文章编号：0253-374X(2024)02-0293-10

DOI：10.11908/j.issn.0253-374x.22220

边缘信息增强的显著性目标检测网络

赵卫东，王　辉，柳先辉

（同济大学 电子与信息工程学院，上海 201804）

摘要：针对显著性目标检测任务中识别结果边缘模糊的问题，提出了一种能够充分利用边缘信息增强边缘像素置信度的新模型。该网络主要有两个创新点：设计三重注意力模块，利用预测图的特点直接生成前景、背景和边缘注意力，并且生成注意力权重的过程不增加任何参数；设计边缘预测模块，在分辨率较高的网络浅层进行有监督的边缘预测，并与网络深层的显著图预测融合，细化了边缘。在6种常用公开数据集上用定性和定量的方法评估了该模型，并且与其他模型进行充分对比，证明设计的新模型能够取得最优的效果。此外，该模型参数量为30.28 M，可以在GTX 1080 Ti显卡上达到31 帧·s^{-1}的预测速度。

关键词：显著性目标检测；注意力机制；边缘检测；深度卷积神经网络

中图分类号：TP391.4　　**文献标志码**：A

Edge Enhancing Network for Salient Object Detection

ZHAO Weidong, *WANG Hui*, *LIU Xianhui*

(College of Electronic and Information Engineering, Tongji University, Shanghai 201804, China)

Abstract: Aiming at the problem of blurred edges in salient object detection, this paper proposes a new method that can fully utilize edge information to enhance the confidence of edge pixels. First, the triple attention module is introduced, which uses the characteristics of the predicted saliency map to directly generate foreground, background and edge attention, and the process of generating attention weights does not add any parameters. Next, the edge prediction module is introduced, which performs supervised edge prediction in the shallowest layer of the network with the biggest feature map, and fuses the predicted edge with the saliency map to refine the edges. Finally, the model is qualitatively and is quantitatively evaluated on six commonly used public datasets, and fully compared with other models, which proves that the proposed model can achieve the best results. The method proposed in this paper has 30.28 M parameters, and can predict saliency maps at 31 frames per second on GTX 1080 Ti graphics card.

Keywords: salient object detection; attention mechanism; boundary detection; deep convolutional neural network

人的视觉系统能选择性地注视不同场景中富含丰富信息的区域[1]，在机器视觉领域中，利用这种视觉选择性注意力机制进行像素级物体检测的方法被称为显著性目标检测(salient object detection，SOD)。由于SOD能够在检测出显著对象的同时保留物体边缘细节，在应用中主要作为一种图像预处理方法。

在SOD发展的早期，大多数模型依赖于图像低层特征和启发式算法[2]，自从深度学习和卷积神经网络的兴起以来，因其强大的特征发现与表达能力，目前几乎所有的典型模型都基于深度卷积神经网络[3]。即使这些模型已经能取得非常优异的成绩，但在网络处理图像的过程中，经过层层下采样，图片的细节信息被大量丢失，使预测图的边缘无法很好地贴合复杂的物体边缘。

1　相关工作

(1)特征融合

为了充分利用来自不同卷积层的信息从而检测不同尺度的物体，一些研究聚焦于如何有效地整合多尺度特征。文献[4]提出了一种具有深监督结构的整体嵌套边缘检测网络来学习多层次的特征。受文献[4]的启发，很多SOD模型都采用了特征融合和深监督的方式。文献[5]设计了一个多尺度融合网络，将高层语

收稿日期：2022-05-13
基金项目：上海市科技计划项目(20DZ2281000)
第一作者：赵卫东，研究员，工学博士，主要研究方向为网络协同制造、机器视觉。E-mail：wd@tongji.edu.cn
通信作者：王　辉，工学硕士，主要研究方向为机器视觉、深度学习。E-mail：2033111@tongji.edu.cn

论文拓展介绍

义信息和低层空间信息结合起来,但使用了传统的超像素预处理或者条件随机场后处理来提高算法效果。文献[6]通过直接连接特征图来聚合高层和低层特征,但递归预测显著图的方法降低了算法时间效率。文献[7]使用金字塔池化模块和多阶段细化机制来整合全局和局部上下文信息。文献[8]设计了一种双向消息结构,可以在多级特征之间传递信息,并使用一个门函数控制消息传输率。文献[9]引入了注意力引导网络以选择性地融合多尺度上下文信息,并用多路径循环反馈模型将全局语义信息从深层传递到浅层。文献[6-9]提出的都是近几年对显著性目标检测效果有较大提升的模型,但都主要关注网络不同层特征的融合,而没有关注检测到的物体边缘模糊的问题。

(2) 注意力机制

注意力机制是近些年的深度神经网络中一个频繁被使用的方法,通过给不同区域的特征赋予不同的权值,达到强调特定信息的目的,在SOD领域,注意力机制也被广泛地应用。文献[10]采用反向注意力来引导残差学习。反向注意力把当前预测的显著区域擦除,从而引导网络从未擦除的区域中有效地学习丢失的细节,实现更完整的预测。文献[11]在反向注意力残差学习的基础上,提出一种级联式的网络,使高层特征和低层特征的输出循环交替优化彼此,但显著增加了训练与预测时间。文献[12]发现,现有的模型大多只考虑显著性检测的一个方面,即前景信息[9,13]或背景信息[10],导致预测不完整。因此,他们提出了一个融合正注意力和反注意力的模块,正注意力增强了显著区域的预测,而反注意力突出了缺失的细节。文献[14]也提出了双注意力模块来整合前景注意力和背景注意力,但文献[12]采用的是自注意力,文献[14]采用的是外注意力。

(3) 显著图细化

显著图边缘模糊的问题也是很多学者工作的重点。文献[15]把基于超像素的过滤器作为网络的一层进行边缘细化。虽然超像素能够很好地提取图像的低层特征,标记边缘,但传统的超像素算法难以并行运算,影响时间效率,而且不易与网络整合。文献[16]提出了一种多分尺度网格结构的网络来捕捉局部和全局线索,并引入了一种边缘损失函数来减少物体边界上的预测错误,但边缘预测只被简单地融合进最终结果,没有充分得到利用。文献[17]使用标签解耦的方式,将显著性物体的边缘和内部分开,分别监督细节解码器和主体解码器,并用交互解码器获得最终的预测结果,能够得到目前最好的显著性检测结果之一,但模型结构复杂。

