

基于特征融合与边界修正显著性目标检测

陈慧, 彭力

引用本文

陈慧, 彭力. 基于特征融合与边界修正显著性目标检测[J]. 计算机科学, 2023, 50(12): 166-174.

CHEN Hui, PENG Li. [Feature Fusion and Boundary Correction Network for Salient Object Detection](#)[J].

Computer Science, 2023, 50(12): 166-174.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[使用RAP生成可传输的对抗网络流量](#)

Generate Transferable Adversarial Network Traffic Using Reversible Adversarial Padding

计算机科学, 2023, 50(12): 359-367. <https://doi.org/10.11896/jsjcx.221000155>

[融合关系传递信息的双图文档级关系抽取方法](#)

Method of Document Level Relation Extraction Based on Fusion of Relational Transfer Information
Using Double Graph

计算机科学, 2023, 50(12): 229-235. <https://doi.org/10.11896/jsjcx.230500010>

[基于迭代非对称盲点网络的低剂量CT重建算法](#)

Low-dose CT Reconstruction Algorithm Based on Iterative Asymmetric Blind Spot Network

计算机科学, 2023, 50(12): 221-228. <https://doi.org/10.11896/jsjcx.230300014>

[面向工业图像异常检测的连续密集标准化流模型](#)

Continuous Dense Normalized Flow Model for Anomaly Detection in Industrial Images

计算机科学, 2023, 50(12): 212-220. <https://doi.org/10.11896/jsjcx.221000183>

[基于双空间共轭自编码器的多时相高光谱异常变化检测](#)

Multi-temporal Hyperspectral Anomaly Change Detection Based on Dual Space Conjugate
Autoencoder

计算机科学, 2023, 50(12): 175-184. <https://doi.org/10.11896/jsjcx.221100092>

基于特征融合与边界修正显著性目标检测

陈慧^{1,2} 彭力¹

1 江南大学物联网工程学院 江苏 无锡 214121

2 无锡职业技术学院物联网技术学院 江苏 无锡 214121

(13489959@qq.com)

摘要 显著性目标检测旨在寻找图像中的视觉显著区域。现有的显著性目标检测方法已经展现出强大的优势,但依然在尺度感知和边界预测方面具有局限性。首先,各类场景中的显著目标存在诸多尺度,使算法难以适应不同尺度变化。其次,显著目标往往具有复杂的轮廓,这使边界像素点的检测变得更为困难。针对以上问题,文中提出了基于特征融合与边界修正的显著性目标检测网络,该网络基于特征金字塔,提取了不同层次显著特征。首先针对目标的尺度多样性设计了由多尺度特征解码模块组成的特征融合解码器,通过逐层融合相邻层特征,提高了网络对目标尺度的感知能力。同时设计了边界修正模块学习显著目标的轮廓特征,以生成边界清晰的高质量显著图。在5个常用显著性目标检测数据集上进行实验,结果表明所提算法在平均绝对误差、F指标和S指标3项定量指标上均能取得较优的结果。

关键词: 显著性目标检测;深度学习;卷积神经网络;特征融合;边界修正

中图分类号 TP391.4

Feature Fusion and Boundary Correction Network for Salient Object Detection

CHEN Hui^{1,2} and PENG Li¹

1 School of Internet of Things Engineering, Jiangnan University, Wuxi, Jiangsu 214121, China

2 College of Internet of Things of Technology, Wuxi Institute of Technology, Wuxi, Jiangsu 214121, China

Abstract Saliency object detection aims to find visually significant areas in an image. Existing salient object detection methods have shown strong advantages, but they are still limited by scale perception and boundary prediction. First of all, there are many scales of salient objects in various scenes, which makes it difficult for the algorithm adapt to different scale changes. Secondly, salient objects often have complex contours, which makes detection of boundary pixels more difficult. To solve these problems, this paper proposes a feature fusion and boundary correction network for salient object detection. This network extracts salient features at different levels on the feature pyramid. Firstly, a feature fusion decoder composed of multi-scale feature decoding modules is designed for the scale diversity of the object. By fusing the features of adjacent layer by layer, the network's ability to perceive the scale is improved. At the same time, a boundary correction module is designed to learn the contour features of salient objects to generate high quality salient images with clear boundaries. Experimental results on five commonly used salient object detection datasets show that the proposed algorithm can achieve better results on the average absolute error, F index and S index.

Keywords Salient object detection, Deep learning, Convolutional neural network, Feature fusion, Boundary correction

1 引言

显著性目标检测旨在寻找图像在视觉上最重要的目标或区域,是目标检测^[1]、语义分割^[2]以及目标跟踪^[3]等计算机视觉任务的基础工作。

传统的显著性目标检测算法使用手工提取特征,例如对比度先验^[4]、前景先验^[5]和背景先验^[6]等,这些方法在处理复杂的场景和目标时难以充分提取特征,检测结果往往不尽人意。近年来,随着卷积神经网络在特征学习方面展现出了巨大的优势,显著性目标检测算法开始演变为基于卷积神经网络的深度学习方法,其大多沿用了全卷积网络(Fully Convo-

lutional Network, FCN)^[7]和特征金字塔网络(Feature Pyramid Network, FPN)^[8]结构,同时设计了各类融合策略来充分学习并有效整合不同层次下的图片信息。

尽管基于卷积神经网络和特征融合的显著性目标检测算法^[9-10]能够有效地学习到目标的深层语义信息,但是依然存在一些问题。首先,待检测的各类复杂场景中目标往往尺度不一,如图1中的动物、人、植物等形态各异,这需要算法能够更好地适应不同目标的尺度变化。此外,各类目标往往具有复杂的边界,如图2中的动物、船以及椅子。有效地区分前景并准确界定目标边界有助于进一步提升检测算法的性能。



