_i^e) \ln(1 - e_i)\} \quad (1)$$

其中, N 表示总像素数, $\mathbf{E}_{\text{initial}}$ 表示一致差异聚合模块输出的显著性边缘预测图, $\mathbf{G}^e \in \{0, 1\}$ 表示显著性边缘真值图, g_i^e 和 s_i 分别表示显著性边缘真值图和显著性边缘预测图的第 i 个像素值。

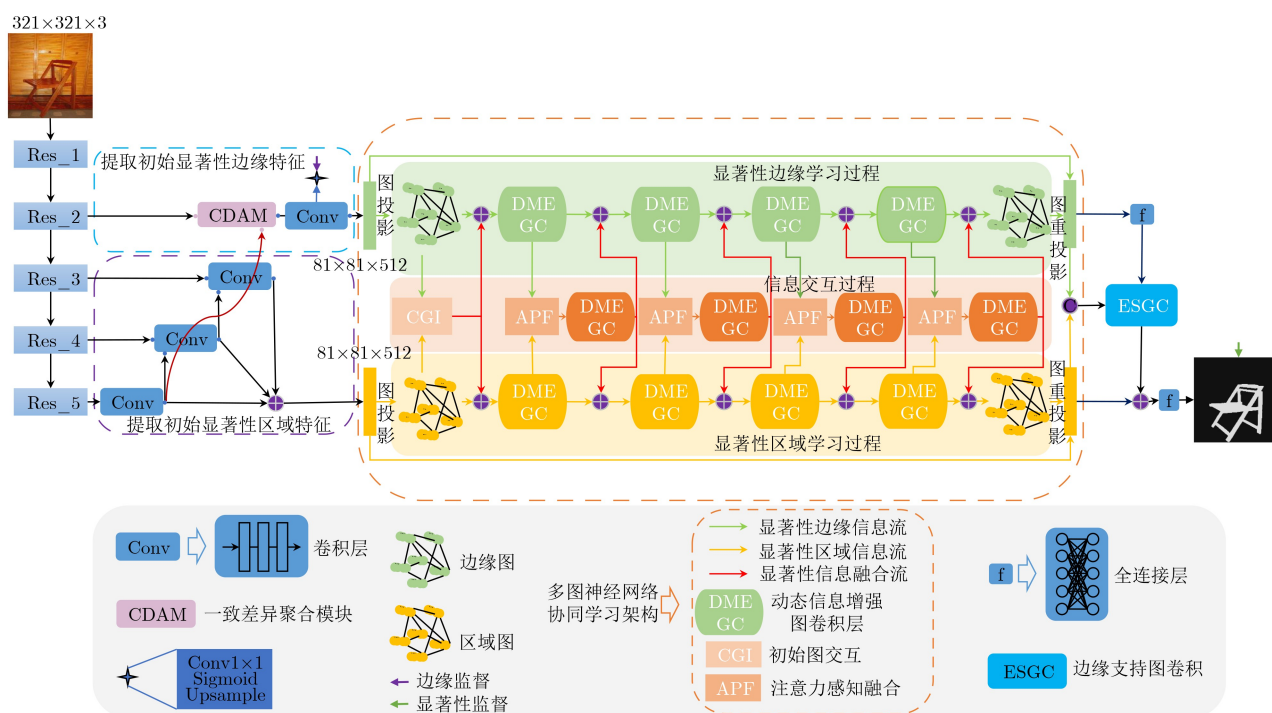


图1 所提方法整体框架

3.1.2 获取初始显著性区域特征

为了获得鲁棒的显著性区域特征、减少噪声的干扰,本文在广泛使用的U-Net框架的基础上,将骨干网络ResNet50的Res_1、Res_2和Res_3输出的多尺度特征分别通过额外添加的3层卷积层进行特征增强,然后采用ReLU激活函数以保证非线性;增强之后的多尺度特征通过相加融合,获得最终的初始显著性区域特征,具体如图1中紫色虚线框所示。

3.2 构建图结构数据

给定输入的显著性边缘特征 $F_e \in \mathbf{R}^{H \times W \times C}$ 和显著性区域特征 $F_r \in \mathbf{R}^{H \times W \times C}$, 首先使用Conv1×1对他们进行降维处理,转换成低维特征 $F_e^l \in \mathbf{R}^{(H \times W) \times C}$ 和 $F_r^l \in \mathbf{R}^{(H \times W) \times C}$; 然后将这两种特征分别通过图投影操作 $f_{G_{proj}}(\cdot)$ 转化为相应的边缘图节点 $V_e \in \mathbf{R}^{C \times N}$ 和区域图节点 $V_r \in \mathbf{R}^{C \times N}$, 即

$$V_e = f_{G_{proj}}(F_e^l) \in \mathbf{R}^{C \times N}, V_r = f_{G_{proj}}(F_r^l) \in \mathbf{R}^{C \times N} \quad (2)$$

通过上述方式,边缘图 $G_e = (V_e, E_e)$ 与区域图 $G_r = (V_r, E_r)$ 构建完成,其中, $V_e = \{v_1^e, v_2^e, \dots, v_N^e\}$ (或 $V_r = \{v_1^r, v_2^r, \dots, v_N^r\}$) 表示图节点集合, E_e (或 E_r) 表示图边集合。

