

DOI:10.11918/202109111

# 高效通道注意力和特征融合的协同显著性检测算法

张德华<sup>1</sup>, 李俊豪<sup>1</sup>, 张静凯<sup>2</sup>, 肖启阳<sup>1</sup>

(1. 河南大学 人工智能学院, 郑州 450046; 2. 河南大学 迈阿密学院, 河南 开封 475004)

**摘要:** 针对现有的协同显著性检测算法在多显著目标复杂场景下表现不佳的问题, 提出了一种基于高效通道注意力和特征融合的协同显著性检测算法。首先, 检测算法利用预训练的深度卷积神经网络对场景进行多尺度特征的提取, 结合边缘显著信息设计了显著性语义特征提取模块, 以避免全卷积神经网络导致边缘信息的缺失; 其次, 通过内积基本原理得到组内图片间的关联性信息并根据其关联程度进行自适应加权, 结合高效通道注意力层设计了协同特征提取算法; 最后, 为了将各级高层语义特征经过协同显著性特征提取之后的结果与浅层次的特征进行融合, 并实现对预测结果进行多分支同步监督, 设计了基于高效通道注意力的特征融合模块。通过对3个经典的数据集进行测试, 并与6种现有的协同显著检测算法进行对比, 结果表明本文所提算法提高了复杂场景中图像的协同显著性检测的精度以及边缘信息的丰富程度, 并具有更优的协同显著性信息检测性能; 通过消融实验进一步验证了所提设计算法各个模块的有效性和必要性。

**关键词:** 深度卷积神经网络; 协同显著性检测; 多尺度特征; 特征提取; 特征融合; 注意力机制

**中图分类号:** TP212

**文献标志码:** A

**文章编号:** 0367-6234(2022)11-0103-09

## Co-saliency detection algorithm with efficient channel attention and feature fusion

ZHANG Dehua<sup>1</sup>, LI Junhao<sup>1</sup>, ZHANG Jingkai<sup>2</sup>, XIAO Qiyang<sup>1</sup>

(1. School of Artificial Intelligence, Henan University, Zhengzhou 450046, China;

2. Miami College, Henan University, Kaifeng 475004, Henan, China)

**Abstract:** Considering the poor performance of existing co-saliency detection algorithms in multiple salient object complex scenarios, a co-saliency detection algorithm with efficient channel attention and feature fusion was proposed. Firstly, the pre-trained deep convolutional neural network was adopted to extract multi-scale features of the images, and a saliency semantic feature extraction module with edge saliency feature was designed to avoid the lack of edge information caused by fully convolutional neural networks. Secondly, the association information between images in the group was obtained based on the inner product principle, and adaptive weighting was carried out according to the association degree; a collaborative feature extraction algorithm was designed in combination with the attention layer of efficient channel. Finally, a feature fusion module based on efficient attention layer was designed, so as to fuse the results of co-saliency feature extraction at high-level semantic features with low-level features, and supervise the predictions with multi-branches simultaneously. Three classic datasets were tested, and six existing collaborative saliency detection algorithms were compared with the proposed algorithm. Results show that the proposed algorithm not only improved the accuracy of collaborative saliency detection and the richness of edge information in complex scenarios, but also had better performance of collaborative saliency detection. The effectiveness and necessity of each designed module were further verified by ablation experiments.

**Keywords:** deep convolutional neural network; co-saliency detection; multi-scale feature; feature extraction; feature fusion; attention mechanism

显著性检测是通过计算机模拟人类视觉系统易被显著物体刺激的特性来检测图像当中的显著物体, 被广泛应用于众多领域, 成为当前图像检测领域

的关键研究之一<sup>[1]</sup>。因传统的显著性检测被认为是一种偏向计算资源的分配, 而且对单一图片中标注显著区域, 无法解决复杂背景下一张图片甚至一组图片包含多个显著目标区域的问题。因此, 为突破传统显著性检测的束缚, 近年来提出了协同显著性检测(co-saliency detection, CD)<sup>[2]</sup>概念, 并成为当前一个热门的研究领域, 目前已广泛应用于如图像裁剪<sup>[2]</sup>、协同分割<sup>[3]</sup>、图像检索<sup>[4]</sup>和视频前景检测<sup>[5]</sup>等许多更复杂的计算机视觉任务当中。

早期研究<sup>[5-6]</sup>主要集中在无监督学习方法, 直

收稿日期: 2021-09-23

基金项目: 国家自然科学基金(61771006, 61802111);

河南省科技厅科技攻关项目(222102220028);

河南省高等学校重点研究计划(20A120005, 22A416004);

河南大学一流学科培育项目(2018YLTD04);

河南省青年人才托举计划(2021HYTP014)

作者简介: 张德华(1984—), 男, 副教授

通信作者: 肖启阳, xqy@henu.edu

接将单幅图的显著特征图作为对应图的特征,然后通过特征匹配和聚类的方法直接将其与其他图的显著特征图进行组合得到最终结果,但其组合的权重是固定不变的,这很明显制约了学习的效果和精度。而文献[7]对此进行了改进,提出了一种秩约束的方法得到自适应组合权重,取得了较好的效果,但由于无监督学习方法依赖人工设计的特征,这些特征常常不包含语义信息<sup>[8]</sup>。因此这些算法在亮度或拍摄角度的改变下等复杂场景中预测结果和真实值差距较大<sup>[9]</sup>。故近年来,在该领域无监督学习的方法研究主要集中于图学习和图优化<sup>[10-12]</sup>。随着计算机算力的提升以及神经网络强大的学习泛化能力,基于深度学习的方法取得了很多不错的效果<sup>[13-16]</sup>,它们很多采用了机器学习领域的先进设计各个模块分别用于提取图片本身的特征和图片组共有的特征,比如文献[16]提出了使用卷积神经网络可能会导致边缘细节信息的缺失,并通过设计联合注意力模块改进这一问题,取得了较好的实验结果。这表明在协同显著性检测当中引入注意力机制是有效的,能够将注意力区域调整至图像中物体所处的区域,进而弱化背景区域。而文献[17]通过整合局部边缘信息和全局位置信息来获得显著边缘特征,文献[18]则从二值分割与边缘映射之间的逻辑关系的角度来克服过往一些研究者只利用边缘特征来改进分割特征的弊端。但像素级特征以及图像级特征的对应关系对显著性检测的精度也起着重要的作用。为此文献[19]考虑了表征图像内部和图像之间的对应关系,并提出了一种具有注意图聚类的自适应图卷积网络。