图1 尺度多变问题

Fig. 1 Problem of variable scale

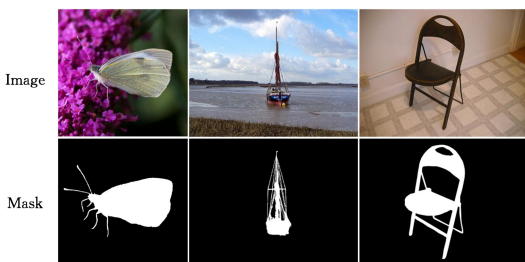


图2 边界复杂问题

Fig. 2 Problem of complex boundary

为了有效解决尺度多变和边界复杂问题,本文提出了基于特征融合与边界修正网络(Feature Fusion and Boundary Correction Network, FFBCN)。网络基于编解码的结构,采用特征金字塔作为编码器来提取不同层次的显著特征。其中,针对目标的尺度多样性,设计了特征融合解码器(Feature Fusion Decoder, FFD),用于多次融合相邻层次特征,以适应各类目标的尺度变化,同时提出组成解码器的多尺度解码模块(Multi-scale Decoding Module, MDM)来充分融合相邻层次的显著特征。针对边界复杂问题,本文还设计了边界修正模块(Boundary Correction Module, BCM),其通过独立监督编码器在各个阶段的目标边界学习情况,来提高网络对显著目标边界的关注度,以更好地划分前背景区域,提升模型的检测性能。本文的主要贡献如下:

1) 针对大跨度特征融合造成的语义污染问题,设计了特征融合解码器来逐层融合相邻特征,避免了跨层特征差异过大对最终检测的不利影响,解决了尺度多变问题。

2) 设计了多尺度解码模块,用于组成解码器,充分融合相邻层次特征,保留特征细化信息。

3) 针对复杂场景及目标边界的预测模糊问题,设计了边界修正模块,在定位显著性目标的同时提高网络对边界预测的关注度,以便更好地区分前背景,解决边界复杂问题,并生成更清晰准确的显著图。

2 相关工作

显著性目标检测分为传统的手工提取特征方法和现阶段主流的基于神经网络的方法。其中传统的显著性目标检测方法往往通过启发性假设,如对比度先验、中心先验、边界先验等方式^[4,6],来提取显著目标的低级特征。而基于神经网络的方法采用深层网络来挖掘目标的深层语义信息,弥补了传统方法在提取高级特征方面的不足。当前流行的显著性目标检测算法基于FCN跳跃连接、FPN^[8]提取多层次特征以及SegNet^[11]编解码结构。

2.1 特征融合

特征融合旨在结合不同层次的显著特征,帮助网络更高效地检测不同尺度的显著目标。PoolNet^[12]聚合了多层次的特征图,并将其用于解码阶段,保留了特征的细节信息。ITSD^[13]采用了双解码结构抑制多层次特征的丢失。EDN^[14]构建了尺度金字塔,通过特征融合恢复深层次特征的细节信息。DPNet^[15]设计了双向解码器,通过其在特征尺度与动态集合之间的路由能力,进行尺度感知。这些方法均使用多个多级跳跃连接来由下往上逐步合并不同层次的特征,但是忽视了包含轮廓和纹理等低级语义信息的底层特征和蕴含着丰富语义信息的高层特征之间较大的差异。如图3所示,F1-F4代表了不同层次的特征图,如果直接将这些特征图跨级结合,则可能导致最终的预测图中出现较多的噪点,这不仅降低了图像质量,还会影响最终的检测精度。本文采用相邻层特征单次高效融合与多次充分融合的方式,在充分利用不同层次显著特征的前提下减少细节丢失,帮助网络更有效地感知多尺度。

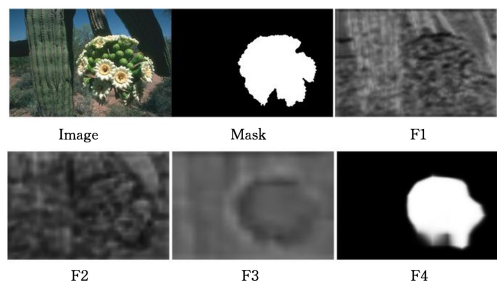


图3 不同层次的特征图

Fig. 3 Feature maps from different levels

2.2 边界感知

有效地区分前背景并准确界定目标边界能够在一定程度上提升检测性能。为了帮助网络学习显著目标的边界信息,BASNet^[16]使用结构相似性指标监督目标边界的学习情况,使网络具有了边界感知能力。F3Net^[17]提出边界权重系数,将其与交叉熵和交并比损失结合,使网络更关注边界像素点的预测情况。Contour-aware^[18]损失指导模型感知显著物体轮廓,捕获局部范围内的上下文细节,从而有效划分显著区域边界。此类通过设计损失函数来提升网络对边界学习关注度的方法在应对复杂边界的情况时具有明显作用,但会忽略网络对其他显著特征的关注。因此,本文设计了单独的网络模块来监督不同编码阶段下的目标边界预测,以提高对目标轮廓学习的关注度,生成边界更清晰的显著图。

本文的特征融合与边界修正网络所提的由多尺度解码模块组成的特征融合解码器和边界修正模块,既弥补了现有方法的缺陷,又提高了检测性能。

3 特征融合与边界修正网络

本文提出的基于特征融合的边界修正显著性目标检测网络的具体结构如图4所示,编码器为ResNet-50的前48层,其通过5个残差块来提取不同层次的显著特征;解码阶段采用由6个多尺度特征解码模块组成的解码器进行特征融合,通过依次融合相邻层特征来感知显著目标的不同尺度。在此