3.3 多图神经网络协同学习

为了充分利用显著性边缘特征与显著性区域特征之间的协同关系来辅助显著性物体检测任务,本文设计了一种多图神经网络协同学习框架。

首先,将边缘图 G_e 与区域图 G_r 进行初始图交互,以捕获他们之间的高阶依赖关系,初步实现图间信息传递,如图2所示。具体来说,首先计算边缘图节点 $V_e \in \mathbf{R}^{C \times N}$ 与区域图节点 $V_r \in \mathbf{R}^{C \times N}$ 之间的相似性矩阵 A^{inter} , 定义为

$$A^{inter} = f_{norm}(V_e^T \times V_r) \in \mathbf{R}^{N \times N} \quad (3)$$

其中, $f_{norm}(\cdot)$ 表示归一化处理操作;然后将初始图交互信息分别传递给边缘图节点 $V_e \in \mathbf{R}^{C \times N}$ 与区域图节点 $V_r \in \mathbf{R}^{C \times N}$, 详细的定义为

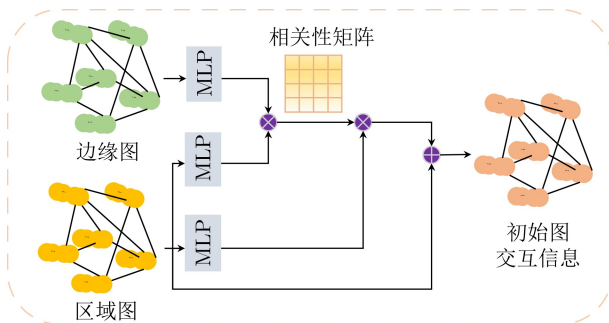


图2 初始图交互示意图

$$\left. \begin{aligned} V_e' &= \lambda(A^{inter} \times V_r^T) + V_e \in \mathbf{R}^{C \times N} \\ V_r' &= \lambda(A^{inter} \times V_e^T) + V_r \in \mathbf{R}^{C \times N} \end{aligned} \right\} \quad (4)$$

其中, λ 是平衡图交互信息重要性的权衡参数。

为了实现图节点嵌入信息的充分有效挖掘,将交互之后的边缘图 G_e 与区域图 G_r 同时送入本文构造的动态信息增强图卷积层。在动态信息增强图卷积层构造的过程中,本文通过考虑不同图节点之间隐式存在的多种语义关系,增强不同图节点之间的信息传递,同时对于每一个图节点而言,采用基于通道的注意力机制,增强其内部不同通道之间的信息传递。动态信息增强图卷积作为图神经网络每层的基本图卷积单元,图数据结构在图神经网络的每层中得到动态优化更新,并在每层中实现图节点嵌入信息的充分挖掘,即同时挖掘当前图节点与其邻居节点之间隐式存在的多种语义关系,增强图节点之间的信息交互,并通过通道注意力机制完成同一图节点内不同通道之间的信息传递,具体结构如图3所示。

具体来说,首先利用欧氏距离计算每个图节点的最近 k 个邻居节点;然后通过边学习函数 $f_e^{r_{learn}}(\cdot)$ 来学习当前图节点 v_i 与其邻居节点 $v_j, j \in N_i^k$ 之间的边权值大小 $e_{i,j}^{r_{learn}}$, 以充分挖掘当前图节点与其邻居节点之间隐式存在的多种语义关系,增强不同图节点之间的信息交互,具体计算过程

$$e_{i,j}^{r_{learn}} = f_e^{r_{learn}}(v_i, v_j), j \in N_i^k, r_{learn} \in [1, R] \quad (5)$$

其中, r_{learn} 表示关系类型索引, R 表示待学习的关系总数, N_i^k 表示当前图节点 v_i 的邻居节点集合,边学习函数 $f_e^{r_{learn}}(\cdot)$ 可以采用多层感知机MLP来拟合学习,定义为

$$f_e^{r_{learn}}(v_i, v_j) = \text{MLP}^{r_{learn}}(\text{concat}(v_i, v_j - v_i)), \quad r_{learn} \in [1, R] \quad (6)$$

利用来自多个可学习边的信息更新当前图节点 v_i 的嵌入信息,即

$$v_i' = \max_{j \in N_i^k} \sum_{r_{learn}=1}^R e_{i,j}^{r_{learn}} \quad (7)$$

最后,为了实现同一图节点内不同通道之间的信息传递,计算图节点之间基于通道的注意力归一化系数矩阵,具体可以表示为

$$A = \text{softmax}(V'^T V') \in \mathbf{R}^{C \times C} \quad (8)$$

其中, $\text{softmax}(\cdot)$ 操作对基于通道的注意力系数矩阵 $V'^T V'$ 的结果按列进行归一化,每列中的元素表示从一个通道到其他通道的相关系数。因此,同一图

节点内不同通道之间的信息传递可以通过下式完成

$$\mathbf{V}^{\text{intra}} = \mathbf{V}' \times \mathbf{A} \in \mathbf{R}^{N \times C} \quad (9)$$

另外，为了提高同一图节点内部特征的稳定性，采用残差结构来结合原始图节点特征 \mathbf{V}' 和 $\mathbf{V}^{\text{intra}}$ ，最终输出的图节点特征

$$\hat{\mathbf{V}} = \alpha \times \mathbf{V}^{\text{intra}} + \mathbf{V}' \in \mathbf{R}^{N \times C} \quad (10)$$