从上面可以看出,在协同显著性检测中,边缘信息的丰富程度、像素级特征以及图像级特征对多显著目标复杂场景下的预测精度起着非常重要的影响。因此,针对现有协同显著性检测算法在多显著目标复杂场景下表现不佳的问题,本文提出了一种基于高效通道注意力和特征融合的协同显著性检测算法(efficient channel attention and feature fusion-CD, ECAF-CD)。

## 1 相关工作及评价标准

### 1.1 相关工作

在早期的非深度学习的方法当中,研究者们往往使用基于聚类的超像素分割方法来提取图像的特征<sup>[5,8,20]</sup>,这样做的好处是将图像进行了区域分割,降低了维度,并且可以在运算过程中去除异常点。通过对图中超像素的聚类,可以获得图中所含物体的信息,用于后续寻找协同显著性特征。而这种方

法的弊端在于必须指定超像素的个数,并且其所含的特征往往是有关颜色、材质等细节信息,不能提取到高层次的语义信息。

基于深度学习算法的卷积神经网络的特征提取方法因其表现出的优越性能,被广泛用于图像的特征提取。比如广泛用于图像特征提取预训练的VGG16模型<sup>[14-16,21]</sup>,其优势在于预训练的模型是在海量图像当中训练得到的,其表现成熟稳定,不需要指定参数。此外,由于模型的层数多,可以得到含有不同层次的特征,比如低层次的细节特征以及高层次的语义特征。其中,高层次的语义信息可以被用来进一步探索图像之间的关联性,从而实现协同显著性目标的检测。但深度的卷积神经网络容易导致边缘信息的缺失,这在考虑复杂场景时会影响系统的预测精度。

注意力机制最早用于机器翻译,而这种思想也被应用到了计算机视觉当中,用于模拟人眼集中注意力,过滤掉无用的信息。经过数年的发展,注意力机制目前发展出了空间域注意力<sup>[22]</sup>、通道注意力<sup>[23]</sup>和混合型。Gao等<sup>[16]</sup>为了解决使用卷积神经网络可能会导致边缘细节信息的缺失问题,在其模型当中引入了注意力机制并设计了协同注意力模块,它的作用是将图中存在物体的区域的权值调大,而让其余区域的权值调小,从而模拟实现了人眼注意力集中的特点。这表明在协同显著性检测当中引入注意力机制是有效的,能够将注意力区域调整至图像中物体所处的区域,弱化背景区域。

由于任务的复杂性,需要通过不同的方式计算多种不同的特征,它们都代表了图像当中所含的各种信息,但为了得到最终的结果,则要将这些不同的特征进行融合,形成预测图。基于深度学习的方法往往需要进行特征融合,因为在网络结构当中,低层特征的分辨率相对更高,并且含有较多的位置细节信息和干扰,高层特征含有较多的语义信息而对位置细节信息不敏感,所以将二者融合能够有效提高预测的精度和质量。而特征融合的基本操作是相加和拼接,在此基础上研究者们提出了新的结构,比如U-Net<sup>[24]</sup>以及特征金字塔网络(feature pyramid network, FPN)<sup>[25]</sup>。

U-Net主要由压缩路径和扩展路径构成,形成对称结构,在压缩路径当中,通道数不断翻倍,而图像尺寸缩减为原来的二分之一;在扩展路径当中,通过反卷积和拼接实现了通道数减小,尺寸扩大。这种多级特征融合的技术已经被广泛应用在检测和分割领域。而FPN也采用了类似U-Net的思想,先通过压缩获取语义信息,然后再通过相加实现图像尺

寸的恢复,其在不同尺度的目标下表现都比较稳定。

文献[21,26]在协同显著检测算法中引入了类似 U-Net 结构的特征融合方法,根据其结果可以发现多级特征的修正下,预测图中物体所在位置较为准确,在语义和细节上较传统的方法更好。而文献[14]在其结构当中使用了特征金字塔注意力结构,提高了抑制干扰的能力以及精度。

由于现实生活中,存在很多复杂场景,其中的干扰物相对较多,因此在显著性目标检测中,协同信息的提取不仅要考虑边缘信息,还需要考虑像素级特征以及图像级特征对系统性能的影响。为此,本文提出了一种有效的应用于多显著目标复杂场景下的基于高效通道注意力和特征融合的协同显著性检测算法。该算法通过结合边缘信息设计显著性特征提取模块提高边缘信息的丰富程度进而提高预测精度和鲁棒性;同时采用内积衡量机制并根据关联程度进行自适应加权,结合高效通道注意力层设计了协同特征提取模块和特征融合模块。所提算法在协同信息提取方法充分考虑了边缘信息、像素级特征以及图像级特征对系统性能的影响,用同样的评价标准计算程序,与现有主要文献[5,17-19,21]相对比,验证了本文算法的有效性。

## 1.2 评价标准

为有效评价协同显著性检测算法的优劣,常使用多种定量分析指标进行全方位的分析,比如精准率-召回率(precision-recall, P-R)、平均 F-measure、平均绝对误差( $\delta_{MAE}$ )、S-measure 和平均 E-measure 等。

通过对预测二值化后的显著图的遍历,可以根据下式计算得到对应预测图的 P-R 值。

$$\begin{cases} P = \frac{T_p}{T_p + F_p} \\ R = \frac{T_p}{T_p + F_N} \end{cases} \quad (1)$$

式中: $P$  和  $R$  分别为精准率、召回率, $T_p$  为标签所标注的协同显著目标区域, $F_p$  为将背景区域替换为前景区域, $F_N$  为将协同显著目标区域替换为背景区域。由于  $P$  和  $R$  相对独立,难以对算法的表现进行综合且直观的评价,故引入定义为精准率以及召回率的加权调和平均值 F-measure,用于更加全面地评价算法的表现<sup>[27]</sup>,定义为

$$F_m = \frac{(1+\beta^2)P \times R}{\beta^2 \times P + R} \quad (2)$$

式中  $\beta^2$  一般定义为 0.3,用于强调准确率<sup>[28]</sup>。

为最直观的评估协同显著性检测算法的优劣,

$\delta_{MAE}$  用来衡量预测结果和真值图之间的平均绝对误差,其数学表达式如下:

$$\delta_{MAE} = \frac{1}{W \times H} \sum_{x=1}^W \sum_{y=1}^H | \hat{P}(x,y) - \hat{G}(x,y) | \quad (3)$$

