

## 特征混合增强与多损失融合的显著性目标检测

李春标, 谢林柏<sup>+</sup>, 彭 力

物联网技术应用教育部工程研究中心(江南大学 物联网工程学院), 江苏 无锡 214122

+ 通信作者 E-mail: xie\_linbo@jiangnan.edu.cn

**摘要:**针对当前显著性目标检测算法存在的特征缺失和区域一致性差的问题,基于全卷积神经网络提出一种特征混合增强与多损失融合的显著性目标检测算法。该算法包含上下文感知预测模块(CAPM)和特征混合增强模块(FHEM)。首先利用上下文感知预测模块提取图像多尺度特征信息,并且在预测模块中嵌入空间感知模块(SAM)以进一步提取图像高层语义信息,然后利用特征混合增强模块对预测模块产生的全局特征信息和细节特征信息进行有效的整合,并利用特征聚合模块(FAM)对整合的信息进行特征增强。此外,提出了一种新的区域增强损失函数(RA),并将此损失函数与已有的二进制交叉熵(BCE)损失函数、结构化相似度(SSIM)损失函数融合,以多损失融合的方式监督网络保持前景区域的完整性以及增强区域像素一致性。在五个包含多个显著性目标和复杂背景的图像数据集上对算法进行验证,实验结果表明,该算法有效地提高了复杂场景下显著性目标的检测精度,改善了显著图特征缺失与区域一致性差的问题。

**关键词:**卷积神经网络;显著性目标检测;特征混合增强;区域增强损失

**文献标志码:**A **中图分类号:**TP391.4

## Salient Object Detection with Feature Hybrid Enhancement and Multi-loss Fusion

LI Chunbiao, XIE Linbo<sup>+</sup>, PENG Li

Engineering Research Center of Internet of Things Technology Applications (School of Internet of Things Engineering, Jiangnan University), Ministry of Education, Wuxi, Jiangsu 214122, China

**Abstract:** To tackle the problem of missing features and poor regional consistency in existing salient object detection algorithms, a salient object detection network which uses feature hybrid enhancement and multi-loss fusion based on fully convolutional neural network is proposed. The network includes a context-aware prediction module (CAPM) and a feature hybrid enhancement module (FHEM). First, the context-aware prediction module is used to extract the multi-scale feature information of the image, in which the spatial-aware module (SAM) is embedded to further extract the high-level semantic information of the image. Furthermore, the feature hybrid enhancement module is used to effectively integrate the global feature information and the detailed feature information generated by the prediction module, and the integrated feature is enhanced through embedded feature aggregation module (FAM). In addition, the multi-loss fusion method is used to supervise the network, which combines the binary cross-entropy (BCE) loss function, the structured similarity (SSIM) loss function and the proposed regional augmentation (RA) loss function. The network with the multi-loss fusion method can maintain the integrity of the foreground region and enhance the regional pixel consistency. The algorithm is verified on five image datasets with multiple salient objects and complex backgrounds. Experimental results demonstrate that the algorithm effectively improves

**基金项目:**国家重点研发计划(2018YFD0400902);国家自然科学基金(61873112)。

This work was supported by the National Key Research and Development Program of China (2018YFD0400902), and the National Natural Science Foundation of China (61873112).

**收稿日期:**2021-04-20 **修回日期:**2021-06-07

the detection accuracy of saliency objects in complex scenes, and alleviates the problem of saliency map features missing and poor regional consistency.

**Key words:** convolutional neural network; salient object detection; feature hybrid enhancement; regional augmentation loss

显著性目标检测<sup>[1]</sup>旨在从输入图像上识别出最引人注目的对象。出于降低计算复杂度的目的,显著性目标检测<sup>[2]</sup>已经被广泛应用于语义分割<sup>[3]</sup>、图像检索<sup>[4]</sup>、目标识别<sup>[5]</sup>、场景分类<sup>[6]</sup>等领域。

图像显著性目标<sup>[7-9]</sup>的特征包括低层级基本特征与高层级语义特征,低层级基本特征主要包括纹理<sup>[10]</sup>、颜色<sup>[11]</sup>、对比度<sup>[12]</sup>等,高层级特征则揭示目标与背景之间内在的语义信息。许多传统的显著性目标检测方法<sup>[10-12]</sup>依赖于手工制作的特征和预定义的先验知识。这些方法利用图像的低层级信息以自底向上的方式进行检测。然而,由于这些先验信息在复杂的自然场景中往往是不准确的,导致了较差的检测结果。

最近,基于卷积神经网络的算法<sup>[13-18]</sup>通过利用高层级语义信息,已经取得了更优越的检测结果。Liu等<sup>[13]</sup>使用胶囊网络进行显著性目标检测,在模型中加入了部分物体关系,采用双流策略来实现胶囊网络,有效降低了噪声分配的可能性。Kuen等<sup>[14]</sup>提出了一个循环注意卷积-反卷积网络,利用空间变压器和循环网络单元,实现了显著图的逐步细化。Zhang等<sup>[15]</sup>提出了Amulet网络,首先将多层级特征整合到多个分辨率下,然后在每个分辨率下进行特征整合,最后进行显著性融合获得最终的显著图。Luo等<sup>[16]</sup>提出了一种非局部深度特征网络,使用多分辨率网格结构将局部信息与全局信息结合起来进行显著性目标检测,并且实现了一个边缘感知的损失函数。

由于深度学习的快速发展,这一领域已经取得了很大的成功,但是复杂场景下的显著性目标检测仍有以下挑战:(1)一方面由于显著性是基于图像全局定义的,算法必须整合全局信息;另一方面,精细的显著性目标检测,又要考虑局部、边缘等细节。当前的显著性目标检测方法并不能对这两方面进行有效的权衡。(2)当前许多显著性目标检测方法使用二进制交叉熵(binary cross-entropy, BCE)作为损失函数。但是交叉熵损失函数是像素级的,它不考虑邻域的像素值,因此不能有效地处理全局信息。这导致了区域一致性差和特征缺失的问题。

为了应对以上挑战,本文提出了一种特征混合增强与多损失融合的显著性目标检测算法。首先设计

上下文感知预测模块(context-aware prediction module, CAPM),它由多组编码器-解码器组成,并且为了有效提取全局信息嵌入了空间感知模块(spatial-aware module, SAM)。然后设计特征混合增强模块(feature hybrid enhancement module, FHEM),对预测模块产生的全局特征信息和细节特征信息进行有效的整合,并利用特征聚合模块(feature aggregation module, FAM)对整合的信息进行特征增强。最后提出了区域增强损失函数(regional augmentation, RA),通过结合二进制交叉熵损失函数和结构化相似度损失函数(structural similarity, SSIM)<sup>[19]</sup>,监督网络保持前景区域的完整性,增强区域像素一致性。