其中， α 控制 $\mathbf{V}^{\text{intra}}$ 对原始图节点特征的影响。

对边缘图节点 $\hat{\mathbf{V}}_e \in \mathbf{R}^{N \times C}$ 与区域图节点 $\hat{\mathbf{V}}_r \in \mathbf{R}^{N \times C}$ 的有效融合可以为显著性边缘特征与显著性区域特征的学习过程提供互补线索。例如，可以通过引入显著性区域线索为显著性边缘特征学习过程提供语义信息，也可以通过引入显著性边缘线索为显著性区域特征学习过程提供细节信息。为此，本文设计了一种注意力感知融合模块，其自适应地为不同图节点分配不同的权重，并融合显著性边缘特征与显著性区域特征以实现空间信息的编码。因此，将经过动态信息增强图卷积层处理之后的边缘图节点 $\hat{\mathbf{V}}_e \in \mathbf{R}^{N \times C}$ 与区域图节点 $\hat{\mathbf{V}}_r \in \mathbf{R}^{N \times C}$ 同时送入本文设计的注意力感知融合模块，实现对两种特征的互补融合，为显著性边缘特征与显著性区域特征的学习过程提供互补线索，如图4所示。

具体地，首先将边缘图节点 $\hat{\mathbf{V}}_e \in \mathbf{R}^{N \times C}$ 与区域图节点 $\hat{\mathbf{V}}_r \in \mathbf{R}^{N \times C}$ 按照对应图节点逐通道相加得到初始融合图节点 $\hat{\mathbf{V}}^F \in \mathbf{R}^{N \times C}$ ，计算方法为

$$\hat{\mathbf{V}}^F = (\hat{\mathbf{V}}_e + \hat{\mathbf{V}}_r) \in \mathbf{R}^{N \times C} \quad (11)$$

其次，将初始融合图节点 $\hat{\mathbf{V}}^F \in \mathbf{R}^{N \times C}$ 输入到一个由两个线性层组成的注意力子模块，融合全局节点信息，生成逐节点的注意力权值，即

$$\text{Att} = \sigma(\text{ReLU}(\hat{\mathbf{V}}^F \times \mathbf{W}_1) \times \mathbf{W}_2) \in \mathbf{R}^{N \times 1} \quad (12)$$

其中， $\mathbf{W}_1 \in \mathbf{R}^{C \times C}$ 和 $\mathbf{W}_2 \in \mathbf{R}^{C \times 1}$ 是两个线性层共享的可学习参数；然后，通过为不同图节点分配不同的注意力权值以获得新的融合图节点 $\hat{\mathbf{V}}^{F'} \in \mathbf{R}^{N \times C}$ ，具体计算为

$$\hat{\mathbf{V}}^{F'} = (\hat{\mathbf{V}}^F \otimes \text{Att}) \times \mathbf{W}_3 \in \mathbf{R}^{N \times C} \quad (13)$$

其中， \otimes 表示逐行的加权乘法操作，对加权之后的图节点进行一次线性变换 $\mathbf{W}_3 \in \mathbf{R}^{C \times C}$ ，生成更具判别性的图节点嵌入信息；最后，将获得的新的融合图节点 $\hat{\mathbf{V}}^{F'} \in \mathbf{R}^{N \times C}$ 分别与输入的边缘图节点 $\hat{\mathbf{V}}_e \in \mathbf{R}^{N \times C}$ 和区域图节点 $\hat{\mathbf{V}}_r \in \mathbf{R}^{N \times C}$ 进行相加融合，送入到下一个动态信息增强图卷积层，以相同方式实现对两种图节点嵌入信息的进一步挖掘与协同优化。

3.4 损失函数

在训练过程中，本文选用二元交叉熵损失作为损失函数，结合对初始显著性边缘特征监督训练的损失函数 L^e ，整体损失函数 L_{total} 定义为

$$L_{\text{total}} = L^e(\mathbf{E}_{\text{initial}}, \mathbf{G}^e) + L^s(\mathbf{S}_{\text{final}}, \mathbf{G}^s) \quad (14)$$

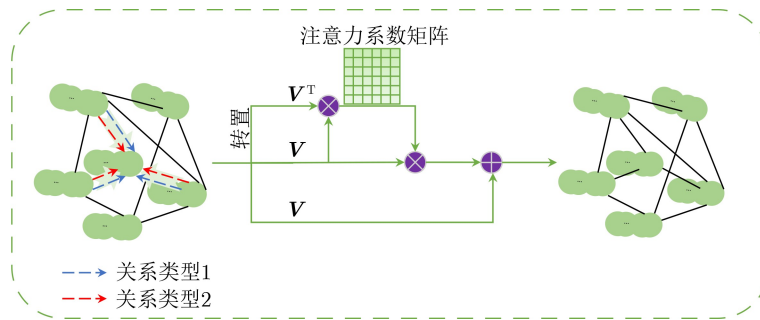


图3 动态信息增强图卷积模块

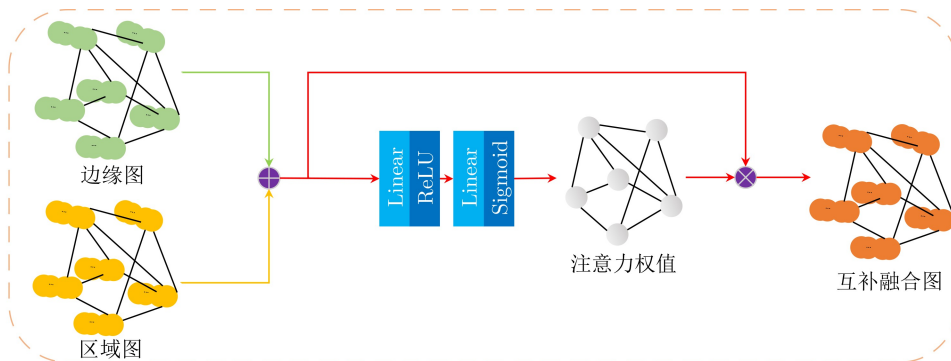


图4 注意力感知融合模块

其中, E_{initial} 表示一致差异聚合模块输出的显著性边缘预测图, $G^e \in \{0, 1\}$ 表示显著性边缘真值图, S_{final} 表示模型最终输出的显著性预测图, $G^s \in \{0, 1\}$ 表示显著性真值图。