基础上,边界修正模块学习显著目标的边界信息,提高了网络

对目标轮廓特征的关注度。

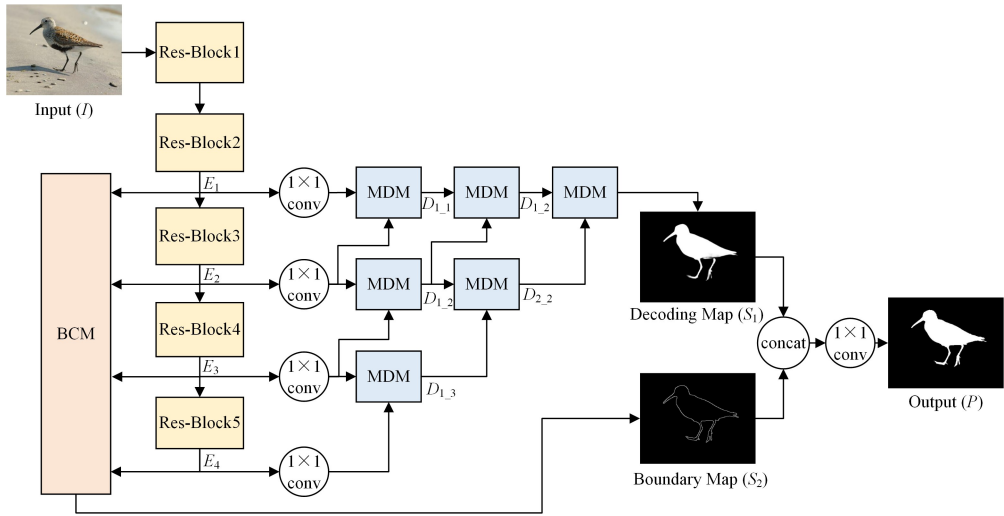


图4 特征融合和边界修正网络

Fig. 4 Feature fusion and boundary correction network

在边界修正网络中,给定的输入图 I 将首先经过编码器进行基础特征提取,该过程将生成不同层次的特征图 E_i ($i = 1, 2, 3, 4$),特征图 E_i 将被送入多尺度特征解码器中,经过多次融合生成解码图 S_1 。此外,边界修正模块也将学习 E_4 以生成边界图 S_2 。最后将解码图 S_1 与边界图 S_2 级联结合,由 1×1 卷

积降维生成最终显著图 P 。其中, S_1 和 S_2 作为深层次监督用于优化训练过程。所提网络的具体参数如表 1 所列,其中 $C(x * x, c)$ 代表尺寸为 $x * x$ 、个数为 c 的卷积核,编码器中下采样操作由尺寸为 $1 * 1$ 、步长为 2 的卷积实现,Concat 为级联操作,边界修正模块中的上采样采用双线性插值法。

表 1 网络参数

Table 1 Parameters of network

主干网络	模块	网络层	中间输出	
编码器 ResNet-50	C1	$C(7 * 7, 64)$ 下采样		
	C2	$[C(1 * 1, 64), C(3 * 3, 64), C(1 * 1, 64)] \times 3$ 下采样	E_1	
	C3	$[C(1 * 1, 64), C(3 * 3, 64), C(1 * 1, 64)] \times 4$ 下采样	E_2	
	C4	$[C(1 * 1, 64), C(3 * 3, 64), C(1 * 1, 64)] \times 6$ 下采样	E_3	
	C5	$[C(1 * 1, 512), C(3 * 3, 512), C(1 * 1, 2048)] \times 3$ 下采样	E_4	
特征融合 解码器	MDM1_1	$C(3 * 3, 64), C(3 * 3, 64)$ 上采样, Concat($H_{1,1}, L_{1,1}$)	$H_{1,1}, L_{1,1}$ $D_{1,1}$	
	MDM1_2	$C(3 * 3, 64), C(3 * 3, 64)$ 上采样, Concat($H_{1,2}, L_{1,2}$)	$H_{1,2}, L_{1,2}$ $D_{1,2}$	
	MDM1_3	$C(3 * 3, 64), C(3 * 3, 64)$ 上采样, Concat($H_{1,3}, L_{1,3}$)	$H_{1,3}, L_{1,3}$ $D_{1,3}$	
	MDM2_1	$C(3 * 3, 64), C(3 * 3, 64)$ 上采样, Concat($H_{2,1}, L_{2,1}$)	$H_{2,1}, L_{2,1}$ $D_{2,1}$	
	MDM2_2	$C(3 * 3, 64), C(3 * 3, 64)$ 上采样, Concat($H_{2,2}, L_{2,2}$)	$H_{2,2}, L_{2,2}$ $D_{2,2}$	
	MDM3_1	$C(3 * 3, 64), C(3 * 3, 64)$ 上采样, Concat($H_{3,1}, L_{3,1}$)	$H_{3,1}, L_{3,1}$ S_1	
	边界修正 模块 BCM		上采样 $E_3 + B_4$	B_4
			$C(3 * 3, 64)$ 上采样	B_3
			$E_2 + B_3$ $C(3 * 3, 64)$	B_3
		上采样 $E_1 + B_2$	S_2	
		$C(3 * 3, 64)$		
输出		Concat(S_1, S_2) $C(1 * 1, 1)$	P	

3.1 特征融合解码器

由多尺度感知模块组成的解码器旨在深度融合不同层次的显著特征,使网络能够有效地感知不同尺度下的显著目标。传统的基于 FCN 的方法通常采用多级跳跃连接,然而直接跨级结合可能会使最终的显著图中出现许多噪点,且容易丢失细节信息,影响网络的检测性能。所提的基于多尺度感知模块的解码器采用多次融合邻接层特征的方法来避免跨层融合造成的特征污染,同时减少细节信息的丢失。

3.2 特征融合

如图 4 所示,特征融合解码器由 6 个多尺度解码模块组成,为三层解码结构。编码阶段生成的特征图 $E_i (i=1,2,3,4)$ 将通过 3 个多尺度解码模块对相邻层特征进行融合,来生成解码图 $D_{1,i} (i=1,2,3)$,之后经过 2 个多尺度解码模块生成解码图 $D_{2,i} (i=1,2)$,最后由 1 个多尺度解码模块生成最终的解码图 S_1 。

多尺度解码模块旨在融合相邻层次的显著特征,其具体结构如图 5 所示。首先采用 3×3 卷积、归一化和 ReLU 操作对输入的低层次特征与高层次特征进行学习,之后通过上采样操作使两特征图分辨率保持一致,随后使用级联操作进行特征融合,最后通过 3×3 卷积、归一化和 ReLU 操作学习显著目标的多尺度形态并输出多尺度特征图。

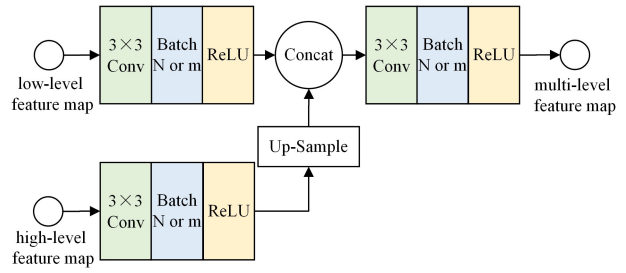


图 5 多尺度解码模块

Fig. 5 Multi-scale decoding module

3.3 边界修正模块

边界修正模块旨在通过监督网络对显著目标边界的学习情况来强调其轮廓特征。其具体结构如图 6 所示。对于不同层次的特征图 $E_i (i=1,2,3,4)$,边界修正模块首先使用 1×1 卷积进行降维,随后使用上采样操作统一图像分辨率,之后通过像素点相加的方式以自底向上的顺序对相邻层特征进行融合。为了提高网络鲁棒性,加入 3×3 卷积、归一化和 ReLU 操作来学习融合后的显著特征,最终输出涵盖显著目标轮廓信息的边界图 S_2 。边界修正模块通过 Canny 算法^[19]处理真值图,生成目标边界标签来监督训练,如图 7 所示。