2 边缘信息增强的显著性目标检测网络

2.1 总体结构

本文模型的骨干网络为去掉全连接层的ResNet-50[18]。图像的特征经过逐层下采样,得到分辨率小、语义信息丰富的特征图,此特征图虽然丢失了大量的细节信息,但保留了高准确度的物体位置信息。较浅层的特征虽然语义信息不足,但具有更丰富的细节信息,尤其是边缘信息[19]。为了能够充分融合深层和浅层的互补特征,本文受到文献[10]的启发,设计了一种自顶向下逐层优化的残差学习网络。最深层的特征经过多尺度上下文模块(multi-scale context module,MSCM)[10]输出粗略的预测,再逐层地向上传递,浅层特征通过预测残差丰富预测图的细节。每一层预测残差时经过三重注意力模块(triple attention module,TAM),通过前景、背景、边缘三重注意力充分提取信息。最浅层特征用于预测边缘,经过边缘预测模块(edge prediction module,EPM)预测残差,与上一层的结果融合,得到最终预测结果。网络的总体结构如图1所示,为展示方便,其中的显著图、残差图经过缩放处理,使每层的输出图看起来大小相同。

ResNet-50网络各层输出的特征定义为$X_i(i=1,2,3,4,5)$。假设输入的图像I大小为$H\times W\times 3$,则第i层特征的大小为$\frac{H}{2^i}\times\frac{W}{2^i}\times c_i$,其中,$c_i$为特征通道数。在计算过程中,第5层的$X_5$经多尺度上下文模块MSCM输出最小、最粗糙的显著图预测O_5;在第i层($i=4,3,2$),TAM利用X_i和$\mathrm{Up}_{\times 2}(O_{i+1})$($\mathrm{Up}_{\times 2}$表示双倍上采样)输出残差$E_i$,与$\mathrm{Up}_{\times 2}(O_{i+1})$相加,获得比前一层更精细的显著图预测;在最上层,EPM利用最大、细节最丰富的特征X_1预测边缘,并输出残差E_1,与$\mathrm{Up}_{\times 2}(O_2)$相加后得到网络的最终预测结果。显著图的真值为$G_s$,在训练中监督每层输出的显著图预测;显著边缘图的真值为G_e,在训练中监督EPM中的边缘预测。

2.2 三重注意力模块

在自顶向下逐层补充信息、优化显著图的过程中,由于来自深层的显著图中已有一部分语义信息,故如果直接用每层的特征对显著图进行优化,会被大量的冗余信息干扰。如果可以舍弃这些冗余,就能提高信息利用率,进而提高优化效果。为此,本文提出三重注

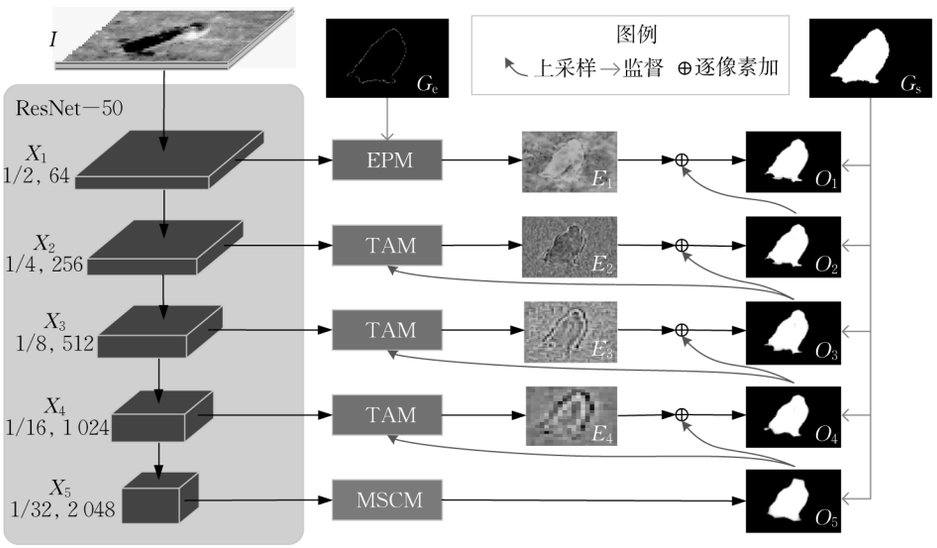

图 1　网络总体结构图

Fig. 1　Overall architecture of network

意力模块即TAM，通过前景、背景、边缘三重注意力引导网络从各层特征中充分提取信息。前景注意力又称正注意力，可以突出并强化显著区域的预测；背景注意力又称负注意力，可以通过突出非显著区域补充丢失的细节信息；边缘注意力突出了物体边缘，补充了复杂的边缘细节信息。TAM的结构如图2所示。

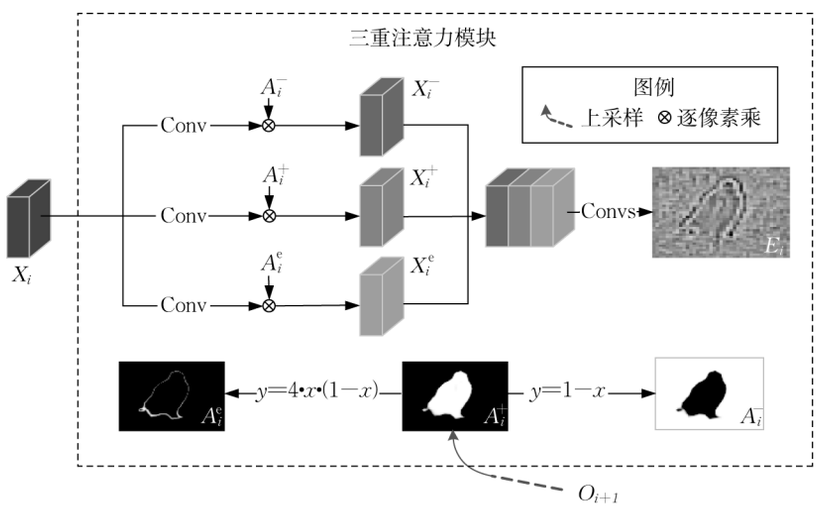

图 2　TAM示意图

Fig. 2　Illustration of TAM

第i层TAM的输入为X_i与O_{i+1}。O_{i+1}经过两倍上采样后为$\mathrm{Up}_{\times 2}(O_{i+1})$，记作$A_i^+$。$A_i^+$即正注意力，负注意力$A_i^-$用公式$y=1-x$得到，边缘注意力$A_i^e$用公式$y=4\cdot x\cdot(1-x)$得到。特征$X_i$经过三个分支分别获得上述三种注意力，生成正特征$X_i^+$、负特征$X_i^-$、边缘特征$X_i^e$，公式表达为