4 实验结果与分析

4.1 数据集和评价指标

本文采用4个公开的面向显著性物体检测任务的基准数据集进行实验, 分别是DUTS-TE, ECSSD, PASCAL-S和DUT-OMRON, 这4个数据集文件中均含有手工标记的真值图。本文使用4种常用的量化评价指标来评估提出的方法, 分别是精度-召回率(Precision-Recall, P-R)曲线、S-measure(Structure similarity measure, S_α)、加权F-measure(weighted F-measure, F_β^ω)、平均绝对误差(Mean Absolute Error, MAE)。

4.2 实验设置

本文方法是基于PyTorch深度学习框架, 通过PyTorch Geometric机器学习库, 在NVIDIA TITAN RTX(24 GB)上实现。本文采用ResNet50预训练模型, 在DUTS-TR训练数据集^[22]上完成训练。在图结构数据的构建过程中, 为了选取合适的图节点数目 N 和邻居节点数目 k , 本文使用DUTS-TR训练数据集在不同的参数值设置下训练了一组

网络变量, 并在 S_α , F_β^ω 和MAE指标下完成性能测试, 如表1所示。

综合考虑不同参数设置下的评价指标值, 本文选取的图节点数目 N 和邻居节点数目 k 分别为32和8。

4.3 性能比较

为了保证与现有显著性物体检测方法比较的公平性, 所有竞争方法的实验结果均由原论文作者提供, 或者是由使用推荐参数设置的开源代码生成。其中, 参与对比的8种基于卷积神经网络的显著性物体检测方法包括PoolNet^[16], BASNet^[17], EGNet^[15], AFNet^[14], DFI^[13], LDF^[18], PFSNet^[12]和DCN^[19]。

4.3.1 定量比较

图5展示了不同方法与对比方法的P-R曲线图。从图5的P-R曲线可以看出, 与其他显著性物体检测方法相比, 当召回率较高时, 本文方法在大多数情况下具有更高的准确率。例如, 在DUTS-TE测试数据集上, 当召回率为0.8时, 本文方法获得了更高的准确率, 约为97%。另外, 本文方法在场景复杂的PASCAL-S数据集和DUT-OMRON数据集上的整体性能均优于其他对比方法, 进一步说明本文方法具有较强的泛化能力, 适用于更多复杂的场景。表2展示了本文方法与现有的8种先进的方法分别在指标 S_α , F_β^ω 和MAE上的比较结果。其

表1 参数 N 和 k 在不同设置下的性能结果

	ECSSD			PASCAL-S		
	$S_\alpha(\uparrow)$	$F_\beta^\omega(\uparrow)$	MAE(\downarrow)	$S_\alpha(\uparrow)$	$F_\beta^\omega(\uparrow)$	MAE(\downarrow)
本文 ($N = 48, k = 8$)	0.933	0.924	0.024	0.881	0.841	0.047
本文 ($N = 32, k = 8$)	0.932	0.926	0.024	0.886	0.850	0.047
本文 ($N = 16, k = 8$)	0.929	0.923	0.028	0.879	0.830	0.057
本文 ($N = 32, k = 16$)	0.932	0.928	0.024	0.879	0.851	0.047
本文 ($N = 32, k = 8$)	0.932	0.926	0.024	0.886	0.850	0.047
本文 ($N = 32, k = 4$)	0.931	0.922	0.026	0.875	0.833	0.052

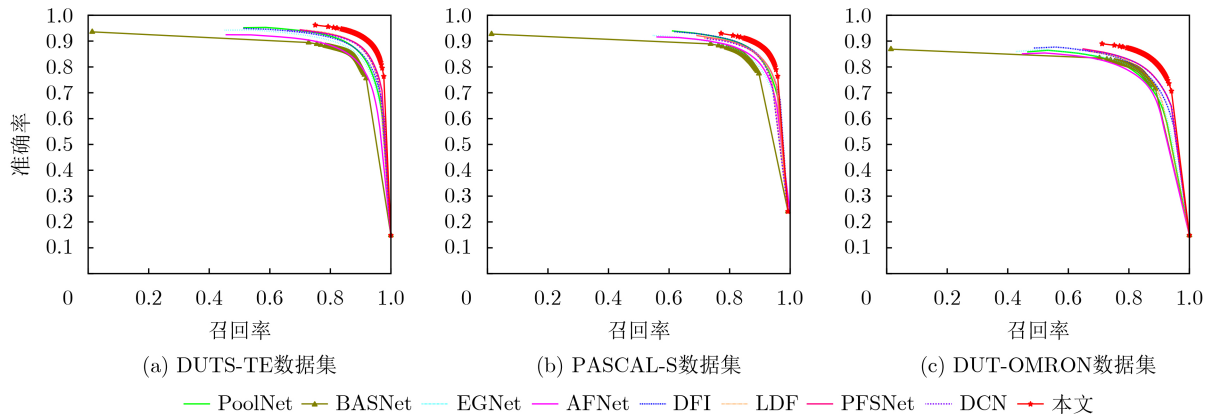


图5 9种方法在3个标准数据集上的P-R曲线图

中，最好的前3个性能指标分别用红色、蓝色、绿色表示。从表2的3个量化指标可以看出，本文方法在4个公开基准测试数据集上的所有性能指标均明显优于对比方法，尤其是在 S_α 和MAE这两项指标上，说明本文方法的输出显著图具有优秀的结构相似度和精度。例如，在PASCAL-S测试数据集上，与排名第2的方法相比，本文方法在 S_α 、 F_β^ω 和MAE上分别获得了约1.7%，2.7%和16.4%的性能提升。虽然本文方法在DUTS-TE测试数据集上的 F_β^ω 指标值排名第2，这可能是因为DUTS-TE数据集的多张图片中存在多个显著性物体且背景结构