式中  $W$  和  $H$  为预测图形的宽度和高度,而  $\hat{P}(x,y)$  和  $\hat{G}(x,y)$  分别为归一化到  $[0,1]$  后预测图和真值图在  $(x,y)$  的像素值。计算所得的  $\delta_{MAE}$  越小,则表示算法的预测性能越好。

由于对目标的检测不仅要考虑预测值与真值图之间的差异,还有考虑目标的整体结构,因此 S-measure<sup>[29]</sup> 用于衡量算法的区域感知和对对象感知能力,即用于衡量协同显著性检测算法的稳定性和准确性,其定义为

$$S_m = \alpha \cdot S_o + (1 - \alpha) \cdot S_r \quad (4)$$

式中: $S_o$  为面向物体结构相似性度量, $S_r$  指面向区域相似性度量, $\alpha$  为平衡参数,一般设置为 0.5。

除此之外,在评价预测图和标签值时为了兼顾像素级和图像级属性<sup>[30]</sup>,定义平均 E-measure 为

$$\begin{cases} E_m = \frac{1}{W \times H} \sum_{x=1}^W \sum_{y=1}^H \varphi_{FM}(x,y) \\ \varphi_{FM} = f(\xi_{FM}) \\ \xi_{FM} = \frac{2\varphi_{GT} \circ \varphi_{FM}}{\varphi_{GT} \circ \varphi_{FM} + \varphi_{FM} \circ \varphi_{FM}} \end{cases} \quad (5)$$

式中: $f(\cdot)$  为凸函数, $f(x) = \frac{1}{4}(1+x)^2$ , $\varphi_{GT}$  和  $\varphi_{FM}$  分别为标签值的偏差矩阵和预测图的偏差矩阵, $\circ$  为 Hadamard 乘积。

## 2 网络结构设计

本文提出的基于高效通道注意力和特征融合的协同显著性目标检测算法模型见图 1,主要包括语义特征提取模块、协同特征提取模块以及特征融合模块。相对现有文献,其进步点有三:第一,结合边缘信息设计的显著性检测模块对初步所提取的特征进行修正和过滤,以提供相对丰富的边缘信息,提高预测的精度,并为后续协同特征提取提供丰富信息;第二,利用修正后的特征得到初步的协同显著语义信息,并利用自适应加权机制结合高效通道注意力机制得到各级的协同显著预测结果;第三,对各级的预测结果通过特征融合模块得到最终的预测图,同时利用训练集中的真值对每一级产生的预测图进行监督训练,实现对预测结果进行多分支同步监督,智能优化各级网络的结构参数。其相应的实现流程见图 2。

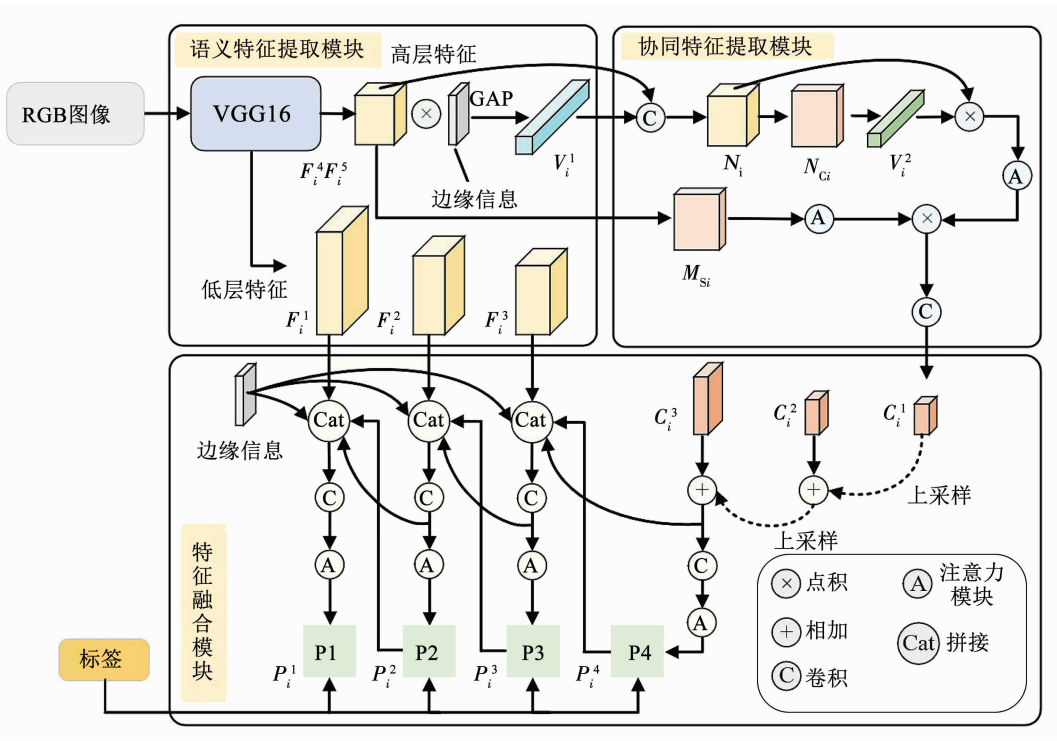


图 1 高效通道注意力和特征融合协同显著性检测算法模型

Fig. 1 Co-saliency detection algorithm model with efficient channel attention and feature fusion

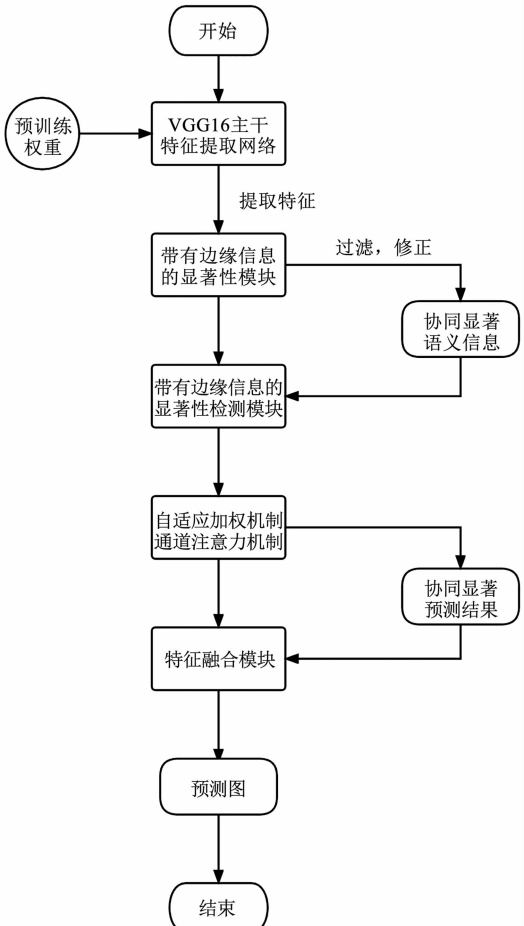


图 2 高效通道注意力和特征融合协同显著性检测算法实现流程图

Fig. 2 Flow chart of co-saliency detection algorithm with efficient channel attention and feature fusion