$$X_i^+=\mathrm{Conv}(X_i)\cdot A_i^+ \tag{1}$$

$$X_i^-=\mathrm{Conv}(X_i)\cdot 1-A_i^+ \tag{2}$$

$$X_i^e=\mathrm{Conv}(X_i)\cdot 4\cdot A_i^+\cdot(1-A_i^+) \tag{3}$$

式中："•"表示逐元素乘；Conv表示连续的卷积、批归一化[20]、线性整流[21]操作。X_i在三个分支上分别经过一次Conv，可以起到通道选择的作用，增强注意力的效果。三重特征最终融合并生成残差E_i，如下：

$$E_i=\mathrm{Convs}\left(\mathrm{Concat}\left(X_i^+,X_i^-,X_i^e\right)\right) \tag{4}$$

式中：Concat表示沿着通道维度的连接；Convs表示连续的Conv操作。残差E_i由TAM输出后，与$\mathrm{Up}_{\times 2}(O_{i+1})$相加即可得到本层的显著图预测结果，这体现了自顶向下逐层优化的思想。

三种注意力中，正注意力与负注意力分别强调了前景与背景，而边缘注意力则强化了边缘细节，下

面对边缘注意力的计算进行详细的解释。由于在显著图预测结果中,显著性区域各像素的值是接近1的,只在靠近边缘的地方小于1,而且是渐渐由1平滑地过渡到0,即非显著区域。因此,把值接近0.5的像素点认为是恰好在边缘上,而将值接近0或1的像素点认为是远离边缘的。在TAM中,使用公式$y=4\cdot x\cdot(1-x)$将显著预测图转化为边缘预测图,并保证值域仍为[0,1]。如图3所示,显著预测图中白色的显著区域和黑色的非显著区域经过转换后,都变为了边缘预测图中的黑色区域,而灰色的边缘区域经过转换,则变成了边缘预测图中白色或灰色的边缘区域。

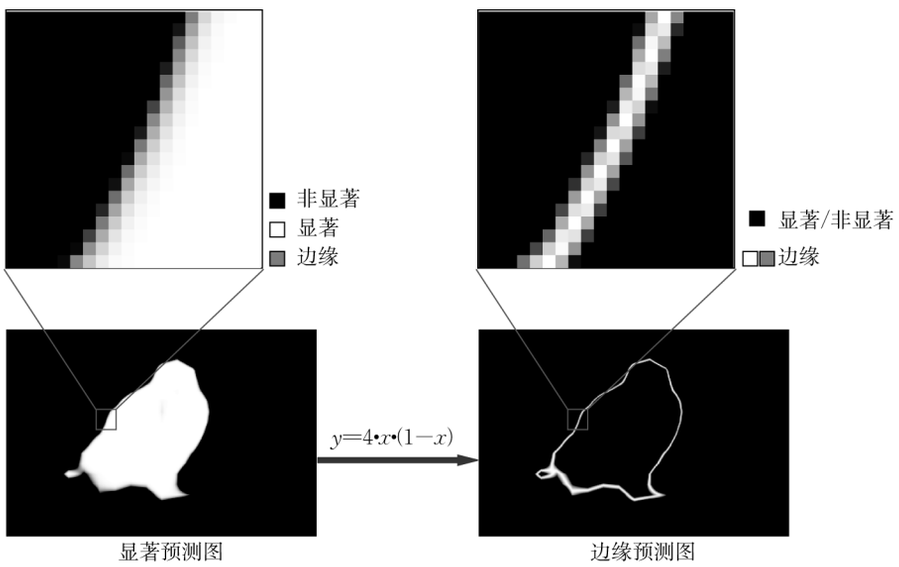

图3 边缘注意力生成示意图

Fig. 3 Illustration of edge attention generation

2.3 边缘预测模块

通过逐层优化的方式可以得到细节越来越丰富的显著图,在此基础上,本文进一步提出边缘预测模块即EPM,在细节信息最丰富的网络第1层,用监督的方式获得显著性物体的边缘,并优化显著图,得到边缘更加清晰的预测结果。TAM中的边缘注意力来自网络内部,而EPM从外部获取边缘信息,两者互为补充,共同增强边缘信息。EPM的结构如图4所示。

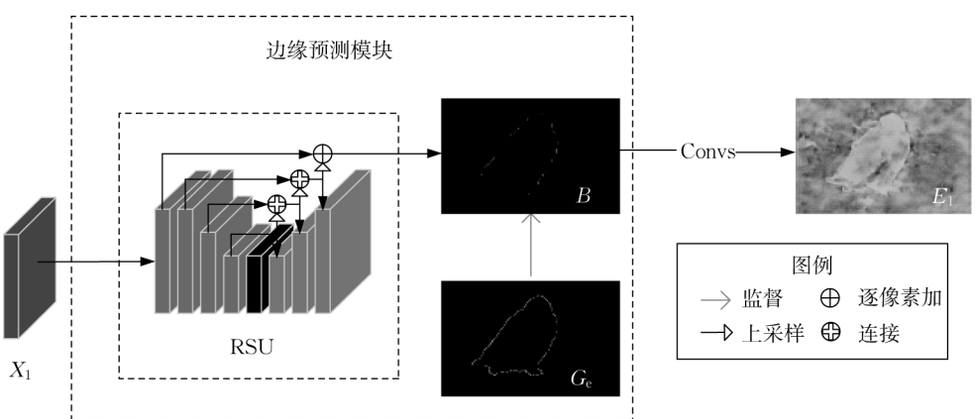

图4 EPM示意图

Fig. 4 Illustration of EPM

EPM的输入为X_1,X_1经过RSU(residual u-blocks)[22]输出边缘预测B,以真值边缘图G_e监督。边缘预测B经过多层卷积生成残差E_1。该步骤用公式表达为

$$E_1=\mathrm{Convs}(\mathrm{EP}(X_1)) \tag{5}$$

式中:EP(edge prediction)表示用于预测边缘的网络,本文采用的是RSU。RSU内部为U型结构,可以在不降低特征图大小的前提下提取出多尺度特

征，保留充足的边缘信息。EPM输出的残差E_1与第2层显著图相加，得到最终的预测结果。

2.4　损失函数

本文使用深监督的方式，对每个尺度的显著图输出进行监督，损失函数定义为

$$L_{\mathrm{s}}=\sum_{i=1}^{5}L_{\mathrm{BCE}}(P_i,G_{\mathrm{s}})+L_{\mathrm{IoU}}(P_i,G_{\mathrm{s}}) \tag{6}$$