复杂，先进的基于卷积神经网络的方法可以产生更多的判别性特征，而这些特征在从复杂的背景中分离显著性物体或者区分多个显著性物体方面比结构信息发挥更重要的作用，但是本文方法在 S_α 和MAE指标上是最佳的，尤其是 S_α 指标，进一步说明本文方法在捕获丰富的上下文结构信息方面更具有优势。

4.3.2 定性比较

图6展示了本文方法与8种对比方法在一些具有代表性的测试图像上检测结果的视觉比较样例，以更加直观的方式展示本文方法与对比方法的性能比

表2 9种方法在4个标准数据集上的 S_α 、 F_β^ω 和MAE指标

方法	DUTS-TE			ECSSD			PASCAL-S			DUT-OMRON		
	$S_\alpha(\uparrow)$	$F_\beta^\omega(\uparrow)$	MAE(\downarrow)	$S_\alpha(\uparrow)$	$F_\beta^\omega(\uparrow)$	MAE(\downarrow)	$S_\alpha(\uparrow)$	$F_\beta^\omega(\uparrow)$	MAE(\downarrow)	$S_\alpha(\uparrow)$	$F_\beta^\omega(\uparrow)$	MAE(\downarrow)
PoolNet	0.883	0.807	0.040	0.921	0.896	0.039	0.851	0.799	0.075	0.836	0.729	0.055
BASNet	0.866	0.803	0.048	0.916	0.904	0.037	0.836	0.795	0.077	0.836	0.751	0.057
EGNet	0.879	0.798	0.044	0.919	0.892	0.041	0.847	0.791	0.078	0.836	0.727	0.056
AFNet	0.867	0.785	0.046	0.914	0.887	0.042	0.850	0.797	0.071	0.826	0.717	0.057
DFI	0.886	0.817	0.039	0.927	0.906	0.035	0.866	0.819	0.065	0.839	0.736	0.055
LDF	0.892	0.845	0.034	0.924	0.915	0.034	0.862	0.825	0.061	0.839	0.751	0.052
PFSNet	0.900	0.898	0.036	0.927	0.912	0.031	0.844	0.791	0.063	0.802	0.743	0.055
DCN	0.892	0.840	0.035	0.928	0.920	0.032	0.862	0.825	0.062	0.845	0.760	0.051
本文	0.920	0.893	0.027	0.932	0.926	0.024	0.886	0.850	0.047	0.867	0.807	0.048