2.1 语义特征提取

在本文中,对每一组输入的图片进行 5 个尺度的特征提取,给定输入 RGB 图像  $\{I_i, i = 1, 2, \dots, n\}$ ,  $I_i \in \mathbb{R}^{224 \times 224}$ ,其中  $n$  为一组图像的数量,为了便于输入网络,将其大小重置为  $224 \times 224$ 。可以将输入图片组的特征表示为  $\{F_i^j, i = 1, 2, \dots, n; j = 1, 2, 3, 4, 5\}$ ,其中  $i$  为图片的序号,  $n$  为一组图片的数量,  $j$  为特征尺度,  $1 \sim 5$  分别表示由低至高 5 个尺度,其特征通道数分别为 64、128、256、512、512。在本文中,考虑到 VGG16 的浅层次特征大多含有物体的边缘轮廓等细节信息和背景信息,高层次特征含有更加抽象的有关物体的语义特征,因此选取  $F_i^4$  和  $F_i^5$  两个尺度用于协同显著信息的提取,而剩余的 3 个尺度特征  $\{F_i^1, F_i^2, F_i^3\}$  用于最后融合细节信息生成预测图。考虑到显著性检测算法已经较为成熟,能够有效地检测出图片当中的显著物体信息,所以先采用显著性检测算法对图片进行检测,然后将其与所提取的特征做点乘,从而筛选出显著区域。

受启发于通道注意力机制<sup>[31]</sup>,并考虑到显著区域可能是不准确的,在本文中,对于每一个尺度特征,需要计算每一个通道对后续运算的权重,从而使得模型能够集中于图片中的特定显著物体,并根据对应图像的边缘信息对各个尺度的特征进行修正,然后对其进行全局平均池化操作,将其压缩得到一个值,用于量化每个尺度特征各个通道的贡献程度,其具体操作定义为

$$V_i^j = \text{GAP}(F_i^j) = \frac{1}{W \times H} \sum_{x=1}^H \sum_{y=1}^W F_i^j(x, y) \odot \varepsilon_i(x, y) \quad (6)$$

式中:  $\odot$  代表矩阵的点乘运算,  $\varepsilon_i(x, y)$  为由显著性检测算法得到的对应图片的边缘显著信息, 将其与特征相乘即可滤除干扰信息并得到含有图中显著物体的特征信息, 为下一步协同特征提取提供语义信息。

## 2.2 协同特征提取

协同特征提取模块见图 1, 由上下两个分支组成, 输入特征分别经过上下两个分支后融合得到最后结果。

图像经过语义特征提取模块之后, 本文选取  $F_i^4$  和  $F_i^5$  两个含有高层语义特征的信息, 然后利用式(6)计算得到的值作为卷积核对图像特征进行卷积运算, 这样做是为了用单幅图的语义信息过滤图像特征。将过滤后的高层语义图像特征的通道压缩一半, 同时将其大小再压缩一半, 然后将其分上下两路操作: 图片级相关性特征提取和像素级相关性特征提取。受启发于文献[21]通过计算内积来衡量相似度的方法, 本文通过采用内积的基本原理得到组内图片之间的关联性信息, 具体来说, 将  $N_i \in \mathbb{R}^{N \times H \times W}$  展开为  $\mathbb{R}^{N \times HW}$ , 其中  $N$  为一组图片的数量,  $H$  和  $W$  分别为特征的高度和宽度, 然后计算出权重矩阵  $M_{Ci} \in \mathbb{R}^{N \times N}$ , 将其求和后通过 softmax 函数映射后得到关系向量  $V_i^2 \in \mathbb{R}^{N \times 1}$ , 具体操作过程为

$$\begin{cases} M_{Ci} = N_i^T N \\ V_i^2 = \text{softmax}(\sum (M_{Ci})) \end{cases} \quad (7)$$

根据  $V_i^2$  可以得到当前同组图片之间的相关度信息, 将  $V_i^2$  和特征结合即可得到加权后的协同特征。同样, 本文对图片自身的特征各个像素进行相似度计算, 最后将其维度改变至下一步适合运算的大小, 具体可表示为

$$\begin{cases} \hat{F} = \text{reshape}(F) \rightarrow \mathbb{R}^{C \times HW} \\ \bar{M}_{Si} = \hat{F}^T \hat{F} \\ M_{Si} = \text{reshape}(\bar{M}_{Si}) \rightarrow \mathbb{R}^{HW \times H \times W} \end{cases} \quad (8)$$

式中:  $C$  是特征通道的数目,  $\bar{M}_{Si} \in \mathbb{R}^{HW \times HW}$ ,  $F \in \mathbb{R}^{C \times H \times W}$  代表输入的特征信息。最后, 将所得的图片级相关特征和每张图片像素级特征分别经由高效注意力通道后相乘, 再经过卷积操作后得到协同显著图。文献[32]中提供了一种用前景和背景融合后加强预测结果的方法, 可以使得背景和目標物体区分更明显, 会将无用的背景信息滤除, 进一步加强前景物体。而本文采用了类似策略, 分别对图像的前

景和背景信息提取协同显著性特征, 然后融合得到最终的结果, 最后对高层语义特征的协同显著特征提取得到了  $C_i^1$ 、 $C_i^2$  和  $C_i^3$  用于进一步的特征融合。

## 2.3 特征融合

该模块右侧采用了类似 U-Net 的设计, 其相应算法可由下式表示:

$$L_n = \begin{cases} C_j^n, n=1 \\ L_{n-1} + \text{upsample}(C_j^n), n \geq 2 \end{cases} \quad (9)$$