式中：P_i表示上采样到输入图像大小的各层显著预测图；G_{s}表示真值预测图；L_{BCE}表示二元交叉熵(binary cross entropy)损失；L_{IoU}表示交并比(intersection over union)损失。

BCE是图像分割领域常用的衡量显著图与真值逐像素误差的方法，计算方法如下：

$$L_{\mathrm{BCE}}(P_i,G_{\mathrm{s}})=-\sum_{(r,c)}G_{\mathrm{s}}(r,c)\log\big(P_i(r,c)\big)+\big(1-G_{\mathrm{s}}(r,c)\big)\log\big(1-P_i(r,c)\big) \tag{7}$$

式中：(r,c)表示像素坐标。

IoU的含义是两个图形相交部分与合并后图形面积的比，用于在对象级别衡量预测显著目标与真实显著目标之间的误差，计算方法如下：

$$L_{\mathrm{IoU}}(P_i,G_{\mathrm{s}})=1-\frac{\sum_{(r,c)}G_{\mathrm{s}}(r,c)P_i(r,c)}{\sum_{(r,c)}P_i(r,c)+G_{\mathrm{s}}(r,c)-G_{\mathrm{s}}(r,c)P_i(r,c)} \tag{8}$$

对EPM中预测边缘的监督采用基础的BCE损失：

$$L_{\mathrm{e}}=L_{\mathrm{BCE}}(\mathrm{P}_1,\mathrm{G}_{\mathrm{e}}) \tag{9}$$

将显著图损失与边缘损失结合，得到最终损失函数：

$$L=\omega_{\mathrm{s}}L_{\mathrm{s}}+\omega_{\mathrm{e}}L_{\mathrm{e}} \tag{10}$$

式中：ω_{s}与ω_{e}在本文中都取1。

3　实验与分析

3.1　训练细节

本文使用PyTorch实现模型，并用ResNet-50的预训练模型初始化。在训练中，使用Adam优化器，参数为默认参数(betas=(0.9, 0.999), epsilon=1×10^{-8}, weight decay=0)。批大小为14，初始学习率为5×10^{-5}，每30代衰减至10 %，共训练50代。本文使用DUTS[23]数据集的训练集作为本文模型的训练集。在图片被输入网络之前，先缩放到336×336，并进行标准化，将取值范围限制到[0,1]。为充分利用训练集，提高模型泛化能力，本文使用色彩抖动、随机裁剪、随机水平翻转的数据增强方法。

3.2　数据集

为了充分评估本文提出模型的泛化能力，本文选取了6个被广泛使用的数据集用于评估，包括HKU-IS[24]、ECSSD[25]、PASCAL-S[26]、SOD[27]、DUT-OMRON[28]、DUTS，其中，对DUTS仅使用测试集进行评估。所有6个数据集都有逐像素的标注，并且每张图都至少有一个显著目标。这6个数据集都是现代SOD模型进行评估的常用数据集，具有如下特征：来自于多种多样的自然场景，拍摄距离、光照条件等不一；显著性目标或背景常常具有复杂的纹理；显著性目标常常具有复杂的轮廓；部分图片中有多个显著性目标，种类可能相同，也可能不同，其中，HKU-IS的所有图片都具有多个显著性目标。

3.3　评估方法

本文使用F-measure[29]、MAE(mean absolute error，平均绝对误差)、S-measure[30]、PR(precision-recall，精确率-召回率)曲线、F-measure曲线评估提出的模型。

F-measure是综合地考虑精确率和召回率的一种评估方法，定义如下：

$$F_\beta=\frac{(1+\beta^2)\cdot P\cdot R}{\beta^2\cdot P\cdot R} \tag{11}$$

式中：P和R分别代表精确率和召回率；β^2按经验设为0.3从而给予精确率更多的权重。在本文中报告的是平均F-measure，计算方法为每张显著图的阈值取所有像素平均值的两倍。

MAE的计算方法是，取显著图与真值之间的逐像素误差的平均值：

$$E_{\mathrm{MAE}}=\frac{1}{H\cdot W}\sum_{r=1}^{H}\sum_{c=1}^{W}\big|S(r,c)-G(r,c)\big| \tag{12}$$

式中：H和W表示图片的高与宽；S和G表示显著图和真值图；(r,c)表示像素坐标。数据集的MAE通过计算所有图片的MAE的平均值得到。

S-measure用于评估预测显著图与真值之间的结构相似度，由式(13)计算：

$$S=\alpha\cdot S_{\mathrm{o}}+(1-\alpha)\cdot S_{\mathrm{r}} \tag{13}$$

式中：S_{o}表示目标结构相似度；S_{r}表示区域结构相似度；α按经验设为0.5。

PR曲线是用于评估概率图的基本方法，精确率和召回率是通过比较数据集中所有图片的所有像素的预测结果和真值而来。在PR曲线上，每一个点代表[0,1]之间的某个阈值下的一对精确率和召回率。