## 1 方法概述

本文提出区域增强网络,包含上下文感知预测模块和特征混合增强模块。上下文感知预测模块能够有效地提取全局信息和细节信息,获得初步的显著性特征图。特征混合增强模块可以将解码器得到的全局信息和细节信息进行充分的筛选提取,最终获得结构更加完整清晰的显著图。此外,网络中使用的混合损失函数可以引导网络学习更多前景细节信息,缓解特征缺失和增强区域像素的一致性。

### 1.1 上下文感知预测模块

上下文感知预测模块被设计成U型结构<sup>[20-22]</sup>,U型结构可以有效地获取低层和高层的信息。对于主干网络提取的多尺度信息,U型结构使用自顶向下的方式分为多个层级进行逐级融合。其中逐级融合的方法保证了上采样恢复出来的特征更精细,能更好地保留低层细节信息,多层级特征的融合保证了可以汇集多尺度信息,能更好地保留高层全局信息。网络结构图如图1所示。

图1中,六对编码器和解码器分别被记作 $E^{(l)}$ 和 $D^{(l)}$ ( $l \in \{1, 2, 3, 4, 5, 6\}$ )。

采用ResNet-34作为主干网络,去除其最后两个全连接层和最后一个池化层,并对输入层进行修改。ResNet模型首先用于图像分类任务,由于其优异的性能,广泛应用于各种计算机视觉任务中。由

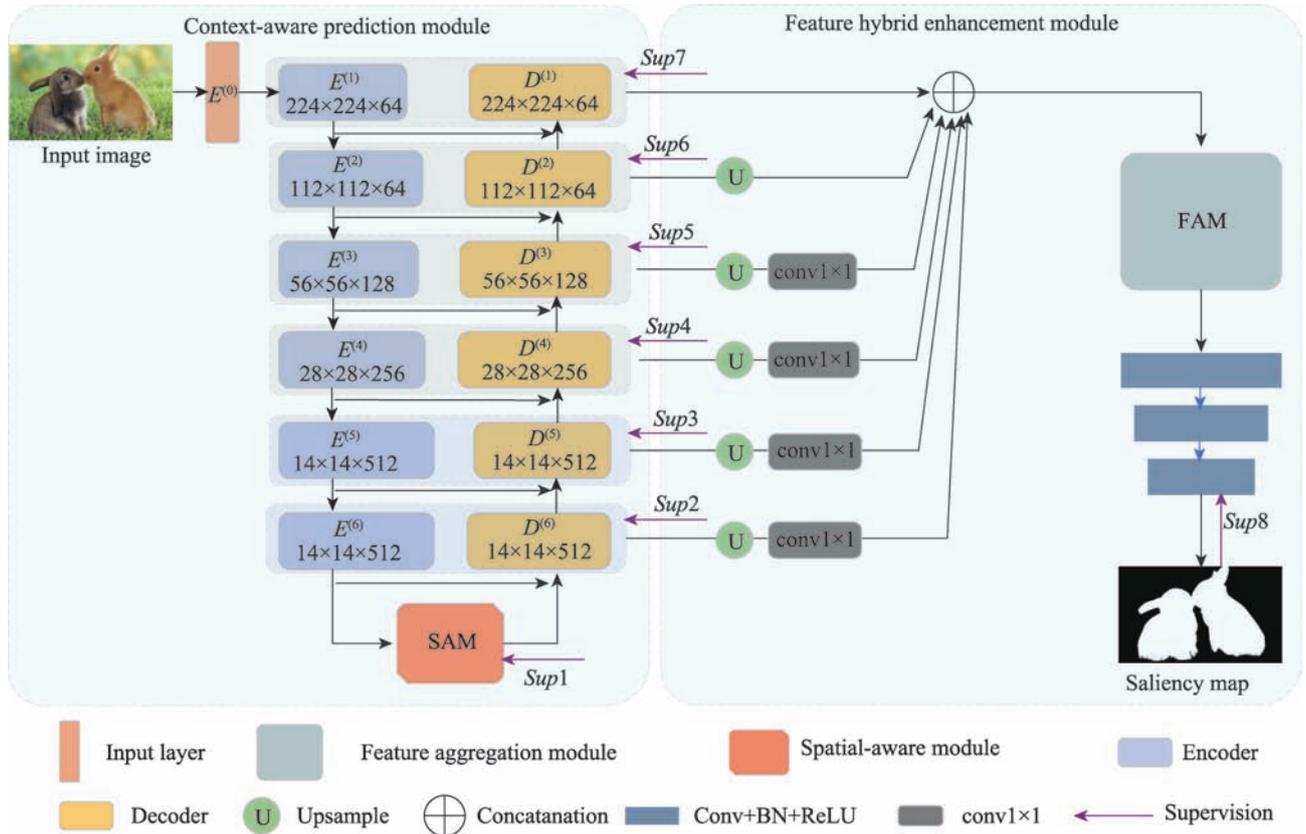


图1 显著性目标检测网络框架

Fig.1 Framework of salient object detection network

于ResNet的输入层使用的卷积核大小为 $7 \times 7$ , stride被设置为2,这引入了过多的参数,损失了图像的细节信息。因此使用64个大小为 $3 \times 3$ 的卷积核重新构造输入层 $E^{(0)}$ 。在ResNet-34这4组卷积块后进行池化操作pool4,此时特征图大小为 $14 \times 14$ 。此外,通过加入两组卷积块( $E^{(5)}$ 、 $E^{(6)}$ ),提高模型容量以充分提取高层级语义信息,这两组卷积块由3个拥有512个卷积核的普通残差块组成。为了提高显著图的清晰度,不再使用池化模块,保持特征图大小不再变化。

编码器与解码器连接处增加一个空间感知模块以进一步捕获全局语义信息,见图2。它包含3个膨胀率不同的 $3 \times 3$ 卷积层和一个 $1 \times 1$ 卷积层,在经过这些卷积层后产生4个大小相同的特征图,将其拼接,经过批归一化和ReLU激活函数后,得到和输入大小相同的特征图。

预测模块的解码器和编码器几乎完全对称,共有6个解码器( $D^{(1)} \sim D^{(6)}$ )。每个解码器由3个卷积块组成,其中每个卷积块依次由卷积层、批归一化层、激活函数层组成。每个解码器的输入由前一个阶段的解码器的输出与对应的编码器拼接产生。如果特