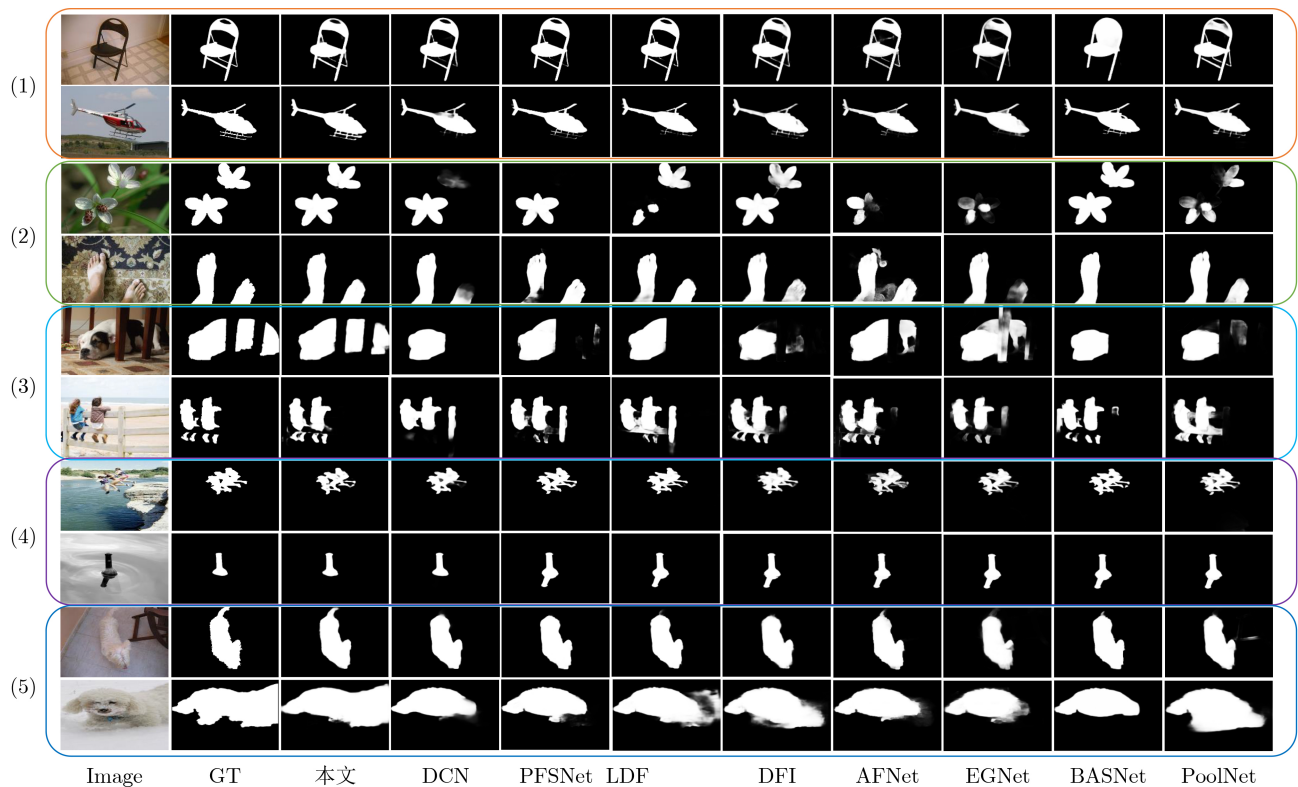


图6 视觉比较结果

较。其中包括一些极具挑战性的复杂情况，(1)显著性物体内存在背景区域；(2)视觉场景中存在多个显著性物体；(3)显著性物体被遮挡；(4)背景复杂且显著性物体形状小；(5)前景与背景对比度低。

如图6所示，从这些复杂场景下的显著图视觉样例可以看出，与本文提出的方法相比，大多数现有的先进方法都没能正确地分割显著性物体，存在显著性物体边缘模糊、检测完整性差以及复杂背景被错误凸显等问题。然而本文方法可以准确预测具有挑战性的复杂场景中最显著的物体和连贯的细节，例如，显著性物体被遮挡(第5、第6行)、背景中存在显著性物体的倒影(第8行)以及低对比度(第9、第10行)。这些可视化对比结果进一步证明了本文方法优秀的显著性物体检测能力，验证了本文方

法在具有挑战性的数据集上的鲁棒性和泛化能力。

4.4 消融实验

为了探讨考虑图节点之间不同关系类型数量对显著性物体检测性能的影响，以及证明本文所提出的动态信息增强图卷积模块和注意力感知融合模块的有效性，本小节在ECSSD数据集上分别进行消融实验。

图7和表3分别展示了在ECSSD数据集上的消融分析结果。其中，AMF表示注意力感知融合模块；DMEGC表示动态信息增强图卷积模块；B表示基线模型，其图卷积单元由基础的图卷积算子GCN^[20]代替，且不包括可学习的注意力感知融合模块； R 表示动态信息增强图卷积模块中待学习的关系类型数量。

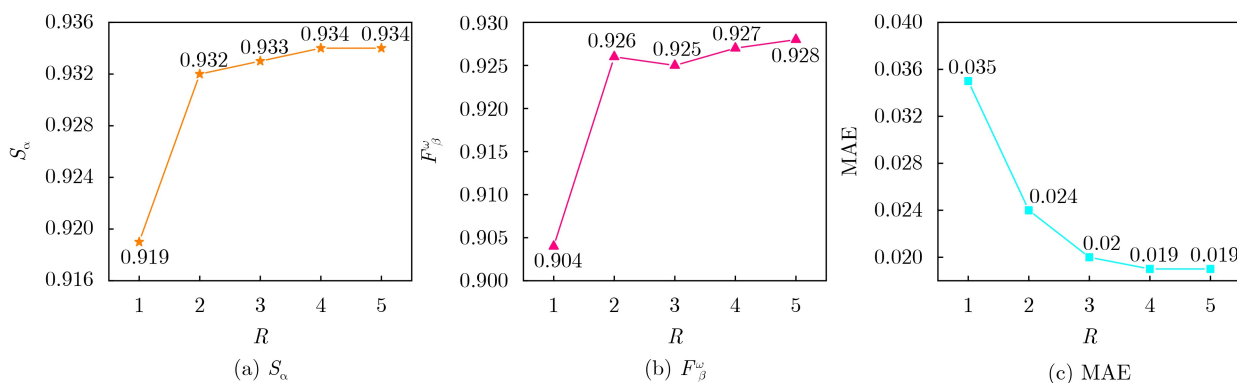


图7 不同关系类型数量对性能指标 S_α 、 F_β^ω 和MAE的影响曲线

首先，本小节通过实验验证了动态信息增强图卷积模块中不同关系类型数量对显著性物体检测性能的影响，如图7所示。可以观察到，关系类型数量取2, 3, 4, 5时，实现了近似的性能结果，并且相比于只考虑一种关系类型的情况，性能指标有了大幅度的提升，说明融合丰富的关系类型有利于检测性能的提升。综合考虑计算成本和检测精度，在本实验中将关系类型数量 R 设置为2。

从表3可以发现，本文所提出的动态信息增强图卷积模块代替基线模型中的图卷积模块之后，使得评估指标 S_α 、 F_β^ω 和MAE分别实现了21.0%，21.4%和76.9%的性能提升，验证了本文所提图卷积模块对图节点嵌入信息充分挖掘的有效性；进一步

地，引入注意力感知融合模块，实现两种图节点嵌入信息的互补融合，为两种学习过程提供互补的线索，最终得到的模型在评估指标 S_α 、 F_β^ω 和MAE上分别实现了2.4%，3.5%和37.5%的性能提升，说明注意力感知模块对信息挖掘过程起到了良好的推动作用。