式中:  $L_n$  为经过融合之后的特征,  $C_j^n$  为协同显著特征。

将高层次语义特征采用上式融合后, 经过一个卷积和注意力模块得到初步协同显著性预测图  $P_i^4$ 。为了进一步修正预测图, 本模块通过将低层次富含细节的特征  $F_i^1$ 、 $F_i^2$  和  $F_i^3$  融合提取到的高层语义特征, 分别得到协同显著性预测图  $P_i^3$  至  $P_i^1$ 。其具体实现算法如下:

$$P_i^n = \begin{cases} A(\text{conv}(L_3)), n=4 \\ A(\text{conv}(\text{Cat}(\varepsilon_i, F_i^n, P_i^{n-1}, p_i^{n-1}))), n=1, 2, 3 \end{cases} \quad (10)$$

式中:  $\varepsilon_i$  为当前图片对应的边缘信息,  $A(\cdot)$  为注意力模块,  $\text{conv}(\cdot)$  为卷积操作,  $\text{Cat}(\cdot)$  为合并操作,  $L_3$  为融合后的特征, 由式(10)得到。

由于简单的通道融合容易引入干扰从而降低预测图的质量, 因此在融合后加入了卷积层和注意力模块。本文采用的高效通道注意力层能够进一步改善网络的预测性能, 更好地利用各部分特征信息, 本文的消融实验验证了其有效性。

得到协同显著性预测图之后, 通过施加 4 个监督分支, 增强了监督的深度, 并且可以最大程度地利用训练标签, 加快网络的收敛。对于每一个分支, 本文定义损失函数如下:

$$L(M, G) = 1 - \frac{1}{n} \sum_{i=1}^n \frac{\sum_{j=1}^{h_w} (\min(P_i, G_i))_j}{\sum_{j=1}^{h_w} (\max(P_i, G_i))_j} \quad (11)$$

式中:  $P_i$  和  $G_i$  分别指的是图片的预测图和真值标签,  $h_w$  为图片高度  $H$  和宽度  $W$  的乘积值。而网络的总损失由每一个回路的损失相加取平均得到, 具体可以表示为

$$L_{\text{final}} = \frac{1}{n} \sum_{j=1}^n L_n(P_i^j, G) \quad (12)$$

式中  $n$  取 4, 对应本文的 4 条分支。

## 3 实验方法及结果

### 3.1 实验设置

本文所涉及图片均为 RGB 类型, 实验系统环

境为 Windows10,虚拟环境管理软件是 Anaconda3,所使用的 GPU 为 NVIDIA RTX 3060 Laptop(6 GB 显存),所使用的编译器为 PyCharm Community Edition 2021,所用的库主要为 PyTorch。所使用的优化方法是 Adam 优化器<sup>[33]</sup>, $\beta_1$  和  $\beta_2$  分别设置为 0.9 和 0.999,训练周期为 60,学习率调整为  $10^{-5}$ ,权重衰减率为  $10^{-4}$ 。定量评价指标采用平均绝对误差  $\delta_{MAE}$ 、F-measure<sup>[27]</sup>、E-measure<sup>[34]</sup> 和 S-measure<sup>[29]</sup> 4 种。

为了有效评估协同显著性检测模型的优劣,研究者提出了多种针对此类任务的基准数据集,其中最常用的有 COCO9213 数据集、CoSal2015<sup>[30]</sup>、iCoSeg<sup>[35]</sup> 和 MSRC<sup>[36]</sup>。特别是从复杂的日常生活场景中截取的 COCO 数据集,含有超过 30 万张图片,250 万个标签以及 91 个类别,是非常理想的用作训练的数据集。为了有效减少训练时间,本文训练时采用的数据集为 COCO9213,此为 COCO 数据集的子集,经常用于协同显著性检测算法的训练<sup>[37-38]</sup>。它包括 9 213 张图片,含有 90 个类别,平均每个类有 102 张图片,并且每张图片都有其对应的像素级真值(ground truth),所有的训练图片和真值标签大小均被置为  $224 \times 224$ 。训练集以及测试数据集的信息见表 1。

表 2 不同算法在 3 个数据集上的表现

Tab. 2 Performance of different algorithms on three datasets

| 算法      | 年份 - 来源   | CoSal2015             |                       |                       |                       | iCoSeg                |                       |                       |                       | MSRC                  |                       |                       |                       |
|---------|-----------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
|         |           | $\delta_{MAE}$        | $F_m$                 | $E_m$                 | $S_m$                 | $\delta_{MAE}$        | $F_m$                 | $E_m$                 | $S_m$                 | $\delta_{MAE}$        | $F_m$                 | $E_m$                 | $S_m$                 |
| CBCS    | 2013-TIP  | 0.230 7               | 0.383 5               | 0.520 0               | 0.552 9               | 0.172 2               | 0.569 7               | 0.630 1               | 0.670 3               | 0.313 8               | 0.416 0               | 0.473 5               | 0.493 6               |
| GICD    | 2020-ECCV | <b><u>0.075 5</u></b> | <b><u>0.819 9</u></b> | <b><u>0.871 2</u></b> | 0.831 1               | 0.068 8               | <b><u>0.818 2</u></b> | <b><u>0.880 2</u></b> | 0.818 8               | 0.179 9               | 0.706 7               | 0.734 0               | 0.687 4               |
| GCAGC   | 2020-CVPR | <b><u>0.081 3</u></b> | 0.792 0               | 0.835 6               | <b><u>0.844 4</u></b> | 0.076 7               | 0.799 4               | 0.847 7               | <b><u>0.863 7</u></b> | 0.198 5               | 0.661 9               | 0.705 4               | 0.691 7               |
| CBCS-s  | 2013-TIP  | 0.254 6               | 0.470 4               | 0.588 3               | 0.595 4               | 0.195 6               | 0.606 5               | 0.674 0               | 0.697 2               | 0.303 3               | 0.530 3               | 0.556 3               | 0.581 3               |
| EGNet   | 2020-ICCV | 0.098 3               | 0.749 8               | 0.821 9               | 0.815 2               | <b><u>0.067 6</u></b> | 0.771 2               | 0.844 2               | 0.832 3               | <b><u>0.162 2</u></b> | <b><u>0.719 4</u></b> | <b><u>0.759 9</u></b> | <b><u>0.738 4</u></b> |
| SCRN    | 2020-ICCV | 0.088 7               | <b><u>0.793 7</u></b> | <b><u>0.850 0</u></b> | <b><u>0.839 8</u></b> | <b><u>0.054 0</u></b> | <b><u>0.835 3</u></b> | <b><u>0.889 1</u></b> | <b><u>0.873 9</u></b> | <b><u>0.156 1</u></b> | <b><u>0.749 8</u></b> | <b><u>0.777 9</u></b> | <b><u>0.750 1</u></b> |
| ECAF-CD |           | <b><u>0.072 9</u></b> | <b><u>0.817 5</u></b> | <b><u>0.878 3</u></b> | <b><u>0.838 7</u></b> | <b><u>0.057 4</u></b> | <b><u>0.836 0</u></b> | <b><u>0.902 9</u></b> | <b><u>0.854 0</u></b> | <b><u>0.148 5</u></b> | <b><u>0.797 4</u></b> | <b><u>0.806 0</u></b> | <b><u>0.762 0</u></b> |