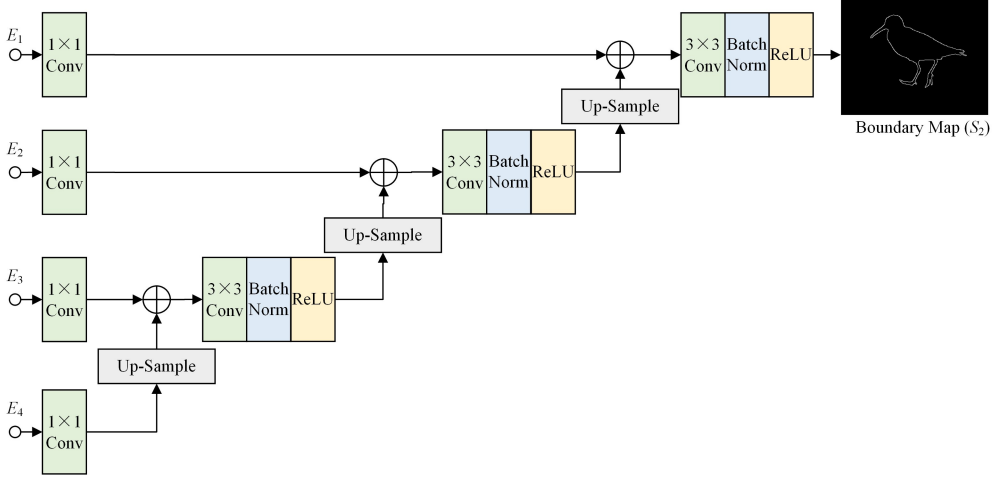


图 6 边界修正模块

Fig. 6 Boundary correction module

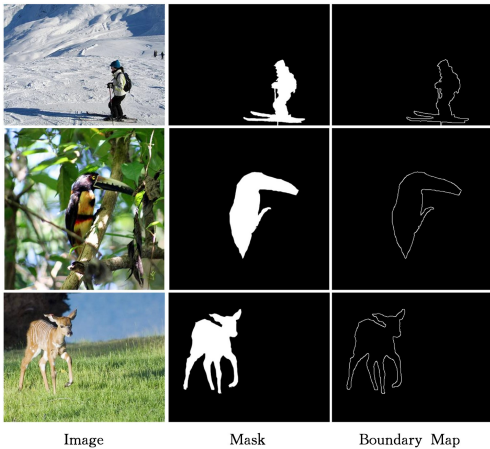


图 7 边界标签示例

Fig. 7 Examples of boundary labels

3.4 损失函数

在显著性目标检测中,由式(1)定义的交叉熵损失 L_{BCE} 常被用于监督网络检测性能,其通过计算单像素点的损失来衡量预测准确度。

$$L_{BCE}(P, G) = - \sum_{i=1}^H \sum_{j=1}^W (P_{ij} \log_2(G_{ij}) + (1 - P_{ij}) \log_2(1 - G_{ij})) \quad (1)$$

其中, G_{ij} 和 P_{ij} 分别为真值图 G 和预测图 P 在位置 (i, j) 的值。

由式(2)定义的交并比损失 L_{IoU} 衡量了真值图与预测图之间的整体差异,可以帮助网络更好地学习显著目标的整体信息。

$$L_{IoU}(P, G) = 1 - \frac{\sum_{i=1}^H \sum_{j=1}^W (G_{ij} * P_{ij})}{\sum_{i=1}^H \sum_{j=1}^W (G_{ij} + P_{ij} - G_{ij} * P_{ij})} \quad (2)$$

交叉熵损失与交并比损失往往协同工作,用于监督显著

目标的学习情况,如式(3)所示。

$$L(P, G) = \omega \times L_{\text{BCE}}(P, G) + (1 - \omega) \times L_{\text{IoU}}(P, G) \quad (3)$$

其中, ω 为超参数,用于协调 $L_{\text{BCE}}(P, G)$ 与 $L_{\text{IoU}}(P, G)$ 的权重。

此外,结合深度学习方法,本文同样分别监督了解码图 S_1 和边界图 S_2 的学习情况,如式(4)和式(5)所示。

$$L(S_1, G) = L_{\text{BCE}}(S_1, G) + L_{\text{IoU}}(S_1, G) \quad (4)$$

$$L(S_2, E) = L_{\text{BCE}}(S_2, E) + L_{\text{IoU}}(S_2, E) \quad (5)$$

其中, E 为目标边界标签。

综上,定义网络最终的损失计算为式(6),在监督最终输出显著图的同时还关注了目标边界的学习情况。

$$L = L(P, G) + \frac{1}{2} (L(S_1, G) + L(S_2, E)) \quad (6)$$

其中, i 为超参数,用于调整多层次监督的权重。

4 实验和分析

本章首先介绍实验采用的 5 个数据集与实验细节,介绍常用在显著性目标检测中的评价指标。然后,通过与现阶段先进的显著性目标检测算法的检测精度、显著图可视化结果以及 P-R 曲线进行对比,来对实验结果进行定量与定性的分析。最后,进行消融实验来验证所提方法的有效性。

4.1 数据集和实验细节

所提算法在 ECSSD^[20], PASCAL-S^[21], HKU-IS^[22], DUTS^[23] 和 DUT-OMRON^[24] 这 5 个数据集上进行了实验。ECSSD 包含 1 000 张图像, PASCAL-S 包含来自 VOC 2010^[25] 分割比赛的 850 张图像, HKU-IS 包含 4 447 张高质量且具有多显著区域的图像。DUTS 来自于 SUN^[26] 和 ImageNet DET^[27], 其中包括具有 10 553 张图像的 DUTS-TR 和 5 019 张图像的 DUTS-TE。DUT-OMRON 包含 5 168 张复杂场景下的图像,具有更强的挑战性。