5 结束语

本文提出了一种多图神经网络协同学习的显著性物体检测方法，相对于基于深度卷积神经网络的显著性物体检测方法，本文利用图神经网络在捕获全局上下文结构信息方面的优势，充分考虑上下文潜在的多种语义关系，对信息进行充分挖掘，解决了目前大多数基于深度卷积神经网络的显著性物体检测模型中不可逆结构导致的物体边缘模糊、检测完整性差以及背景被错误凸显等问题。本文创新性设计的多图神经网络协同学习框架在分别采用基于动态信息增强图卷积单元的图神经网络学习显著性边缘特征与显著性区域特征的同时，通过可学习的注意力感知融合模块完成两种特征的互补融合，为

表3 不同模块的性能影响

模块			ECSSD		
B	DMEGC($R=2$)	AMF	$S_\alpha(\uparrow)$	$F_\beta^\omega(\uparrow)$	MAE(\downarrow)
✓			0.725	0.708	0.189
✓	✓		0.911	0.898	0.040
✓	✓	✓	0.932	0.926	0.024

两种学习过程提供互补线索, 使得信息得到充分挖掘的同时实现信息交互最大化。为了实现复杂场景下的显著性区域进一步准确定位, 本文在显著性区域特征上显式地编码显著性边缘信息。通过在公开基准测试数据集上与8种先进方法进行对比实验, 本文提出的方法取得了最佳的检测结果, 实验结果表明, 基于多图神经网络协同学习框架在非欧氏空间下进行显著性物体检测是一种有效的方法。

参考文献

- [1] VAN DE SANDE K E A, UIJLINGS J R R, GEVERS T, *et al.* Segmentation as selective search for object recognition[C]. 2011 International Conference on Computer Vision, Barcelona, Spain, 2011: 1879–1886. doi: [10.1109/ICCV.2011.6126456](https://doi.org/10.1109/ICCV.2011.6126456).
- [2] BORJI A, CHENG Mingming, HOU Qibin, *et al.* Saliency object detection: A survey[J]. *Computational Visual Media*, 2019, 5(2): 117–150. doi: [10.48550/arXiv.1411.5878](https://doi.org/10.48550/arXiv.1411.5878).
- [3] 刘桂池. 基于显著性和稀疏表示学习的光学遥感图像目标检测与分类[D]. [博士学位论文], 郑州大学, 2020. doi: [10.27466/d.cnki.gzzdu.2020.000029](https://doi.org/10.27466/d.cnki.gzzdu.2020.000029).
LIU Guichi. Saliency and sparse representation learning based optical remote sensing image object detection and classification[D]. [Ph. D. dissertation], Zhengzhou University, 2020. doi: [10.27466/d.cnki.gzzdu.2020.000029](https://doi.org/10.27466/d.cnki.gzzdu.2020.000029).
- [4] 康凯, 杨磊, 李红艳. 基于视觉显著性的小型无人机目标检测方法[J]. *光学与光电技术*, 2020, 18(3): 40–44. doi: [10.19519/j.cnki.1672-3392.2020.03.008](https://doi.org/10.19519/j.cnki.1672-3392.2020.03.008).
KANG Kai, YANG Lei, and LI Hongyan. A novel small UAV detection method based on visual saliency[J]. *Optics & Optoelectronic Technology*, 2020, 18(3): 40–44. doi: [10.19519/j.cnki.1672-3392.2020.03.008](https://doi.org/10.19519/j.cnki.1672-3392.2020.03.008).
- [5] REN Zhixiang, GAO Shenghua, CHIA L T, *et al.* Region-based saliency detection and its application in object recognition[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2014, 24(5): 769–779. doi: [10.1109/TCSVT.2013.2280096](https://doi.org/10.1109/TCSVT.2013.2280096).
- [6] FENG Wei, HAN Ruize, GUO Qing, *et al.* Dynamic saliency-aware regularization for correlation filter-based object tracking[J]. *IEEE Transactions on Image Processing*, 2019, 28(7): 3232–3245. doi: [10.1109/TIP.2019.2895411](https://doi.org/10.1109/TIP.2019.2895411).
- [7] RAMANISHKA V, DAS A, ZHANG Jianming, *et al.* Top-down visual saliency guided by captions[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017: 3135–3144. doi: [10.1109/CVPR.2017.334](https://doi.org/10.1109/CVPR.2017.334).
- [8] ZHOU Lian, ZHANG Yuejie, JIANG Yugang, *et al.* Re-caption: Saliency-enhanced image captioning through two-phase learning[J]. *IEEE Transactions on Image Processing*, 2020, 29: 694–709. doi: [10.1109/TIP.2019.2928144](https://doi.org/10.1109/TIP.2019.2928144).
- [9] HOU Qibin, CHENG Mingming, HU Xiaowei, *et al.* Deeply supervised salient object detection with short connections[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019, 41(4): 815–828. doi: [10.1109/TPAMI.2018.2815688](https://doi.org/10.1109/TPAMI.2018.2815688).
- [10] ZHANG Pingping, LIU Wei, LU Huchuan, *et al.* Saliency object detection with lossless feature reflection and weighted structural loss[J]. *IEEE Transactions on Image Processing*, 2019, 28(6): 3048–3060. doi: [10.1109/TIP.2019.2893535](https://doi.org/10.1109/TIP.2019.2893535).
- [11] 张文明, 姚振飞, 高雅昆, 等. 一种平衡准确性以及高效性的显著性目标检测深度卷积神经网络模型[J]. *电子与信息学报*, 2020, 42(5): 1201–1208. doi: [10.11999/JEIT190229](https://doi.org/10.11999/JEIT190229).
ZHANG Wenming, YAO Zhenfei, GAO Yakun, *et al.* A deep convolutional network for saliency object detection with balanced accuracy and high efficiency[J]. *Journal of Electronics & Information Technology*, 2020, 42(5): 1201–1208. doi: [10.11999/JEIT190229](https://doi.org/10.11999/JEIT190229).
- [12] MA Mingcan, XIA Changqun, and LI Jia. Pyramidal feature shrinking for salient object detection[C]. The Thirty-Fifth AAAI Conference on Artificial Intelligence, Palo Alto, USA, 2021: 2311–2318. doi: [10.1609/aaai.v35i3.16331](https://doi.org/10.1609/aaai.v35i3.16331).
- [13] LIU Jiangjiang, HOU Qibin, and CHENG Mingming. Dynamic feature integration for simultaneous detection of salient object, edge, and skeleton[J]. *IEEE Transactions on Image Processing*, 2020, 29: 8652–8667. doi: [10.1109/TIP.2020.3017352](https://doi.org/10.1109/TIP.2020.3017352).
- [14] FENG Mengyang, LU Huchuan, and DING Errui. Attentive feedback network for boundary-aware salient object detection[C]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, USA, 2019: 1623–1632. doi: [10.1109/CVPR.2019.00172](https://doi.org/10.1109/CVPR.2019.00172).
- [15] ZHAO Jiaying, LIU Jiangjiang, FAN Dengping, *et al.* EGNet: Edge guidance network for salient object detection[C]. 2019 IEEE/CVF International Conference on Computer Vision, Seoul, Korea (South), 2019: 8778–8787. doi: [10.1109/ICCV.2019.00887](https://doi.org/10.1109/ICCV.2019.00887).
- [16] LIU Jiangjiang, HOU Qibin, CHENG Mingming, *et al.* A simple pooling-based design for real-time salient object detection[C]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, USA, 2019: 3912–3921. doi: [10.1109/CVPR.2019.00404](https://doi.org/10.1109/CVPR.2019.00404).
- [17] QIN Xuebin, FAN Dengping, HUANG Chenyang, *et al.* Boundary-aware segmentation network for mobile and web applications[EB/OL]. <https://arxiv.org/abs/2101.04704>, 2021.
- [18] WEI Jun, WANG Shuhui, WU Zhe, *et al.* Label decoupling framework for salient object detection[C]. 2020 IEEE/CVF