在表 2 中, $\delta_{MAE}$ 、 $F_m$ 、 $E_m$  和  $S_m$  分别表示平均绝对误差、F-measure、E-measure 和 S-measure,黑体双下划线、黑体单下划线和黑体下划波浪线分别表示第 1 名、第 2 名和第 3 名。从结果可以看出,得益于强大的边缘信息捕捉能力,本文所提算法 ECAF-CD 展现了很好的显著性检测能力。在 MSRC 数据集上的综合表现完胜其他 3 种显著性检测算法和 3 种协同显著性检测算法。当图片只包含单个显著物体

表 1 数据集的信息

Tab. 1 Dataset information

| 数据集       | 年份   | 类别/类 | 图像数/张 | 平均每组图像/<br>(张/类) |
|-----------|------|------|-------|------------------|
| COCO9213  | 2014 | 90   | 9 213 | 102              |
| MSRC      | 2005 | 8    | 240   | 30               |
| iCoSeg    | 2010 | 38   | 643   | 17               |
| CoSal2015 | 2015 | 50   | 2 015 | 40               |

3.2 定量分析

本文所提算法在 3 个经典数据集上进行了测试。为了验证本文算法对比显著性检测算法在协同信息提取上更有优势,本文选用了 CBCS-s<sup>[5]</sup>、EGNet<sup>[17]</sup> 和 SCRN<sup>[18]</sup> 3 种典型的显著性检测算法进行对照实验。同时为了验证本文所提算法在协同信息提取上更有优势,还与现有的协同显著性检测算法 CBCS<sup>[5]</sup>、GICD<sup>[21]</sup> 和 GCAGC<sup>[19]</sup> 进行了对比研究。为了保证实验环境的一致性,所有算法均在相同的系统环境下运行实现,并且采用同样的评价标准计算程序,其中 CBCS、CBCS-s 算法结果采用了其相应作者提供的 MATLAB 源代码在 MATLAB2016b 上直接运行得到,GICD、EGNet 和 SCRN 均采用了作者提供的 Python 开源代码和预训练模型得到预测图,而 GCAGC 采用了作者提供的 3 个测试集的预测图。实验所得的结果见表 2。

时,虽然 SCRN 算法善于检测物体细节,并在 3 个数据集的测试当中都超过了其他协同显著性检测算法,但本文所提算法 ECAF-CD 在作为全面评价算法的定量指标  $F_m$  以及兼顾像素级和图像级属性平均  $E_m$  指标上均优于 SCRN 算法。因此本文所提算法不但提高了边缘信息的丰富程度,也提高了多显著目标场景下的预测精度。本文所提算法更直观的性能优势也可从视觉比较图 3 中得到验证。





图 3 算法 ECAF-CD 在 3 个数据集当中的视觉比较

Fig. 3 Visual comparison of algorithm ECAF-CD on three datasets

3.3 消融实验

为了进一步验证本文所提算法的各个模块和边缘信息对预测的贡献和影响,本文评估了算法在缺失各个模块或特征的情况下的表现,对照组为完整的本文所提出的算法模型。对于缺失边缘信息的模型,本文将其置为全 1 的矩阵;而对于缺失注意力模块的模型,本文将注意力模块替换为卷积层和激活函数的传统组合;对于缺失像素级相关特征的模型,本文将用图像级相关特征代替,并调整其尺寸使之

适应网络结构;对于缺失图像级相关特征的模型,本文将用像素级相关特征进行代替,并调整其尺寸。每改变一次网络的结构,都会将模型进行重新训练,统一训练周期为 60,并且除网络结构不同外,其余训练参数配置均保持一致,测试所选用的模型均为 60 个训练周期当中训练误差最小的,测试所用的程序和定量评价标准计算方法均一致。所选用的测试数据集和衡量标准和上文保持一致。实验的结果见表 3。

表 3 本文 ECAF-CD 模型的消融实验

Tab. 3 Ablation experiment of ECAF-CD model

| 模型        | CoSal2015      |                |                |                | iCoSeg         |                |                |                | MSRC           |                |                |                |
|-----------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
|           | $\delta_{MAE}$ | $F_m$          | $E_m$          | $S_m$          | $\delta_{MAE}$ | $F_m$          | $E_m$          | $S_m$          | $\delta_{MAE}$ | $F_m$          | $E_m$          | $S_m$          |
| 缺失边缘信息    | 0.136 8        | 0.694 9        | 0.772 0        | 0.773 5        | 0.104 0        | 0.726 7        | 0.821 7        | 0.800 0        | 0.151 6        | 0.776 8        | 0.804 8        | <b>0.765 9</b> |
| 缺失注意力模块   | 0.074 6        | <b>0.819 6</b> | 0.875 1        | 0.832 9        | 0.069 0        | 0.806 4        | 0.863 1        | 0.810 9        | 0.170 7        | 0.767 4        | 0.775 9        | 0.723 5        |
| 缺失像素级相关特征 | 0.074 1        | 0.814 0        | 0.870 9        | <b>0.839 4</b> | 0.062 7        | 0.807 8        | 0.878 3        | 0.834 5        | 0.161 3        | 0.734 4        | 0.770 1        | 0.723 4        |
| 缺失图像级相关特征 | 0.105 5        | 0.721 8        | 0.803 4        | 0.805 1        | 0.075 5        | 0.760 8        | 0.840 6        | 0.819 0        | 0.156 5        | 0.752 9        | 0.755 9        | 0.742 0        |
| 对照组       | <b>0.072 9</b> | 0.817 5        | <b>0.878 3</b> | 0.838 7        | <b>0.057 4</b> | <b>0.836 0</b> | <b>0.902 9</b> | <b>0.854 0</b> | <b>0.148 5</b> | <b>0.797 4</b> | <b>0.806 0</b> | 0.762 0        |