与PR曲线类似，F-measure曲线上的每一个点

代表[0,1]之间的某个阈值下的F-measure。

3.4 结果对比

本文选取了近几年表现最优异的若干SOD模型，在最常用的数据集上进行定量评估，并与本文提出的方法进行对比，如表1与图5。表1中，F、M和S分别代表F-measure、MAE和S-measure，F-measure和S-measure越高越好，MAE越低越好，最好的结果以加粗表示，次之的结果以下划线表示，第三的结果以斜体与下划线表示。参数量表示整个网络的参数数量，单位为百万(M)，FPS(frames per second)表示该模型在GTX 1080 Ti显卡上预测时每秒可以处理的图片数量。PiCANet选择以ResNet为骨干网络的模型进行评估，CAGNet使用完整的CAGNet-V评估，RASNet使用v2版本做评估。本文模型在5个数据集上达到了最佳MAE，其中，HKU-IS、PASCAL-S和DUT-OMRON分别降低了0.1%、0.5%和0.4%；在5个数据集上达到了至少第二的S-measure，在2个数据集上达到了至少第二的F-measure。可知，本文模型在MAE和S-measure上较有优势。在复杂度方面，本文模型参数量处于中等水平，预测速度可以初步满足一般场景的实时性要求。在对比模型中，HVPNet与SAMNet是以轻量化为目标设计的，但也明显损失了预测效果。

表1 F-measure、MAE以及S-measure

Tab. 1 F-measure, MAE, and S-measure

模型	参数量/M	FPS	HKU−IS			ECSSD			PASCAL−S		
			F	M	S	F	M	S	F	M	S
AFNet[31]	21.08	19	0.888	0.036	0.905	*0.908*	0.042	0.913	**0.820**	0.070	0.849
Amulet[6]	33.15	10	0.841	0.051	0.886	0.868	0.059	0.894	0.757	0.100	0.818
BASNet[32]	87.06	32	*0.896*	0.032	*0.909*	0.880	*0.037*	0.916	0.771	0.076	0.838
BDMP[8]	22.09	—	0.871	0.039	0.907	0.868	0.045	0.911	0.762	*0.074*	0.844
CAGNet[33]	20.98	—	0.905	0.033	0.897	0.915	0.042	0.898	0.819	0.076	0.827
HVPNet[34]	1.23	312	0.872	0.045	0.899	0.889	0.052	0.904	0.784	0.089	0.830
PAGRN[9]	—	—	0.886	0.048	0.887	0.894	0.061	0.889	*0.799*	0.089	0.822
PiCANet[13]	32.85	5	0.870	0.043	0.904	0.886	0.046	0.917	0.792	0.076	**0.854**
RASNet[10]	24.59	40	**0.906**	0.030	0.915	**0.923**	0.034	*0.925*	—	—	—
SAMNet[35]	1.33	332	0.871	0.045	0.898	0.891	0.050	0.907	0.778	0.092	0.826
SRM[7]	43.74	12	0.874	0.046	0.887	0.892	0.054	0.895	0.792	0.084	0.834
U2Net[36]	44.02	30	*0.896*	*0.031*	**0.916**	0.892	**0.033**	**0.928**	0.770	*0.074*	*0.845*
UCF[37]	23.98	12	0.823	0.062	0.875	0.844	0.069	0.883	0.726	0.116	0.806
本文模型	30.28	31	0.905	**0.029**	**0.916**	0.900	**0.033**	0.926	0.797	**0.065**	**0.854**

模型	参数量/M	FPS	SOD			DUTS			DUT−OMRON		
			F	M	S	F	M	S	F	M	S
AFNet[31]	21.08	19	—	—	—	0.793	0.046	0.867	0.739	*0.057*	0.826
Amulet[6]	33.15	10	0.741	0.144	0.755	0.678	0.085	0.804	0.647	0.098	0.781
BASNet[32]	87.06	32	0.744	*0.112*	0.772	0.791	0.048	0.866	0.756	*0.057*	*0.836*
BDMP[8]	22.09	—	0.761	0.106	0.790	0.745	0.050	0.862	0.692	0.064	0.810
CAGNet[33]	20.98	—	—	—	—	0.822	*0.045*	0.852	0.744	*0.057*	0.807
HVPNet[34]	1.23	312	*0.779*	0.122	0.765	0.749	0.058	0.849	0.721	0.065	0.831
PAGRN[9]	—	—	0.770	0.145	0.720	0.784	0.056	0.839	0.711	0.071	0.775
PiCANet[13]	32.85	5	**0.785**	**0.103**	**0.793**	0.759	0.051	0.869	0.717	0.065	0.832
RASNet[10]	24.59	40	—	—	—	**0.831**	**0.037**	**0.884**	0.763	0.055	*0.836*
SAMNet[35]	1.33	332	0.780	0.124	0.762	0.745	0.058	0.849	0.717	0.065	0.830
SRM[7]	43.74	12	0.780	0.126	0.745	0.753	0.059	0.836	0.707	0.069	0.780
U2Net[36]	44.02	30	0.769	0.106	*0.789*	0.792	*0.045*	*0.874*	*0.761*	0.055	**0.847**
UCF[37]	23.98	12	0.737	0.148	0.763	0.631	0.112	0.782	0.621	0.120	0.760
本文模型	30.28	31	0.773	**0.103**	0.788	*0.813*	0.039	0.880	**0.765**	**0.051**	0.838

对上述算法在数据集DUT-OMRON、DUTS、ECSSD、HKU-IS、PASCAL-S、SOD上绘制了F-measure曲线和PR曲线，结果如图5和图6所示。曲线的位置越靠上说明效果越好，粗实线是本文测试结果，可以看出其基本上都在最高的位置。不过，在DUT-OMRON数据集中，本文方法不如U2Net，在DUTS数据集中，本文方法不如RASNet，说明本文方法在特定场景下的泛化能力仍有提升空间。

从上述数据集中选取了6张有代表性的图片进行测试，在各算法之间进行定性对比，如图7，第一列是原图，第二列是真值图，第三列是本文结果，随后是对比模型的结果。其中，从a、b、d、e看出，本文的算法可以