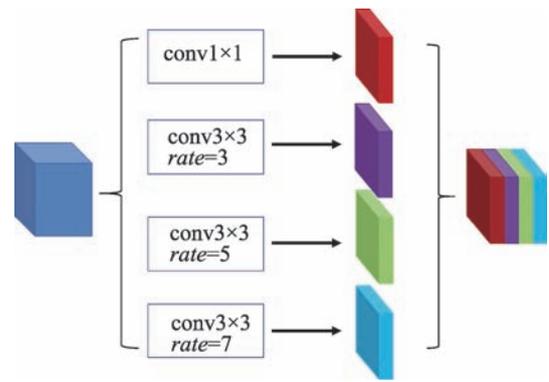


图2 空间感知模块

Fig.2 Spatial-aware module

征图大小不匹配,则对解码器产生的特征图进行上采样,使其大小一致。预测模块的编码器与解码器关系见式(1)。

$$D_i^{(l)} = \begin{cases} E_o^{(l)} \oplus U^{(l+1)}(D_o^{(l+1)}), & l = 1, 2, 3, 4 \\ E_o^{(l)} \oplus D_o^{(l+1)}, & l = 5 \\ E_o^{(l)} \oplus G_o^{(bridge)}, & l = 6 \end{cases} \quad (1)$$

其中, $E_o^{(l)}$ 、 $D_i^{(l)}$ 、 $D_o^{(l)}$ 分别指的是第 $l$ 个编码器的输出、解码器的输入、解码器的输出, $U^{(l+1)}$ 是对第 $l+1$

个解码器的输出进行  $2 \times$  上采样操作,  $G_o^{(\text{bridge})}$  是空间感知模块的输出,  $\oplus$  是拼接操作。

### 1.2 特征混合增强模块

为了对预测模块输出特征图的细节信息与全局信息进行充分的挖掘与整合,得到细节清晰、结构完整的显著图。在预测模块的解码器端引入特征混合增强模块(FHEM),利用特征聚合模块(FAM)充分聚合多尺度信息,并对得到的信息进行有效的整合,生成最终的显著图。

注意力机制是一种增强深度卷积神经网络性能的手段,一些研究<sup>[23-24]</sup>已经验证了注意力机制对显著性目标检测任务的有效性。但是他们所使用的是通道注意力(channel attention, CA)模块,激励函数使用的是全连接神经网络,这使得模型容易发生过拟合,影响模型的泛化性能,并且为了减少参数量进行了通道降维,这对于特征的深层次融合和细节保留是致命的。最近 Wang 等<sup>[25]</sup>提出了高效通道注意力(efficient channel attention, ECA)模块,使用一维卷积作为激励函数,极大地缓解了以上问题。因此使用 ECA 模块构建了特征聚合模块。

如图 1 所示,上下文感知预测模块输出 ( $D_o^{(1)} \sim D_o^{(6)}$ ) 的通道维数并不相同,且高层特征图存在大量冗余。使用  $1 \times 1$  卷积统一到  $D_o^{(1)}$  的维度数 64,然后将降维后的特征图进行拼接,得到特征图  $X \in \mathbb{R}^{W \times H \times C}$ 。

如图 3 所示,将  $X$  展开为  $[x_1, x_2, \dots, x_C]$ , 其中  $x_i \in \mathbb{R}^{W \times H}$  是  $X$  的第  $i$  层 ( $i$ -th) 特征图,  $C$  是通道总数。首先,对每个  $x_i$  使用全局平均池化得到通道级特征向量  $v \in \mathbb{R}^C$ 。之后,使用卷积核大小为  $k$  的一维卷积来捕获通道级依赖,  $k$  的取值见式(2)。

$$k = \psi(C) = \left\lfloor \frac{\lfloor bC \rfloor}{r} + \frac{b}{r} \right\rfloor_{\text{odd}} \quad (2)$$

其中,  $\lfloor \cdot \rfloor_{\text{odd}}$  表示距离  $t$  最近的奇数,设置  $r$  和  $b$  分别为 2 和 1。通过非线性映射  $\psi$ ,高维通道具有更大范围的相互作用,而低维通道具有更小范围的相互作用。

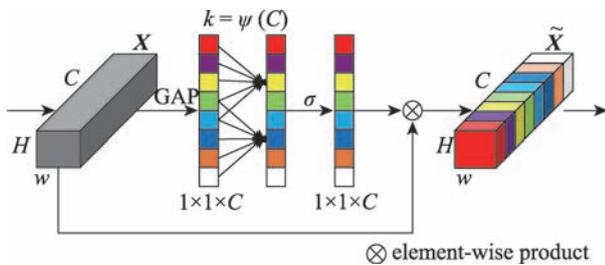


图3 高效通道注意力模块

Fig.3 Efficient channel attention module

然后通过使用 Sigmoid 运算,对经过编码的通道级特征向量进行归一化操作,将其映射到  $[0,1]$ 。

$$v' = \sigma(\varphi(v, W)) \quad (3)$$

其中,  $W$  指的是注意力模块的参数;  $\varphi$  指的是一维卷积运算;  $\sigma$  指的是 Sigmoid 运算。

用  $v'$  对特征图  $X$  进行加权得到输出  $\tilde{X}$ 。

$$\tilde{X} = v' \cdot X \quad (4)$$

然后对得到的融合增强特征进行 3 组卷积操作得到最终的显著预测图  $S$ , 卷积核个数分别为 256、64、32, 每组卷积同样依次由卷积层、批归一化层、激活函数层组成。

CA 模块所使用的激励函数为全连接网络。

$$f_{\{w_1, w_2\}}(X) = w_2 \text{ReLU}(w_1 X) \quad (5)$$

其中,  $w_1$  与  $w_2$  的参数量分别为  $C \times \frac{C}{a}$  和  $\frac{C}{a} \times C$ ,  $a$  为降维比例因子。与 ECA 模块激励函数一维卷积函数的参数量  $k$  相比,其比值  $\frac{2C^2}{ka}$ 。以本文算法为例,当  $C=512, k=3, a=16$  时, CA 模块激励函数的参数量远大于 ECA 模块激励函数的参数量。

### 1.3 损失函数

整体损失函数被定义为:

$$L = \sum_{k=1}^K \alpha_k \ell^{(k)} \quad (6)$$

其中,  $\ell^{(k)}$  指的是第  $k$  层的输出,  $K$  指的是总输出的个数,  $\alpha_k$  指的是每个损失的权值,全部设置为 1。该模型有 8 个输出,即  $K=8$ , 包括 7 个预测模块输出和 1 个增强模块输出。

定义  $\ell^{(k)}$  为混合损失:

$$\ell^{(k)} = \ell_{\text{bce}}^{(k)} + \ell_{\text{ssim}}^{(k)} + \ell_{\text{ra}}^{(k)} \quad (7)$$