实验采用 DUTS-TR 作为训练集, DUTS-TE 和其他 4 个数据集作为测试集。训练前使用水平翻转与随机裁剪进行数据增强, ResNet 加载 ImageNet 预训练模型。初始学习率设为 0.05, 最小学习率为 0.0005。采用 SGD 优化器, Batchsize 设为 32, 共训练 64 轮。

4.2 评估指标

实验采用 P-R 曲线、平均绝对误差 (MAE)、 F_β 值、S 值 4 种评价指标来检验模型性能。

1) P-R 曲线即准确率-召回率曲线,由不同阈值下的准确

率和召回率画出,用于刻画准确率 (Precision) 与召回率 (Recall) 之间的关系。其中,准确率和召回率的计算方式如式(7)所示。

$$\begin{cases} Precision = \frac{TP}{TP + FP} \\ Recall = \frac{TP}{TP + FN} \end{cases} \quad (7)$$

其中, TP, FP, FN 分别代表显著区域被预测为显著区域、非显著区域被预测为显著区域、显著区域被预测为非显著区域的像素点数。

2) 平均绝对误差 MAE 用于计算预测图和真值图之间的平均像素误差,如式(8)所示。其越小说明预测效果越好。

$$MAE = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W |P(i, j) - G(i, j)| \quad (8)$$

其中, $P(i, j)$ 和 $G(i, j)$ 分别代表预测图 P 和真值图 G 在位置 (i, j) 上的像素值。

3) F_β 值为准确率和召回率的加权调和平均,计算式如式(9)所示。

$$F_\beta = \frac{(1 + \beta^2) \times Precision \times Recall}{\beta^2 \times (Precision + Recall)} \quad (9)$$

其中, β 根据文献[28]的研究设置为 0.3。

4) S 指标用于评估预测图与真值图间的相似性,包含了基于区域的结构相似性 S_r 与基于目标的结构相似性 S_o 。计算方法如式(10)所示。依照文献[29],设置 α 为 0.5。

$$S = (1 - \alpha) \times S_r + \alpha \times S_o \quad (10)$$

4.3 精度对比

表 2 列出了在 5 个数据集上所提算法与当前主流的基于深度学习的显著性目标检测算法的 MAE、 F_β 值、S 值的对比结果。对比的 13 种方法包括 TDUB^[30], AFNet^[31], PoolNet^[32], CPR-R^[33], GCPA^[34], GateNet^[35], ITSD^[13], MINet^[28], MFNet^[36], DNA^[37], EANet^[38], ICON-R^[39] 和 DCENet^[40]。

得益于网络对多尺度特征的充分学习与对边界信息的关注,所提算法在整体上取得了较好的结果。在数据集 ECSSD 上, MAE 只比 ICON-R 高 0.001, 同时 F_β 和 S 指标取得了更好的结果。在数据集 PASCAL-S 上,所提算法在 MAE 和 F_β 上取得了最好的结果,仅在 S 指标上比 DCENet 低 0.002。在数据集 DUTS-TE 和 HKU-IS 上,所提算法在各指标上均取得了最好的结果。在最具挑战性的数据集 DUT-OMRON 上,所提算法的 MAE 最低,但相比 ITSD 在 F_β 上的 0.793 和 MFNet 在 S 指标上的 0.742, 还存在一定的进步空间。

表 2 5 个数据集上的定量对比

Table 2 Quantitative comparison on five datasets

Methods	ECSSD			PASCAL-S			DUTS-TE			HKU-IS			DUT-OMRON		
	MAE ↓	F_β ↑	S ↑	MAE ↓	F_β ↑	S ↑	MAE ↓	F_β ↑	S ↑	MAE ↓	F_β ↑	S ↑	MAE ↓	F_β ↑	S ↑
TDUB(CVPR 19)	0.042	0.908	0.913	0.071	0.779	0.844	0.048	0.767	0.865	0.038	0.878	0.907	0.061	0.739	0.837
AFNet(CVPR 19)	0.042	0.908	0.913	0.070	0.821	0.844	0.046	0.792	0.867	0.036	0.888	0.905	0.057	0.738	0.826
PoolNet(CVPR 19)	0.039	0.915	0.921	0.074	0.822	0.845	0.040	0.809	0.883	0.032	0.899	0.916	0.055	0.747	0.835
CPR-R(CVPR 19)	0.037	0.917	0.918	0.072	0.824	0.842	0.043	0.805	0.869	0.034	0.891	0.905	0.056	0.747	0.825
GCPA(AAAI 20)	0.035	0.926	0.926	0.062	0.850	0.863	0.037	0.860	0.883	0.029	0.916	0.918	0.055	0.789	0.831
GateNet(ECCV 20)	0.040	0.925	0.919	0.068	0.851	0.856	0.039	0.855	0.883	0.033	0.909	0.914	0.055	0.791	0.836
ITSD(CVPR 20)	0.035	0.928	0.924	0.066	0.851	0.858	0.040	0.855	0.884	0.031	0.911	0.916	0.060	0.793	0.839
MINet(CVPR 20)	0.034	0.930	0.925	0.064	0.852	0.855	0.037	0.860	0.883	0.029	0.916	0.918	0.055	0.789	0.831
MFNet(ICCV 21)	0.084	0.854	0.834	0.115	0.751	0.770	0.076	0.710	0.775	0.059	0.851	0.846	0.087	0.646	0.742
DNA(TIP 21)	0.042	0.897	0.915	0.079	0.852	0.837	0.046	0.794	0.856	0.035	0.876	0.905	0.063	0.722	0.818
EANet(IVC 21)	0.037	0.900	0.916	0.068	0.795	0.849	0.041	0.807	0.863	0.031	0.890	0.909	0.055	0.747	0.836
ICON-R(TPAMI 22)	0.032	0.918	0.929	0.064	0.818	0.861	0.037	0.836	0.888	0.029	0.902	0.920	0.057	0.761	0.844
DCENet(TCSVT 22)	0.035	0.924	0.921	0.061	0.845	0.862	0.038	0.852	0.877	0.029	0.910	0.924	0.055	0.770	0.839
Ours(FFBCN)	0.033	0.934	0.925	0.059	0.860	0.860	0.035	0.869	0.888	0.027	0.923	0.920	0.055	0.788	0.830