表 3 中黑体表示最优值,可以发现在缺失边缘信息情况下,模型表现在 CoSal2015 和 iCoSeg 下降了约 20%,并且在 3 个数据集当中表现均不如对照组。这表明边缘信息对于本模型非常重要,因为其包含的细节信息有利于提高预测的精度。而在缺失注意力模块的情况下,其在 CoSal2015 中表现尚可,但是在 iCoSeg 和 MSRC 中则下降了约 5%,尤其是在 MSRC 的测试当中误差较大,下降了约 20%,这表明本文对于特征的处理以及引入的高效通道注意力层对协同特征精准提取起到了非常关键的作用;

移除像素级相关特征之后,可以观察到在 3 个数据集当中图片整体的表现有所下降,但下降幅度在 CoSal2015 和 iCoSeg 不大,这是因为这两个数据集中干扰物体较少,大多只含一个显著目标,所以缺乏像素级相关特征影响较小,而在 MSRC 当中部分图片包含多个物体,因此缺乏像素级特征后准确率有所下降;在移除图像级相关特征之后,模型的表现性能下降了约 10%,这是因为缺失图像级相关特征之后,图像之间的相关信息无法得到利用,这使得最终的预测受到影响,这表明本文基于内积的基本原理

得到组内图片间的关联性信息并根据关联层度进行自适应加权机制使得图像之间的相关信息得到了充分使用。

## 4 结 论

针对现有的协同显著性检测算法在多显著目标场景下表现性能不佳的问题,本文提出了一种基于高效通道注意力和特征融合的协同显著性检测算法,主要包括语义特征提取模块、协同特征提取模块以及特征融合模块。针对全卷积神经网络可能导致缺失边缘信息,对于每一个尺度特征,计算每一个通道对后续运算的权重,从而使得模型能够集中于图片中的特定显著物体,同时结合边缘信息设计显著性特征提取模块并量化每个尺度特征对各个通道的贡献程度,使得本文算法能够提供相对丰富的边缘信息,过滤掉无用的干扰信息,提高预测的精度和鲁棒性,并为下一步的协同特征提取提供更加丰富的信息;通过采用内积衡量机制的自适应加权机制,本文将前述得到的语义信息的特征图,通过内积得到组内图片间的关联性信息,根据其关联程度进行自适应加权,并引入高效通道注意力层设计了协同特征提取模块;为了将各级高层语义特征经过协同显著性特征提取之后的结果与浅层次的特征进行融合,并实现对预测结果进行多分支同步监督,设计了基于高效通道注意力的特征融合模块。通过在公开的数据集测试,结果表明本文所提算法不仅提高了对复杂场景中 RGB 图像的协同显著性检测精度也提高了边缘信息的丰富程度,其性能指标优于文中 6 种对比算法,同时通过消融实验进一步验证了本文所设计各个模块的有效性。

## 参 考 文 献

- [1] 钱晓亮,白臻,陈渊,等. 协同视觉显著性检测方法综述[J]. 电子学报, 2019, 47(6): 1352  
QIAN Xiaoliang, BAI Zhen, CHEN Yuan, et al. A review of co-saliency detection[J]. Acta Electronica Sinica, 2019, 47(6): 1352. DOI:10.3969/j.issn.0372-2112.2019.06.024
- [2] JACOBS D E, GOLDMAN D B, SHECHTMAN E. Cosaliency: where people look when comparing images[C]//Proceedings of the 23rd Annual ACM Symposium on User Interface Software and Technology. New York: Association for Computing Machinery, 2010: 219. DOI:10.1145/1866029.1866066
- [3] WANG Wenguan, SHEN Jianbing. Higher-order image co-segmentation[J]. IEEE Transactions on Multimedia, 2016, 18(6): 1011. DOI: 10.1109/TMM.2016.2545409
- [4] CHENG Mingming, MITRA N J, HUANG Xiaolei, et al. SalientShape: group saliency in image collections[J]. The Visual Computer, 2014, 30(4): 443. DOI: 10.1007/s00371-013-0867-4
- [5] FU Huazhu, CAO Xiaochun, TU Zhuowen. Cluster-based co-saliency detection[J]. IEEE Transactions on Image Processing, 2013, 22(10): 3766. DOI:10.1109/TIP.2013.2260166
- [6] LI Hongliang, NGAN K N. A co-saliency model of image pairs[J]. IEEE Transactions on Image Processing, 2011, 20(12): 3365. DOI:10.1109/TIP.2011.2156803
- [7] CAO Xiaochun, TAO Zhiqiang, ZHANG Bao, et al. Self-adaptively weighted co-saliency detection via rank constraint[J]. IEEE Transactions on Image Processing, 2014, 23(9): 4175. DOI:10.1109/TIP.2014.2332399
- [8] ZHANG Dingwen, FU Huazhu, HAN Junwei, et al. A review of co-saliency detection algorithms: fundamentals, applications, and challenges[J]. ACM Transactions on Intelligent Systems & Technology, 2018, 9(4): 1. DOI:10.1145/3158674
- [9] FAN Dengping, LIN Zheng, LI Gepeng, et al. Taking a deeper look at co-salient object detection[C]//Proceedings of the 2000 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2020: 2919. DOI:10.1109/CVPR42600.2020.00299
- [10] JIANG Bo, JIANG Xingyue, TANG Jin, et al. Co-saliency detection via a general optimization model and adaptive graph learning[J]. IEEE Transactions on Multimedia, 2020, 23:3193. DOI:10.1109/TMM.2020.3021251
- [11] TSAI C C, LI Weizhi, HSU K J, et al. Image co-saliency detection and co-segmentation via progressive joint optimization[J]. IEEE Transactions on Image Processing, 2018, 28(1): 56. DOI:10.1109/TIP.2018.2861217
- [12] TSAI C C, HSU K J, LIN Y Y, et al. Deep co-saliency detection via stacked autoencoder-enabled fusion and self-trained CNNs[J]. IEEE Transactions on Multimedia, 2019, 22(4): 1016. DOI:10.1109/TMM.2019.2936803
- [13] 吴泽民,王军,胡磊,等. 基于卷积神经网络与全局优化的协同显著性检测[J]. 电子与信息学报, 2018, 40(12): 2896  
WU Zemin, WANG Jun, HU Lei, et al. Co-saliency detection based on convolutional neural network and global optimization[J]. Journal of Electronics & Information Technology, 2018, 40(12): 2896. DOI: 10.11999/JEIT180241
- [14] ZHA Zhengjun, WANG Chong, LIU Dong, et al. Robust deep co-saliency detection with group semantic and pyramid attention[J]. IEEE Transactions on Neural Networks and Learning Systems, 2020, 31(7): 2398. DOI: 10.1109/TNNLS.2020.2967471
- [15] KONG S, FOWLKES C. Low-rank bilinear pooling for fine-grained classification[C]//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 7025. DOI: 10.1109/CVPR.2017.743
- [16] GAO Guangshuai, ZHAO Wenting, LIU Qingjie, et al. Co-saliency detection with co-attention fully convolutional network[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2020, 31(3): 877. DOI:10.1109/TCSVT.2020.2992054
- [17] ZHAO Jiaxing, LIU Jiangjiang, FAN Dengping, et al. EGNNet: edge guidance network for salient object detection[C]//Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2019: 8779. DOI: 10.1109/ICCV.2019.00887
- [18] WU Zhe, SU Li, HUANG Qingming. Stacked cross refinement network for edge-aware salient object detection[C]//Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019: 7264. DOI:10.1109/ICCV.2019.00736