其中,  $\ell_{\text{bce}}^{(k)}$  为 BCE 损失;  $\ell_{\text{ssim}}^{(k)}$  为 SSIM 损失;  $\ell_{\text{ra}}^{(k)}$  为 RA 损失。

BCE 损失是二值分类和分割中使用最广泛的损失:

$$\ell_{\text{bce}} = - \sum_{(r,c)} [G(r,c) \ln S(r,c) + (1 - G(r,c)) \ln (1 - S(r,c))] \quad (8)$$

其中,  $G(r,c) \in \{0,1\}$  是像素  $(r,c)$  真实图中的像素值,  $S(r,c) \in [0,1]$  是显著性目标像素的预测概率。

SSIM 损失起初是针对图像质量评价提出的方法,它可以捕捉到图像的结构信息,因此现在已经整合到训练损失中去学习图像真实图中的结构信息。

定义  $x = \{x_j; j = 1, 2, \dots, N^2\}$ ,  $y = \{y_j; j = 1, 2, \dots, N^2\}$  分别为显著性预测图和二进制真实图裁剪出的大小为

$N \times N$  的图片的像素值。

SSIM 损失被定义为：

$$\ell_{\text{ssim}} = 1 - \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (9)$$

其中,  $\mu_x$ 、 $\mu_y$  和  $\sigma_x$ 、 $\sigma_y$  分别是  $x$ 、 $y$  的均值和标准差,  $\sigma_{xy}$  是它们的协方差, 为了防止分母为 0, 设置  $C_1 = 0.01^2$  和  $C_2 = 0.03^2$ 。

BCE 损失函数是像素级的损失函数, 不考虑邻域的像素值, 对所有像素的权重相等, 有助于在所有像素上收敛, 但是容易导致特征缺失和区域一致性问题。IOU 损失函数是区域级的损失函数, 考虑的是各个区域的像素值, 因此可以增强各个区域的像素一致性, 但是 IOU 损失函数对前景和背景区域的权重相同, 同样容易导致特征缺失问题。因此, 为了增强区域内像素的一致性, 缓解特征缺失, 基于 IOU 损失函数, 本文提出一种新的损失函数——区域增强损失 (RA):

$$\ell_{\text{ra}} = 1 - \frac{\gamma|TP|}{\gamma|TP| + \alpha|FP| + \beta|FN|} \quad (10)$$

$$|TP| = \sum_{r=1}^H \sum_{c=1}^W S(r,c)G(r,c) \quad (11)$$

$$|FP| = \sum_{r=1}^H \sum_{c=1}^W S(r,c) - \sum_{r=1}^H \sum_{c=1}^W S(r,c)G(r,c) \quad (12)$$

$$|FN| = \sum_{r=1}^H \sum_{c=1}^W G(r,c) - \sum_{r=1}^H \sum_{c=1}^W S(r,c)G(r,c) \quad (13)$$

其中,  $|TP|$ 、 $|FP|$ 、 $|FN|$  分别表示真阳性、假阳性和假阴性区域像素值的和,  $\alpha$ 、 $\beta$  和  $\gamma$  分别用于控制  $|FP|$ 、 $|FN|$  和  $|TP|$  的惩罚程度,  $\alpha + \beta = 1$ 。  $G(r,c) \in \{0,1\}$  是像素值  $(r,c)$  的真实图标签,  $S(r,c) \in [0,1]$  是显著性目标的预测概率。当  $\alpha = \beta = \gamma = 1$ , 该损失退化为 IOU 损失。通过增大  $\gamma$ , 调整  $\beta$  大于  $\alpha$ , RA 损失函数可以控制不同区域的惩罚程度, 减少显著图的前景缺失。

## 2 实验与分析

### 2.1 数据集

为了评估本文算法的有效性, 在 5 个数据集 ECSSD、DUT-OMRON、HKU-IS、DUTS、SOD 上评估了本文的算法。ECSSD 包含 1 000 张有着不同尺寸目标的复杂图片。DUT-OMRON 是一个包含 5 168 张图片的有挑战性的数据集, 每张图片中有一个或两个对象, 大多数前景对象在结构上很复杂。HKU-IS 包含 4 447 张图片, 每张图片包含多个显著性物

体, 这些物体的色彩对比度低且与图像边界存在重叠现象。DUTS 是目前最大的显著性检测基准数据集, 它包含 10 553 张训练图片 (DUTS-TR) 和 5 019 张测试图片 (DUTS-TE)。SOD 包含 300 张图片, 最初用于分割领域, 这些图片大多数包含多个显著性对象。

### 2.2 实现细节

本文选择 Pytorch0.4.0 深度学习框架来实现显著性目标检测网络。全部输入图片的大小被调整为  $256 \times 256$ , 然后随机裁剪为  $224 \times 224$ 。为了提高模型的泛化性能, 使用的数据增强手段有: 随机裁剪、水平翻转。使用 ResNet-34 初始化部分网络参数, 其余参数使用 Xavier 方法初始化。使用 Adam 优化器来训练网络, 超参数设置使用默认值, 其中批次大小为 16, 学习率为 0.001,  $\beta_1$  为 0.9,  $\beta_2$  为 0.999, 正则化系数为 0。在不使用验证集的情况下训练网络, 训练损失函数在 500 次迭代后收敛。在测试过程中, 输入图像的大小将调整为  $256 \times 256$  获得其显著图。然后, 将其调整为输入的原始大小输出。训练该网络使用的设备为 E5-2678v3 和 NVIDIA GTX 1080Ti GPU  $\times 4$ 。

### 2.3 评价指标

本文使用三种最常用的指标来评价本文算法: 准确率-召回率曲线 (precision-recall curve, PR)、MAE (mean absolute error)、F-measure<sup>[26]</sup> ( $F_\beta$ )。

准确率-召回率曲线: 首先二值化固定阈值的灰度预测图, 然后生成的二值图和真实图用于计算准确率和召回率, 准确率与召回率计算方式为:

$$\text{precision} = \frac{TP}{TP + FP}, \text{recall} = \frac{TP}{TP + FN} \quad (14)$$

其中,  $TP$ 、 $FP$  和  $FN$  分别代表真阳性、假阳性和假阴性。准确率-召回率曲线通过一组成对的准确率、召回率按照阈值从 0~255 绘制生成, 准确率-召回率曲线与坐标轴围成的面积越大, 性能越好。

F-measure:  $F_\beta$  由准确率和召回率加权平均生成。

$$F_\beta = \frac{(1 + \beta^2)\text{precision} \times \text{recall}}{\beta^2 \text{precision} + \text{recall}} \quad (15)$$