4.4 PR 曲线

图 8 给出了所提算法和其他 6 种算法的 P-R 曲线的

对比。可以看出,所提算法表现出了更好的性能,这也证明其具有更好的鲁棒性。

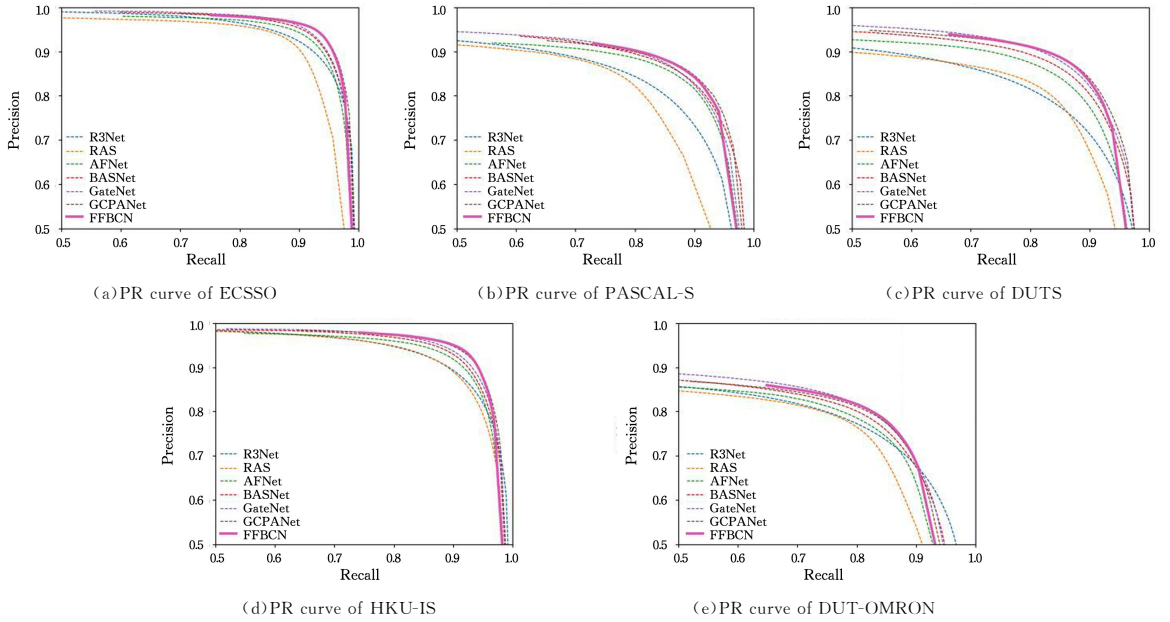


图 8 P-R 曲线对比

Fig. 8 Comparison of P-R curve

4.5 可视化对比

图 9 给出了所提算法与一些优秀显著性目标检测算法的可视化对比结果。

图像中不同尺度的显著目标,这证明了在多尺度解码模块的帮助下,所提算法能够更好的适应不同目标尺度的变化。而由图 9(e)一图 9(h)可以看出,得益于边界修正模块,相比其他的显著性目标检测算法,即使显著目标具有复杂边界,所提算法也能生成更清晰的轮廓。

4.6 调参实验

为了提高检测性能,在数据集 DUTS 和 DUT-OMRON 上对损失函数中的超参数 w 和 i 进行调参实验,其结果如表 3 和表 4 所列。可以看出,当 $w=0.5$ 和 $i=1$ 时,模型的性能最好。

表 3 超参数 w 调参实验

w	DUTS			DUT-OMRON		
	MAE	F_β	S	MAE	F_β	S
0.1	0.041	0.848	0.872	0.064	0.770	0.821
0.2	0.037	0.856	0.877	0.059	0.778	0.827
0.3	0.036	0.860	0.881	0.056	0.785	0.828
0.4	0.037	0.867	0.887	0.056	0.788	0.830
0.5	0.035	0.869	0.888	0.055	0.788	0.830
0.6	0.036	0.862	0.885	0.057	0.782	0.827
0.7	0.038	0.859	0.884	0.057	0.780	0.826
0.8	0.040	0.855	0.876	0.059	0.777	0.824
0.9	0.042	0.871	0.873	0.060	0.770	0.819
1.0	0.042	0.866	0.874	0.063	0.766	0.815

表 4 超参数 i 调参实验

i	DUTS			DUT-OMRON		
	MAE	F_β	S	MAE	F_β	S
1	0.035	0.869	0.888	0.055	0.788	0.830
2	0.036	0.865	0.883	0.058	0.783	0.824
3	0.038	0.859	0.875	0.059	0.779	0.820
4	0.040	0.860	0.875	0.062	0.772	0.820
5	0.041	0.856	0.870	0.062	0.769	0.819



图 9 所提算法同其他先进算法的可视化对比

Fig. 9 Visual comparison between the proposed algorithm and other advanced algorithms

由图 9(a)一图 9(d)可以看出,所提算法能够有效检测出

4.7 消融实验

为了验证所提方法的有效性,在数据集 DUTS 和 DUT-OMRON 上设计了一系列消融实验。实验采用 ResNet-50 编码器和 3 个 VGG 块组成解码器,并将其作为基础网络,之后将 VGG 块替换为多尺度解码模块 MDM,再将解码器替换为特征融合解码器 FFD,最后加入边界修正模块 BBM 来进行比较。实验结果如表 5 所列,相比基础网络,在加入多尺度解码模块 MDM 后,得益于对邻层特征的高效融合,模型在各项指标上都有一定的提升。其中,在数据集 DUTS 上,MAE 由 0.040 降为 0.039, F_{β} 值提升了 0.002, S 值提升了 0.002。而在数据集 DUT-OMRON 上,MAE 由 0.062 降低为 0.060, F_{β} 值提升了 0.003, S 值提升了 0.002。在使用特征融合解码器 FFD 后,模型对多尺度目标具有更强的感知能力,在各项指标上均有显著的提升。在数据集 DUTS 上,MAE 由 0.039 降为 0.037, F_{β} 值提升了 0.010。在数据集 DUT-OMRON 上,MAE 由 0.060 降低为 0.058, F_{β} 值提升了 0.008, S 值提升了 0.007。最后加入边界修正模块 BCM,模型得以有效学习目标的边界信息,性能得到了进一步的提升。其具体体现在 DUTS 数据集上,MAE 降低了 0.002, F_{β} 提升了 0.007, S 值提升了 0.005。而在数据集 DUT-OMRON 上,MAE 降低了 0.003, F_{β} 提高了 0.005,这充分证明了所提方法对准确检测的作用。