- [19] ZHANG Kaihua, LI Tengteng, SHEN Shiwen, et al. Adaptive graph convolutional network with attention graph clustering for co-saliency detection [C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision & Pattern Recognition. Piscataway: IEEE, 2020; 9050. DOI:10.1109/CVPR42600.2020.00907
- [20] 刘政怡, 刘俊雷, 赵鹏. 基于样本选择的 RGBD 图像协同显著目标检测[J]. 电子与信息学报, 2020, 42(9): 2277  
LIU Zhengyi, LIU Junlei, ZHAO Peng. RGBD image co-saliency object detection based on sample selection [J]. Journal of Electronics & Information Technology, 2020, 42(9): 2277. DOI: 10.11999/JETI190393
- [21] ZHANG Zhao, JIN Wenda, XU Jun, et al. Gradient-induced co-saliency detection[C]//Proceedings of the European Conference on Computer Vision. Cham: Springer, 2020; 455. DOI:10.1007/978-3-030-58610-2\_27
- [22] JADERBERG M, SIMONYAN K, ZISSERMAN A, et al. Spatial transformer networks [C]//Proceedings of the 28th International Conference on Neural Information Processing Systems. Cambridge: MIT Press, 2015: 2017
- [23] HU Jie, SHEN Li, SUN Gang, et al. Squeeze-and-excitation networks[C]//Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018; 7132. DOI:10.1109/CVPR.2018.00745
- [24] RONNEBERGER O, FISCHER P, BROX T. U-net: convolutional networks for biomedical image segmentation [C]//Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer, 2015; 234. DOI:10.1007/978-3-319-24574-4\_28
- [25] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017; 936. DOI:10.1109/CVPR.2017.106
- [26] JIN Wenda, XU Jun, CHENG Mingming, et al. ICNet: intra-saliency correlation network for co-saliency detection[J]. Advances in Neural Information Processing Systems, 2020, 33: 18749
- [27] OTSU N. A threshold selection method from gray-level histograms [J]. IEEE Transactions on Systems, Man, and Cybernetics, 1979, 9(1): 62. DOI:10.1109/TSMC.1979.4310076
- [28] ZHANG Dingwen, HAN Junwei, LI Chao, et al. Detection of co-salient objects by looking deep and wide[J]. International Journal of Computer Vision, 2016, 120(2): 215. DOI:10.1007/s11263-016-0907-4
- [29] FAN Dengping, CHENG Mingming, LIU Yun, et al. Structure-measure: a new way to evaluate foreground maps[C]//Proceedings of the 2017 IEEE International Conference on Computer Vision. Piscataway: IEEE, 2017; 4548. DOI:10.1109/ICCV.2017.487
- [30] ZHANG Dingwen, HAN Junwei, LI Chao, et al. Co-saliency detection via looking deep and wide[C]//Proceedings of the 2015 IEEE Conference on Computer Vision & Pattern Recognition. Piscataway: IEEE, 2015; 2994. DOI: 10.1109/CVPR.2015.7298918
- [31] HU Jie, SHEN Li, ALBANIE S, et al. Squeeze-and-excitation networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(8): 2011. DOI:10.1109/TPAMI.2019.2913372
- [32] WANG Ziqin, XU Jun, LIU Li, et al. Ranet: ranking attention network for fast video object segmentation [C]//Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2019; 3978. DOI:10.1109/ICCV.2019.00408
- [33] KINGMA D, BA J. Adam: a method for stochastic optimization [C]//Proceedings of the 2015 International Conference on Learning Representations. San Diego: [s.n.], 2015
- [34] FAN Dengping, GONG Cheng, CAO Yang, et al. Enhanced-alignment measure for binary foreground map evaluation [C]//Proceedings of the 2018 International Joint Conference on Artificial Intelligence. Menlo Park: AAAI Press, 2018; 698. DOI:10.5555/3304415.3304515
- [35] BATRA D, KOWDLE A, PARIKH D, et al. iCoseg: interactive co-segmentation with intelligent scribble guidance[C]//Proceedings of the 2010 IEEE Conference on Computer Vision & Pattern Recognition. Piscataway: IEEE, 2010; 3169. DOI: 10.1109/CVPR.2010.5540080
- [36] WINN J, CRIMINISI A, MINKA T. Object categorization by learned universal visual dictionary [C]//Proceedings of the Tenth IEEE International Conference on Computer Vision. Piscataway: IEEE, 2005, 2: 1800. DOI:10.1109/ICCV.2005.171
- [37] LI Bo, SUN Zhengxing, TANG Lü, et al. Detecting robust co-saliency with recurrent co-attention neural network[C]//Proceedings of the 28th International Joint Conference on Artificial Intelligence. [S.l.]: IJCAI, 2019; 818. DOI:10.24963/ijcai.2019/115
- [38] ZHENG Xiaoju, ZHA Zhengjun, ZHUANG Liansheng. A feature-adaptive semi-supervised framework for co-saliency detection [C]//Proceedings of the 26th ACM International Conference on Multimedia. New York: Association for Computing Machinery, 2018; 959. DOI:10.1145/3240508.3240648

(编辑 苗秀芝)