为了强调准确率的重要性<sup>[26]</sup>, 其中  $\beta^2$  通常被设置为 0.3。

MAE: 计算归一化的显著图和真实图之间的平均绝对误差。

$$\text{MAE} = \frac{1}{H \times W} \sum_{r=1}^H \sum_{c=1}^W |S(r,c) - G(r,c)| \quad (16)$$

其中,  $S$  和  $G$  分别指的是显著图和对应的真实图,  $H$

和  $W$  分别表示显著图的高和宽,  $(r, c)$  指的是对应的像素值。

### 2.4 消融实验

本节逐步分解所提出的方法来揭示每个组件对显著图的贡献。本节所有实验均在 DUT-OMRON 数据集上进行。

#### 2.4.1 结构消融

为了验证本文提出的区域增强网络的有效性, 本文进行了相关的结构消融实验, 表 1 列出了基线网络 U-Net<sup>[22]</sup>、预测模块和增强模块的定量比较结果。如表 1 所示, 本文提出的预测模块在  $\max F_\beta$  (数值越大越好) 和 MAE (数值越小越好) 这两个指标上分别比基线网络提升 0.042、0.021。本文提出的区域增强网络, 在  $\max F_\beta$  和 MAE 这两个指标上与预测模块相比分别提升了 0.019、0.011。为了展现添加不同模块产生的实际效果, 本文进行了更精细的视觉比较。如图 4 所示, 未加入 SAM 模块的预测模块产生的显著图部分与整体存在割裂, 加入了 SAM 的预测模块能够初步进行空间上下文整合, 但是仍然存在特征缺失, 当加入增强模块对预测模块产生的特征图进行增强后, 整个显著图变得完整。此外, 实验过程中尝试使用 CA 模块搭建增强模块, 没有成功。如表 1 所示, ECA 模块在使用较少参数量的情况下取得的结果优于 CA 模块。

表 1 算法使用不同模块性能比较  
Table 1 Performance comparison of different modules in algorithm

Configurations	$\max F_\beta$	MAE
Baseline U-Net	0.742	0.089
CAPM	0.784	0.068
CAPM+FHEM(CA)	0.773	0.072
CAPM+FHEM(ECA)	<b>0.803</b>	<b>0.057</b>

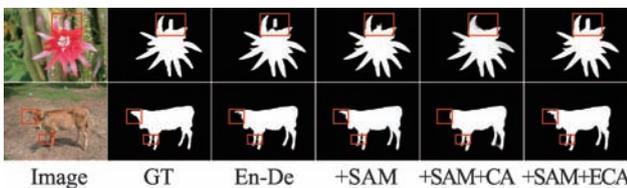


图 4 不同模块显著性检测结果比较

Fig.4 Comparison of saliency detection results of different modules

#### 2.4.2 损失消融

为了证明混合损失的有效性, 基于区域增强网

络, 对不同的损失进行了消融实验。表 2 中的结果表明, 使用混合损失  $\ell_{bsr}$  的网络在  $\max F_\beta$  和 MAE 这两个指标上与使用  $\ell_{bce}$  损失相比分别提升了 0.014、0.012。如图 5 所示, 与使用  $\ell_{bce}$  损失相比, 使用  $\ell_{bsi}$  损失有助于约束前景和背景的像素值, 产生的显著图不易出现大面积的模糊现象。相比使用  $\ell_{bsi}$  损失, 使用  $\ell_{bsr}$  损失产生的显著图更完整, 区域一致性更好。

表 2 算法使用不同损失的性能比较

Table 2 Performance comparison of different losses in algorithm

Configurations	$\max F_\beta$	MAE
CAPM+FHEM+ $\ell_{bce}$	0.789	0.069
CAPM+FHEM+ $\ell_{bs}$	0.787	0.064
CAPM+FHEM+ $\ell_{bsi}$	0.792	0.066
CAPM+FHEM+ $\ell_{bsr}$	<b>0.803</b>	<b>0.057</b>

注:  $\ell_{bsi} = \ell_{bce} + \ell_{ssim} + \ell_{iou}$ ,  $\ell_{bsr} = \ell_{bce} + \ell_{ssim} + \ell_{ra}$ 。

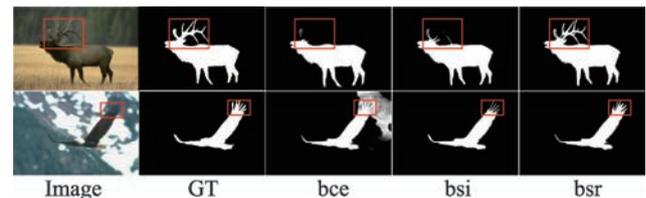


图 5 不同损失显著性检测结果比较

Fig.5 Comparison of saliency detection results of different losses

#### 2.4.3 RA 损失系数实验

为了确定 RA 损失的最佳参数选择, 基于区域增强网络, 采用不同的系数进行了实验。如表 3 所示, 当  $\alpha = 0.3$ ,  $\beta = 0.7$ ,  $\gamma = 2.0$  时,  $\max F_\beta$  与 MAE 取得优值, 相较于 IOU 损失 ( $\alpha = 1.0$ ,  $\beta = 1.0$ ,  $\gamma = 1.0$ ) 分别提升 1.1%、0.9%。

表 3 算法使用不同系数 RA 损失的性能比较

Table 3 Performance comparison of RA loss with different coefficients in algorithm

$\alpha$	$\beta$	$\gamma$	$\max F_\beta$	MAE
1.0	1.0	1.0	0.792	0.066
1.0	1.0	2.0	0.797	0.063
0.1	0.9	2.0	0.756	0.076
0.2	0.8	2.0	0.762	0.073
<b>0.3</b>	<b>0.7</b>	<b>2.0</b>	<b>0.803</b>	<b>0.057</b>
0.4	0.6	2.0	0.801	0.058
0.5	0.5	2.0	0.796	0.061

## 2.5 性能比较

实验过程中与其他9种先进算法 CapsNet<sup>[13]</sup>、RAS<sup>[14]</sup>、Amulet<sup>[15]</sup>、NLDF<sup>[16]</sup>、UCF<sup>[17]</sup>、RFCN<sup>[18]</sup>、ELD<sup>[27]</sup>、DCL<sup>[28]</sup>、LEGS<sup>[29]</sup>进行比较。为了体现公平性,评测结果来自原始论文或者原作者提供的显著图。

### 2.5.1 定量比较

图6为本文算法与其他显著性目标检测算法在PR曲线与F-measure曲线上的比较结果。表4列出了本文算法与其他9种先进算法的  $\max F_\beta$  和MAE性能