表 5 消融实验

Table 5 Ablation experiment

MDM	FFD	BCM	DUTS			DUT-OMRON		
			MAE	F_{β}	S	MAE	F_{β}	S
			0.040	0.850	0.878	0.062	0.772	0.821
✓			0.039	0.852	0.880	0.060	0.775	0.823
✓	✓		0.037	0.862	0.883	0.058	0.783	0.830
✓	✓	✓	0.035	0.869	0.888	0.055	0.788	0.830

为了进一步验证各所提模块的有效性,本节对消融实验生成的显著图进行了可视化。图 10 给出了多尺度解码模块的可视化对比结果,可以看出,得益于多尺度解码模块对邻层特征的有效融合,模型生成的预测图更加清晰。

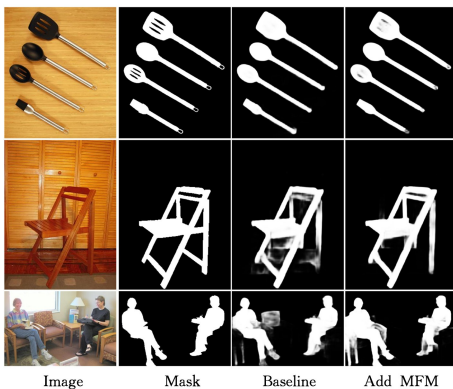


图 10 多尺度解码模块的可视化消融实验

Fig. 10 Visual ablation experiment of multi-scale decoding module

图 11 给出了特征融合解码器的可视化对比结果,可以看出,在使用特征融合解码器融合不同层次特征后,模型能够

有效地检测出不同尺度的显著目标。

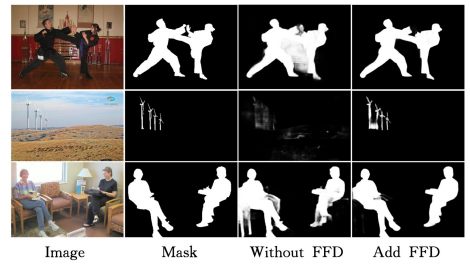


图 11 特征融合解码器的可视化消融实验

Fig. 11 Visual ablation experiment of feature fusion decoder

图 12 给出了加入边界修正模块前后的显著图可视化对比结果。可以看出,在加入边界修正模块后,模型能够更充分地学习到显著目标的边界特征,生成的预测图也更清晰。

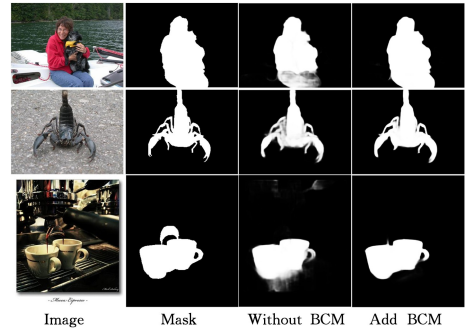


图 12 边界修正模块的可视化消融实验

Fig. 12 Visual ablation experiment of boundary correction module

4.8 时间复杂度分析

表 6 列出了本文方法与其他先进方法在 PASCAL-S 和 DUTS 两个数据集上关于时间复杂度的对比结果。结果显示,在相同的设备上,本文方法每秒传输的帧数仅次于 GC-PA,同时,在定量指标 MAE, F_{β} , S 上优于 PoolNet, ITSD 与 MINet 等方法,这展示了本文方法在速度和精度上的优势。在兼顾效率的同时,本文算法更加注重检测的准确性。

表 6 时间复杂度对比

Table 6 Comparison of time complexity

FPS	PASCAL-S			DUTS			
	MAE	F_{β}	S	MAE	F_{β}	S	
PoolNet	40.5	0.074	0.822	0.845	0.040	0.809	0.883
MINet	31.1	0.064	0.852	0.855	0.037	0.860	0.883
ITSD	47.3	0.066	0.851	0.858	0.040	0.855	0.884
GCPA	51.7	0.062	0.850	0.863	0.037	0.860	0.883
Ours(FFBCN)	48.7	0.059	0.860	0.860	0.035	0.869	0.888

5 缺陷分析

图 13 给出了本文方法的缺陷示例图。对于图像中存在的影子以及相邻的植物等干扰因素较强的背景噪声,本文方法难以有效甄别前背景,容易发生误判。在今后的研究中,将考虑进一步挖掘更深层次的显著特征,提高网络对目标与背景噪声的区别能力。

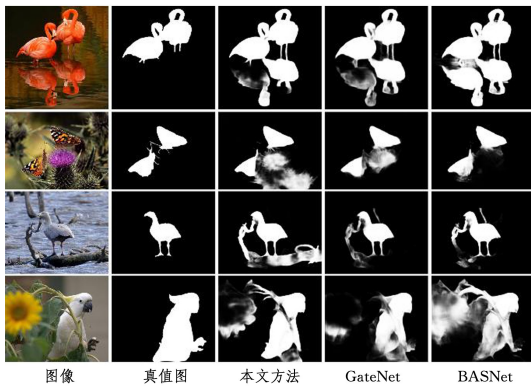


图 13 缺陷示例

Fig. 13 Defect samples

结束语 本文提出了基于特征融合与边界修正显著性目标检测算法,主要提高了网络对目标尺度变化的感知能力以及对复杂边界的关注度,显著提升了检测性能。其中,提出了基于多尺度感知模块的解码器来有效融合不同语义下的特征,并保留了各个阶段的重要信息。同时,提出了边界修正模块及对应的损失函数来提高网络对边界特征的学习效率,通过提高对目标边界的感知能力来提高最终检测精度。通过大量实验证明了所提方法的有效性,未来将考虑进一步监督前背景预测,以减少假阳性。