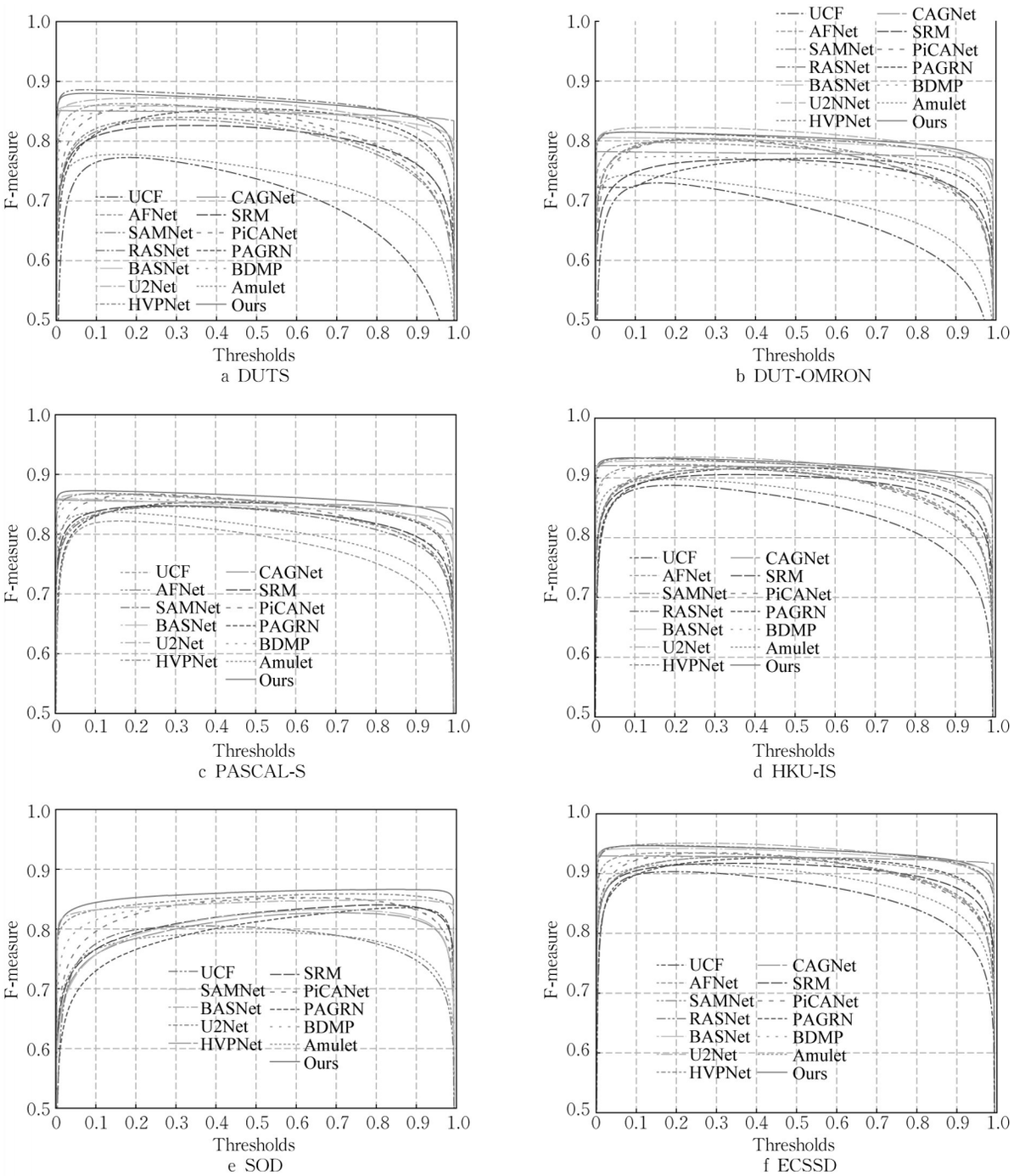

图5　F-measure 曲线

Fig. 5　Curves of F-measure

更完整地预测出显著性目标区域，并有效排除非显著性目标区域；从c、f、g看出，本文的算法预测出的显著性目标有着精细的边缘，验证了边缘信息提取的有效性。

3.5　消融实验

为了充分验证本文所提出创新点的效果，本文进行了消融实验，见表2。依次在网络中添加注意力和EPM，并在ECSSD上评估F-measure、MAE和S-measure。表2中，注意力的N、P、E分别代表负注意力、正注意力和边缘注意力。在无EPM时，从仅有负注意力到三重注意力提升0.35 %，在有EPM时，提升为0.18 %。对比无EPM和有EPM时，三种注意力条件下，F-measure分别提高了0.66 %、1.05 %和0.49 %。最后，从只有负注意力、无EPM到有三重注意力、有EPM，F-measure提高了0.84 %。综上，本文提出的TAM和EPM均对模型的结果起到了提升效果，且两者结合后效果更好。