指标比较结果。

从结果可以看出,本文算法显著优于其他算法。具体的,对于评价指标  $\max F_\beta$ ,本文算法比 CapsNet、RAS 和 Amulet 在5个数据集上平均分别提高了0.066 0、0.017 8、0.050 2。对于评价指标 MAE,本文算法比 CapsNet、RAS 和 Amulet 在5个数据集上平均分别降低了0.008 0、0.014 8、0.032 2。

### 2.5.2 定性比较

显著性目标检测的研究来源于对人类视觉系统

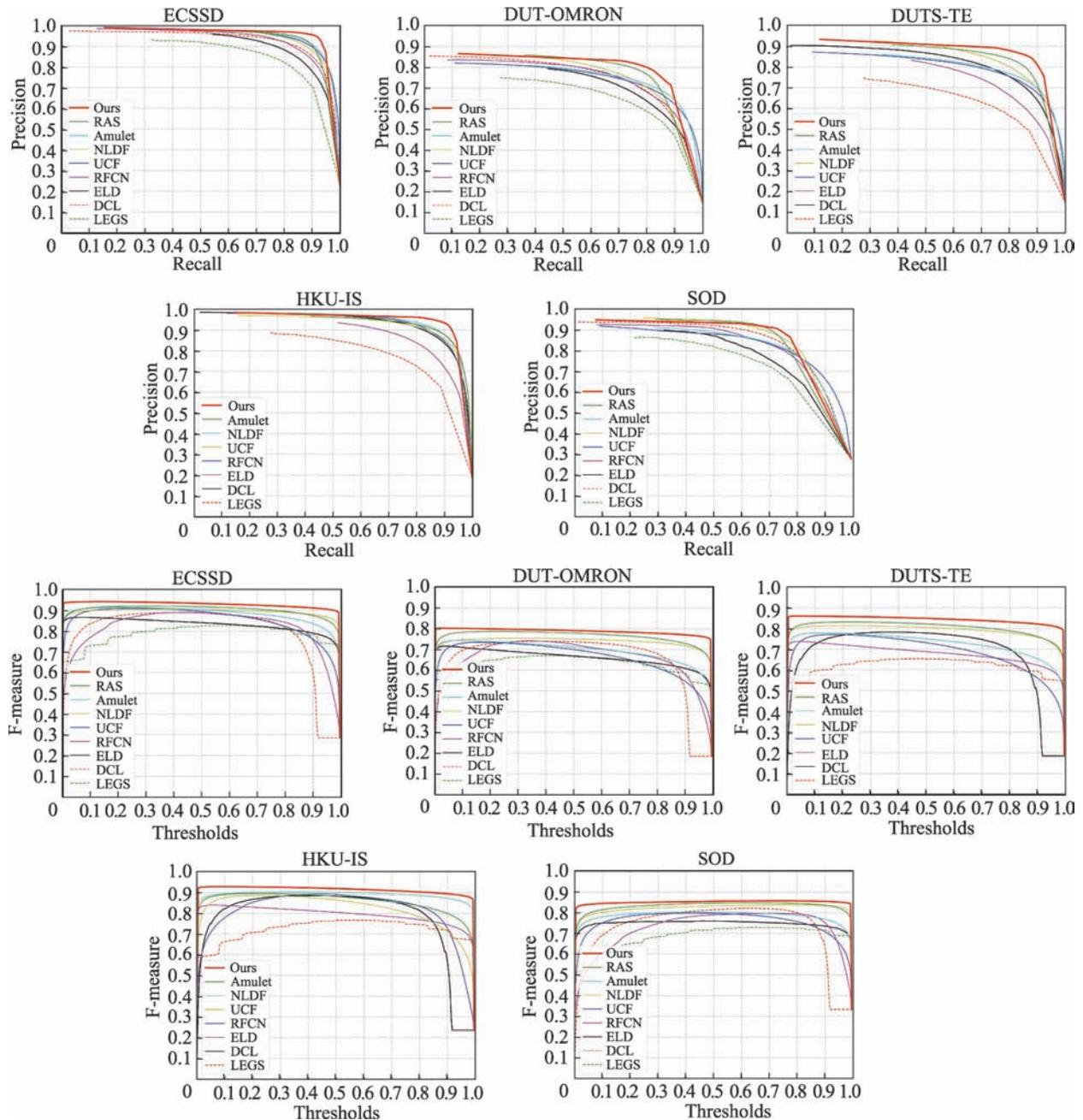


图6 不同算法的PR曲线与F-measure曲线比较

Fig.6 Comparison of PR curves and F-measure curves of different algorithms

表4 在5个数据集上不同算法性能比较

Table 4 Performance comparison of different algorithms on 5 datasets

Model	ECSSD		DUT-OMRON		DUTS-TE		HKU-IS		SOD	
	$\max F_\beta$	MAE								
Ours	<b>0.942</b>	<b>0.036</b>	<b>0.803</b>	<b>0.057</b>	<b>0.860</b>	<b>0.046</b>	<b>0.928</b>	<b>0.031</b>	<b>0.857</b>	<b>0.102</b>
CapsNet	0.887	<u>0.052</u>	0.703	<u>0.063</u>	0.799	<u>0.048</u>	0.880	<u>0.039</u>	—	—
RAS	<u>0.921</u>	<u>0.056</u>	<u>0.786</u>	<u>0.062</u>	<u>0.831</u>	<u>0.059</u>	<u>0.913</u>	<u>0.045</u>	<u>0.850</u>	<u>0.124</u>
Amulet	<u>0.915</u>	0.059	0.743	0.098	0.778	0.084	0.897	0.051	0.806	<u>0.141</u>
NLDF	0.905	0.063	<u>0.753</u>	0.080	<u>0.812</u>	0.066	<u>0.902</u>	0.048	<u>0.841</u>	<u>0.124</u>
UCF	0.911	0.078	0.734	0.132	0.771	0.117	0.886	0.074	0.803	0.164
RFCN	0.890	0.107	0.742	0.111	0.784	0.091	0.892	0.079	0.799	0.170
ELD	0.867	0.079	0.715	0.092	0.738	0.093	0.839	0.074	0.764	0.155
DCL	0.890	0.143	0.739	0.097	0.782	0.088	0.885	0.072	0.823	<u>0.141</u>
LEGS	0.827	0.118	0.669	0.133	0.655	0.138	0.766	0.119	0.734	0.196

的模拟,检测结果的优劣依赖于视觉观感。因此为了进一步阐述本文算法的优越性,图7展示了本文算法与5种先进算法的定性比较结果。

图7第一行具有小目标的自然图像检测结果显示,本文算法能够完整地分割出图像中的小目标,并

且相比其他算法背景的像素更一致;第二、三行具有大目标的图像检测结果显示本文算法能更加完备地检测出航母舰岛和城堡顶部的细节;第四行低对比度图像检测结果显示本文算法能够更好地整合图像全局信息,不会漏检显著性物体;第五行具有复杂边