- Conference on Computer Vision and Pattern Recognition, Seattle, USA, 2020: 13022–13031. doi: [10.1109/CVPR42600.2020.01304](https://doi.org/10.1109/CVPR42600.2020.01304).
- [19] WU Zhe, SU Li, and HUANG Qingming. Decomposition and completion network for salient object detection[J]. *IEEE Transactions on Image Processing*, 2021, 30: 6226–6239. doi: [10.1109/TIP.2021.3093380](https://doi.org/10.1109/TIP.2021.3093380).
- [20] KIPF T N and WELING M. Semi-supervised classification with graph convolutional networks[C]. The 5th International Conference on Learning Representations, Toulon, France, 2017.
- [21] WANG Lijun, LU Huchuan, WANG Yifan, *et al.* Learning to detect salient objects with image-level supervision[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017: 3796–3805. doi: [10.1109/CVPR.2017.404](https://doi.org/10.1109/CVPR.2017.404).
- [22] SHI Jianping, YAN Qiong, LI Xu, *et al.* Hierarchical image saliency detection on extended CSSD[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016, 38(4): 717–729. doi: [10.1109/TPAMI.2015.2465960](https://doi.org/10.1109/TPAMI.2015.2465960).
- [23] LI Yin, HOU Xiaodi, KOCH C, *et al.* The secrets of salient object segmentation[C]. 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, USA, 2014: 280–287. doi: [10.1109/CVPR.2014.43](https://doi.org/10.1109/CVPR.2014.43).
- [24] LI Cuiping, CHEN Zhenxue, LIU Chengyun, *et al.* Saliency detection: Multi-level combination approach via graph-based manifold ranking[C]. 2017 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery, Guilin, China, 2017: 604–609. doi: [10.1109/FSKD.2017.8393339](https://doi.org/10.1109/FSKD.2017.8393339).
- [25] ZHAO Xiaoqi, PANG Youwei, ZHANG Lihe, *et al.* Self-supervised pretraining for RGB-D salient object detection[C]. The Thirty-Sixth AAAI Conference on Artificial Intelligence, Palo Alto, USA, 2022: 3463–3471. doi: [10.1609/aaai.v36i3.20257](https://doi.org/10.1609/aaai.v36i3.20257).
- [26] LI Yin and GUPTA A. Beyond grids: Learning graph representations for visual recognition[C]. The 32nd International Conference on Neural Information Processing Systems, Montréal, Canada, 2018: 9245–9255.
- [27] ZHAI Qiang, LI Xin, YANG Fan, *et al.* Mutual graph learning for camouflaged object detection[C]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, USA, 2021: 12992–13002. doi: [10.1109/CVPR46437.2021.01280](https://doi.org/10.1109/CVPR46437.2021.01280).
- 刘 冰: 男, 副教授, 研究方向为机器学习与图像处理、嵌入式人工智能。
- 王甜甜: 女, 硕士生, 研究方向为计算机视觉。
- 高丽娜: 女, 博士生, 研究方向为计算机视觉。
- 徐明珠: 男, 助理研究员, 研究方向为多媒体信息计算、计算机视觉。
- 付 平: 男, 教授, 研究方向为机器学习与图像处理、信息检测与处理。

责任编辑: 马秀强