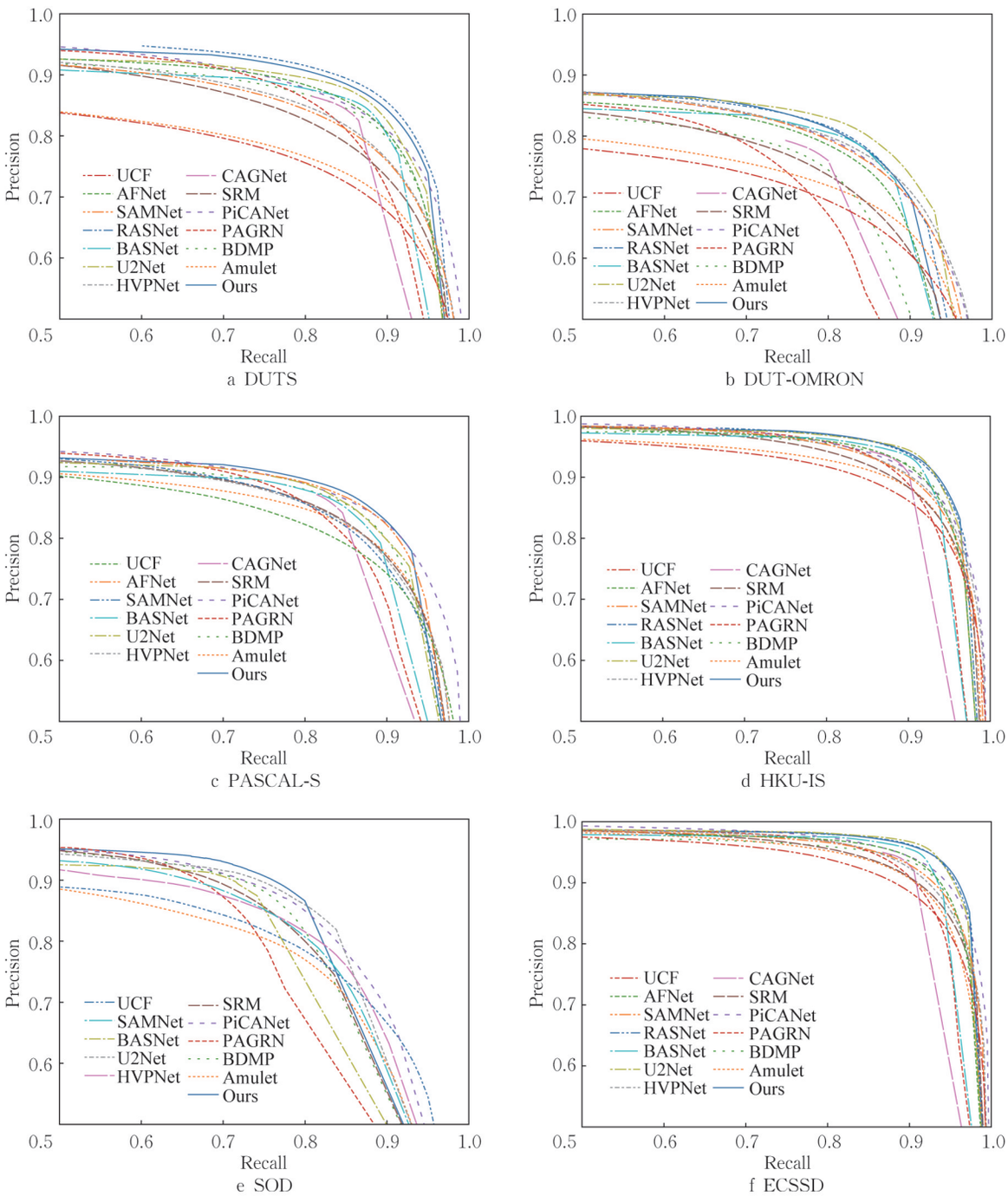

图6 PR曲线

Fig. 6 Curves of PR

单独对边缘融合进行消融实验,对比不融合边缘预测结果(EP)和融合边缘预测结果(EPM)时效果的差异,见表3。其中,基准和表2中只使用负注意力、不使用EPM的网络是一致的;EP代表用边缘真值监督网络第1层输出的边缘预测,但边缘预测结果不再被输出到其他地方;相对于EP,EPM则是用边缘预测结果与深层的显著图融合,进一步优化显著图。

表3中的EP(简)表示使用简单的几次卷积预测边缘,而EP(RSU)表示使用RSU预测边缘。从基准到EP(简)时F-measure下降了,但从基准到EP(RSU)和从EP(RSU)到EPM,F-measure依次提升了0.33 %和0.33 %,总提升为0.66 %。综上,EPM在边缘预测基础上与显著图融合,有助于进一步细化显著预测图,且使用复杂度较高的RSU预测边缘是必要的。

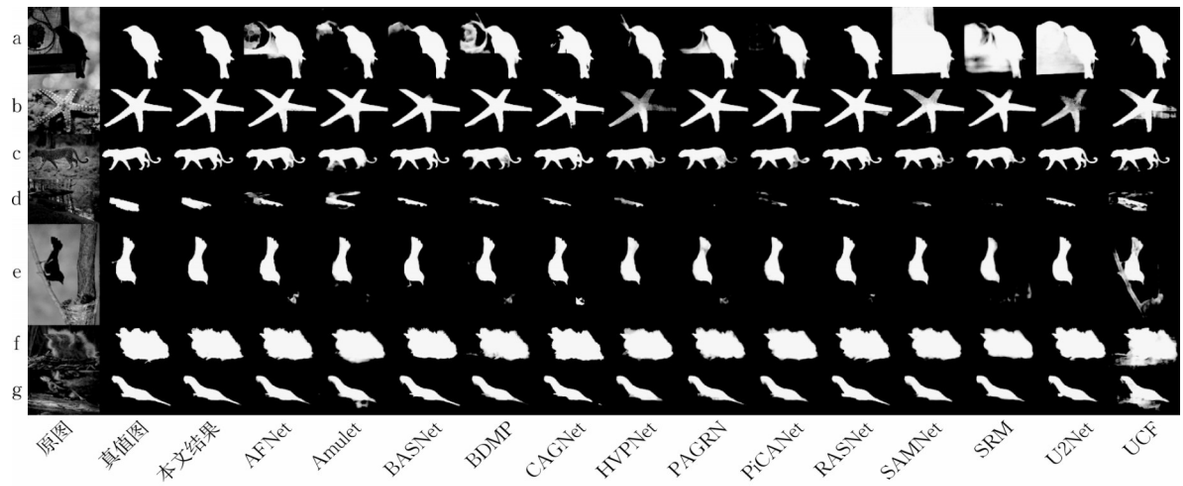

图7　定性对比

Fig. 7　Qualitative comparison

表2　注意力与EPM消融实验

Tab. 2　Ablation study on attention and EPM

注意力			EPM	F−measure	MAE	S−measure
N	P	E				
√				0.891 3	0.036 5	0.919 1
√			√	0.897 9	0.033 9	0.924 4
√	√			0.888 0	0.035 2	0.922 4
√	√		√	0.898 5	0.033 7	0.924 4
√	√	√		0.894 8	0.035 2	0.921 8
√	√	√	√	**0.899 7**	**0.033 0**	**0.925 9**

表3　边缘融合消融实验

Tab. 3　Ablation study on edge fusing

模型	F−measure	MAE	S−measure
基准	0.891 3	0.036 5	0.919 1
EP(简)	0.890 7	0.035 1	0.920 9
EP(RSU)	0.894 6	0.036 8	0.920 2
EPM	**0.897 9**	**0.033 9**	**0.924 4**

4　结语

在本文中，针对常用SOD算法的结果中目标边缘较为模糊的问题，本文提出了一种边缘信息增强的SOD网络。该网络的主体结构是自顶向下逐层优化的，能够提取多尺度的信息。在此基础上，本文引入了两个模块以增强边缘信息的提取。首先，本文提出了TAM，融合了前景、背景和边缘注意力，并且在不增加任何参数的前提下就能从预测图中直接得出；其次，本文提出了EPM，其位于网络最浅层，使用较高分辨率的特征以有监督的方式预测边缘，并于网络深层的预测图融合，保留了更多的边缘细节信息。TAM与EPM互为补充，有效地提高了显著图预测的效果。本文在6个常用SOD数据集上用三种定量指标评估了本文模型，在HKU-IS、PASCAL-S和DUT-OMRON上把MAE分别降低了0.1 %、0.5 %和0.4 %；本文还以定性的方式展示了本文模型与近几年SOD模型的预测结果，体现出本文模型能够更完整地预测显著目标，并且能够精确地预测目标边缘。本文模型参数量为30.28M，可以在GTX 1080 Ti上达到31FPS的预测速度。最后，用消融实验证明了本文提出创新点的有效性。

作者贡献声明：

赵卫东：设计框架、技术指导、论文审定。

王　辉：实验研究、论文撰写。

柳先辉：技术指导、论文审定。

参考文献：

[1] CORBETTA M, SHULMAN G L. Control of goal-directed and stimulus-driven attention in the brain [J]. Nature Reviews Neuroscience, 2002, 3(3): 201.