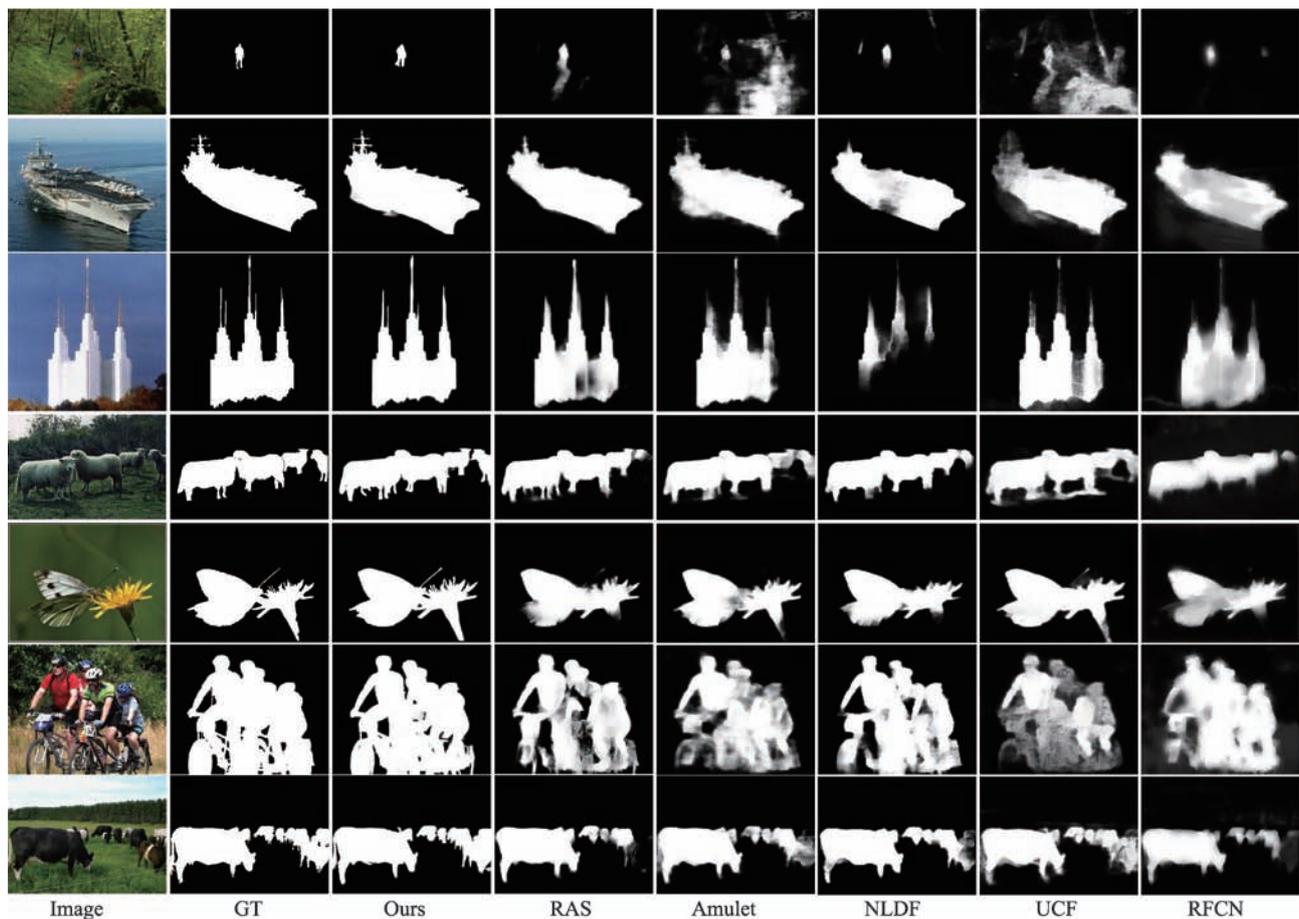


图7 不同算法的显著性目标检测结果对比

Fig.7 Comparison of saliency detection results of different algorithms

缘的图像检测结果显示本文算法能够更好地提取图像细节信息,产生的显著图边缘细节更丰富;第六、七行具有多目标粘连的图像检测结果显示本文算法能够更好地凸出前景和增强区域一致性。

综合来看,本文算法能够充分凸显前景并且抑制背景。通过本文算法可以保持前景的区域像素一致,其内部不会有较大的灰色区域,与显著性物体相似的背景被充分抑制,显著性物体的边缘不会出现大量的模糊现象。

### 3 结论

本文针对显著性目标检测中的特征缺失和区域一致性差的问题,基于全卷积神经网络提出一种特征混合增强与多损失融合的显著性目标检测算法。通过使用编码器-解码器网络初步捕获高层全局信息和低层细节信息,然后使用空间感知模块进一步捕获全局语义信息,最后使用特征混合增强模块将预测模块提取的特征信息做增强处理。同时提出了区域增强损失函数,通过结合二进制交叉熵损失函数和结构化相似度损失函数,以多损失融合的方式监督网络保持前景区域的完整性以及增强区域像素一致性。与其他的基于深度学习的显著性目标检测算法相比,显著目标的边界细节、前景完整度和区域一致性有明显的改善,各项指标有较大的提升。