参考文献

- [1] REN Z,GAO S,CHIA LT,et al. Region-based saliency detection and its application in object recognition [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2013, 24(5):769-779.
- [2] WEI Y,FENG J,LIANG X,et al. Object region mining with adversarial erasing: A simple classification to semantic segmentation approach [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017:1568-1576.
- [3] LIANG P,PANG Y,LIAO C,et al. Adaptive objectness for object tracking [J]. IEEE Signal Processing Letters, 2016, 23(7): 949-53.
- [4] MA Y F,ZHANG H J. Contrast-based image attention analysis by using fuzzy growing [C]//Proceedings of the Eleventh ACM International Conference on Multimedia. 2003:374-381.
- [5] JIANG H,WANG J,YUAN Z,et al. Automatic salient object segmentation based on context and shape prior [C]// Proceedings of British Machine Vision Conference. 2011.
- [6] ZHU W,LIANG S,WEI Y,et al. Saliency optimization from robust background detection [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014: 2814-2821.
- [7] LONG J,SHELHAMER E,DARRELL T. Fully convolutional networks for semantic segmentation [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015:3431-3440.
- [8] LIN T Y,DOLLÁR P,GIRSHICK R,et al. Feature pyramid networks for object detection [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 2117-2125.
- [9] PAN M Y,SONG H H,ZHANG K H,et al. Learning Global Guided Progressive Feature Aggregation Lightweight Network for Salient Object Detection [J]. Computer Science, 2021, 48(6):103-109.
- [10] WU J Y,YANG S,DU J,et al. Review of Bottom-up Salient Object Detection [J]. Computer Science, 2019, 46(3):48-52.
- [11] BADRINARAYANAN V,KENDALL A,CIPOLLA R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(12):2481-2495.
- [12] LIU J J,HOU Q,LIU Z A,et al. Poolnet+: Exploring the potential of pooling for salient object detection [C]// IEEE Transactions on Pattern Analysis and Machine Intelligence. 2022.
- [13] ZHOU H,XIE X,LAI J H,et al. Interactive two-stream decoder for accurate and fast saliency detection [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020:9141-9150.
- [14] WU Y H,LIU Y,ZHANG L,et al. EDN: Salient object detection via extremely-downsampled network [J]. IEEE Transactions on Image Processing, 2022, 31:3125-3136.
- [15] WU Z,LI S,CHEN C,et al. Salient Object Detection via Dynamic Scale Routing [J]. IEEE Transactions on Image Processing, 2022, 31:6649-6663.
- [16] QIN X,ZHANG Z,HUANG C,et al. Basnet: Boundary-aware salient object detection [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 7479-7489.
- [17] WEI J,WANG S,HUANG Q. F³Net: Fusion, Feedback and Focus for Salient Object Detection [C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2020:12321-12328.
- [18] CHEN Z,ZHOU H,LAI J,et al. Contour-aware loss: Boundary-aware learning for salient object segmentation [J]. IEEE Transactions on Image Processing, 2020, 30:431-443.
- [19] CANNY J. A computational approach to edge detection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1986(6):679-698.
- [20] SHI J,YAN Q,XU L,et al. Hierarchical image saliency detection on extended CSSD [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2015, 38(4):717-729.
- [21] LI Y,HOU X,KOCH C,et al. The secrets of salient object segmentation [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014:280-287.
- [22] LI G,YU Y. Deep contrast learning for salient object detection [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016:478-487.
- [23] WANG L,LU H,WANG Y,et al. Learning to detect salient objects with image-level supervision [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017:136-145.
- [24] YANG C,ZHANG L,LU H,et al. Saliency detection via graph

- based manifold ranking [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013;3166-3173.
- [25] EVERINGHAM M, ESLAMI S, VAN GOOL L, et al. The pascal visual object classes challenge; A retrospective [J]. International Journal of Computer Vision, 2015, 111(1):98-136.
- [26] XIAO J, HAYS J, EHINGER K A, et al. Sun database: Large-scale scene recognition from abbey to zoo [C]// 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE, 2010;3485-3492.
- [27] DENG J, DONG W, SOCHER R, et al. Imagenet; A large-scale hierarchical image database [C]// Proceedings of 2009 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2009;248-255.
- [28] PANG Y, ZHAO X, ZHANG L, et al. Multi-scale interactive network for salient object detection [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020;9413-9422.
- [29] FAN D P, CHENG M M, LIU Y, et al. Structure-measure: A new way to evaluate foreground maps [C]// Proceedings of the IEEE International Conference on Computer Vision, 2017;4548-4557.
- [30] WANG W, SHEN J, CHENG M M, et al. An iterative and cooperative top-down and bottom-up inference network for salient object detection [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019;5968-5977.
- [31] FENG M, LU H, DING E. Attentive feedback network for boundary-aware salient object detection [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019;1623-1632.
- [32] LIU J J, HOU Q, CHENG M M, et al. A simple pooling-based design for real-time salient object detection [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019;3917-3926.
- [33] WU Z, SU L, HUANG Q. Cascaded partial decoder for fast and accurate salient object detection [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019;3907-3916.
- [34] CHEN Z, XU Q, CONG R, et al. Global context-aware progressive aggregation network for salient object detection [C]// Proceedings of the AAAI Conference on Artificial Intelligence, 2020;10599-10606.
- [35] ZHAO X, PANG Y, ZHANG L, et al. Suppress and balance: A simple gated network for salient object detection [C]// Proceedings of European Conference on Computer Vision. Springer, 2020;35-51.
- [36] PIAO Y, WANG J, ZHANG M, et al. MFNet; Multi-filter Directional Network for Weakly Supervised Salient Object Detection [C]// Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021;4136-4145.
- [37] FENG M, LU H, YU Y. Residual learning for salient object detection [J]. IEEE Transactions on Image Processing, 2020, 29: 4696-4708.
- [38] CHEN X, ZHANG Q, ZHANG L. Edge-aware salient object detection network via context guidance [J]. Image and Vision Computing, 2021, 110:104166.
- [39] TONG N, LU H, ZHANG Y, et al. Salient object detection via global and local cues [J]. Pattern Recognition, 2015, 48(10): 3258-3267.
- [40] MEI H, LIU Y, WEI Z, et al. Exploring dense context for salient object detection [C]// IEEE Transactions on Circuits and Systems for Video Technology, 2021.



CHEN Hui, born in 1980, postgraduate, associate professor. Her main research interests include electronic information and visual Internet of Things.



PENG Li, born in 1967, Ph.D, professor. His main research interests include networked collaborative control and visual Internet of Things.

(责任编辑: 喻黎)