[2] BORJI A, CHENG M M, JIANG H, *et al*. Salient object detection: A benchmark [J]. IEEE Transactions on Image Processing, 2015, 24(12): 5706.

[3] WANG W, LAI Q, FU H, *et al*. Salient object detection in the deep learning era: An in-depth survey[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 44(6): 3239.

[4] XIE S, TU Z. Holistically-nested edge detection[C]//Proceedings of the IEEE International Conference on Computer Vision. Boston: IEEE, 2015: 1395-1403.

[5] LI G, YU Y. Deep contrast learning for salient object detection [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 478-487.

[6] ZHANG P, WANG D, LU H, *et al*. Amulet: Aggregating multi-level convolutional features for salient object detection [C]// Proceedings of the IEEE International Conference on Computer Vision. Venice: IEEE, 2017: 202-211.

[7] WANG T, BORJI A, ZHANG L, *et al*. A stagewise refinement

model for detecting salient objects in images[C]//Proceedings of the IEEE International Conference on Computer Vision. Venice: IEEE, 2017: 4019-4028.

[8] ZHANG L, DAI J, LU H, *et al*. A bi-directional message passing model for salient object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 1741-1750.

[9] ZHANG X, WANG T, QI J, *et al*. Progressive attention guided recurrent network for salient object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 714-722.

[10] CHEN S, TAN X, WANG B, *et al*. Reverse attention for salient object detection[C]//Proceedings of the European Conference on Computer Vision (ECCV). Munich: Springer Science, 2018: 234-250.

[11] LI T, SONG H, ZHANG K, *et al*. Recurrent reverse attention guided residual learning for saliency object detection [J]. Neurocomputing, 2020, 389: 170.

[12] LI J, PAN Z, LIU Q, *et al*. Complementarity-aware attention network for salient object detection[J]. IEEE Transactions on Cybernetics, 2020, 52(2): 873.

[13] LIU N, HAN J, YANG M H. Picanet: Learning pixel-wise contextual attention for saliency detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 3089-3098.

[14] ZHANG Z, LIN Z, XU J, *et al*. Bilateral attention network for RGB-D salient object detection[J]. IEEE Transactions on Image Processing, 2021, 30: 1949.

[15] HU P, SHUAI B, LIU J, *et al*. Deep level sets for salient object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 2300-2309.

[16] LUO Z, MISHRA A, ACHKAR A, *et al*. Non-local deep features for salient object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 6609-6617.

[17] WEI J, WANG S, WU Z, *et al*. Label decoupling framework for salient object detection [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020: 13025-13034.

[18] HE K, ZHANG X, REN S, *et al*. Deep residual learning for image recognition [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 770-778.

[19] ZEILER M D, FERGUS R. Visualizing and understanding convolutional networks[C]//European Conference on Computer Vision. Cham: Springer, 2014: 818-833.

[20] IOFFE S, SZEGEDY C. Batch normalization: Accelerating deep network training by reducing internal covariate shift [C]// International Conference on Machine Learning. Lille: PMLR, 2015: 448-456.

[21] GLOROT X, BORDES A, BENGIO Y. Deep sparse rectifier neural networks[C]//Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics. Fort Lauderdale: PMLR, 2011: 315-323.

[22] QIN X, ZHANG Z, HUANG C, *et al*. U2-Net: Going deeper with nested U-structure for salient object detection[J]. Pattern Recognition, 2020, 106: 107404.

[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. Honolulu: IEEE, 2017: 136-145.

[24] LI G, YU Y. Visual saliency detection based on multiscale deep CNN features[J]. IEEE Transactions on Image Processing, 2016, 25(11): 5012.

[25] YAN Q, XU L, SHI J, *et al*. Hierarchical saliency detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Portland: IEEE, 2013: 1155-1162.

[26] 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. Columbus: IEEE, 2014: 280-287.

[27] MOVAHEDI V, ELDER J H. Design and perceptual validation of performance measures for salient object segmentation[C]//2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops. San Francisco: IEEE, 2010: 49-56.

[28] 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. Portland: IEEE, 2013: 3166-3173.

[29] MARGOLIN R, ZELNIKMANOR L, TAL A. How to evaluate foreground maps?[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Columbus: IEEE, 2014: 248-255.

[30] 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. Venice: IEEE, 2017: 4548-4557.

[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. Long Beach: IEEE, 2019: 1623-1632.

[32] 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. Long Beach: IEEE, 2019: 7479-7489.

[33] MOHAMMADI S, NOORI M, BAHRI A, *et al*. CAGNet: Content-aware guidance for salient object detection[J]. Pattern Recognition, 2020, 103: 107303.

[34] LIU Y, GU Y C, ZHANG X Y, *et al*. Lightweight salient object detection via hierarchical visual perception learning [J]. IEEE Transactions on Cybernetics, 2020, 51(9): 4439.

[35] LIU Y, ZHANG X Y, BIAN J W, *et al*. SAMNet: Stereoscopically attentive multi-scale network for lightweight salient object detection [J]. IEEE Transactions on Image Processing, 2021, 30: 3804.

[36] QIN X, ZHANG Z, HUANG C, *et al*. U2-Net: Going deeper with nested U-structure for salient object detection[J]. Pattern Recognition, 2020, 106: 107404.

[37] ZHANG P, WANG D, LU H, *et al*. Learning uncertain convolutional features for accurate saliency detection [C]// Proceedings of the IEEE International Conference on Computer Vision. Venice: IEEE, 2017: 212-221.