### 参考文献:

- [1] 凌艳, 陈莹. 多尺度上下文信息增强的显著目标检测全卷积网络[J]. 计算机辅助设计与图形学学报, 2019, 31(11): 2007-2016.  
LING Y, CHEN Y. Salient object detection with multiscale context enhanced fully convolutional network[J]. Journal of Computer-Aided Design & Computer Graphics, 2019, 31(11): 2007-2016.
- [2] 张守东, 杨明, 胡太. 基于多特征融合的显著性目标检测算法[J]. 计算机科学与探索, 2019, 13(5): 834-845.  
ZHANG S D, YANG M, HU T. Salient object detection algorithm based on multi-feature fusion[J]. Journal of Frontiers of Computer Science and Technology, 2019, 13(5): 834-845.
- [3] OH S J, BENENSON R, KHOREVA A, et al. Exploiting saliency for object segmentation from image level labels[C]// Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, Jul 22-25, 2017. Washington: IEEE Computer Society, 2017: 5038-5047.
- [4] GAO Y, WANG M, TAO D C, et al. 3-D object retrieval and recognition with hypergraph analysis[J]. IEEE Transactions on Image Processing, 2012, 21(9): 4290-4303.
- [5] REN Z X, GAO S H, 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.
- [6] ZHANG F, DU B, ZHANG L P. Saliency-guided unsupervised feature learning for scene classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2015, 53(4): 2175-2184.
- [7] 于明, 李博昭, 于洋, 等. 基于多图流形排序的图像显著性检测[J]. 自动化学报, 2019, 45(3): 577-592.  
YU M, LI B Z, YU Y, et al. Image saliency detection with multi-graph model and manifold ranking[J]. Acta Automatica Sinica, 2019, 45(3): 577-592.
- [8] 常振, 段先华, 鲁文超, 等. 基于多尺度的贝叶斯模型显著性检测[J]. 计算机工程与应用, 2020, 56(11): 207-213.  
CHANG Z, DUAN X H, LU W C, et al. Multi-scale saliency detection based on Bayesian framework[J]. Computer Engineering and Applications, 2020, 56(11): 207-213.
- [9] 时斐斐, 张松龙, 彭力. 结合边缘特征先验引导的深度卷积显著性检测[J]. 计算机工程与应用, 2020, 56(14): 199-206.  
SHI F F, ZHANG S L, PENG L. Deep convolution saliency detection combined with edge feature prior-based inspection[J]. Computer Engineering and Applications, 2020, 56(14): 199-206.
- [10] SCHARFENBERGER C, WONG A, FERGANI K, et al. Statistical textural distinctiveness for salient region detection in natural images[C]//Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition, Portland, Jun 23-28, 2013. Washington: IEEE Computer Society, 2013: 979-986.
- [11] KIM J H, HAN D Y, TAI Y W, et al. Salient region detection via high-dimensional color transform[C]//Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, Jun 24-27, 2014. Washington: IEEE Computer Society, 2014: 883-890.
- [12] ZHU W J, LIANG S, WEI Y C, et al. Saliency optimization from robust background detection[C]//Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, Jun 24-27, 2014. Washington: IEEE Computer Society, 2014: 2814-2821.
- [13] LIU Y, ZHANG Q, ZHANG D W, et al. Employing deep part-object relationships for salient object detection[C]//Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision, Seoul, Oct 27-Nov 2, 2019. Piscataway: IEEE, 2019: 1232-1241.
- [14] KUEN J, WANG Z H, WANG G. Recurrent attentional net-

- works for saliency detection[C]//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, Jun 26- Jul 1, 2016. Washington: IEEE Computer Society, 2016: 3668-3677.
- [15] ZHANG P P, WANG D, LU H C, et al. Amulet: aggregating multi-level convolutional features for salient object detection[C]//Proceedings of the 2017 IEEE International Conference on Computer Vision, Venice, Oct 22-29, 2017. Washington: IEEE Computer Society, 2017: 202-211.
- [16] LUO Z M, MISHRA A K, ACHKAR A, et al. Non-local deep features for salient object detection[C]//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, Jul 21-26, 2017. Washington: IEEE Computer Society, 2017: 6593-6601.
- [17] ZHANG P P, WANG D, LU H C, et al. Learning uncertain convolutional features for accurate saliency detection[C]//Proceedings of the 2017 IEEE International Conference on Computer Vision, Venice, Oct 22-29, 2017. Washington: IEEE Computer Society, 2017: 212-221.
- [18] WANG L Z, WANG L J, LU H C, et al. Salient object detection with recurrent fully convolutional networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 41(7): 1734-1746.
- [19] WANG Z, SIMONCELLI E P, BOVIK A C. Multiscale structural similarity for image quality assessment[C]//Proceedings of the 2003 IEEE Asilomar Conference on Signals, Pacific Grove, Nov 9-12, 2003. Washington: IEEE Computer Society, 2003.
- [20] FENG M Y, LU H C, DING E. Attentive feedback network for boundary-aware salient object detection[C]//Proceedings of the 2019 IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, Jun 16-20, 2019. Piscataway: IEEE, 2019: 1623-1632.
- [21] QIN X B, ZHANG Z C, HUANG C Y, et al. BASNet: boundary-aware salient object detection[C]//Proceedings of the 2019 IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, Jun 16-20, 2019. Piscataway: IEEE, 2019: 7471-7481.
- [22] RONNEBERGER O, FISCHER P, BROX T. U-Net: convolutional networks for biomedical image segmentation[C]//Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Oct 5-9, 2015. Cham: Springer, 2015: 234-241.
- [23] ZHAO T, WU X Q. Pyramid feature attention network for saliency detection[C]//Proceedings of the 2019 IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, Jun 16-20, 2019. Piscataway: IEEE, 2019: 3080-3089.
- [24] ZHANG L H, WU J, WANG T T, et al. A multistage refinement network for salient object detection[J]. IEEE Transactions on Image Processing, 2020, 29: 3534-3545.
- [25] WANG Q L, WU B G, ZHU P F, et al. ECA-Net: efficient channel attention for deep convolutional neural networks [C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, Jun 13-19, 2020. Piscataway: IEEE, 2020: 11531-11539.
- [26] ACHANTA R, HEMAMI S, ESTRADA F, et al. Frequency-tuned salient region detection[C]//Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Florida, Jun 20-25, 2009. Washington: IEEE Computer Society, 2009: 1597-1604.
- [27] LEE G, TAI Y W, KIM J. Deep saliency with encoded low level distance map and high level features[C]//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, Jun 26-Jul 1, 2016. Washington: IEEE Computer Society, 2016: 660-668.
- [28] LI G B, YU Y Z. Deep contrast learning for salient object detection[C]//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, Jun 26-Jul 1, 2016. Washington: IEEE Computer Society, 2016: 478-487.
- [29] WANG L J, LU H C, RUAN X, et al. Deep networks for saliency detection via local estimation and global search[C]//Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition, Boston, Jun 7-12, 2015. Washington: IEEE Computer Society, 2015: 3183-3192.



**李春标**(1997—),男,山东聊城人,硕士研究生,主要研究方向为显著性目标检测、深度学习。  
**LI Chunbiao**, born in 1997, M.S. candidate. His research interests include salient object detection and deep learning.



**谢林柏**(1973—),男,湖南永州人,博士,教授,博士生导师,CAA会员,主要研究方向为过程建模与控制、智能检测、系统安全性。  
**XIE Linbo**, born in 1973, Ph.D., professor, Ph.D. supervisor, member of CAA. His research interests include process modeling and control, intelligent detection and system safety.



**彭力**(1967—),男,河北唐山人,博士,教授,博士生导师,CAAI会员,CCF会员,主要研究方向为视觉物联网、行为识别、深度学习。  
**PENG Li**, born in 1967, Ph.D., professor, Ph.D. supervisor, member of CAAI and CCF. His research interests include visual Internet of things, action recognition and deep learning.