

边缘信息引导多级尺度特征融合的显著性目标检测方法

王向军^{1,2}, 李名洋^{1,2}, 王霖^{1,2}, 刘峰^{1,2}, 王玮^{1,2}

1. 天津大学精密测试技术及仪器国家重点实验室, 天津 300072;
2. 天津大学微光机电系统技术教育部重点实验室, 天津 300072)

摘要: 针对基于 FCN 和 U 型网络架构的深度学习显著性目标检测方法提取的显著性图存在边界不清晰和结构不完整的问题, 文中提出了一种基于边缘信息引导多级尺度特征融合网络 (EGMFNet)。EGMFNet 使用多通道融合残差块 (RCFBlock) 以嵌套的 U 型网络架构作为主干模型。同时, 在网络的较低层级引入具有边缘信息引导的全局空间注意力模块 (EGSAM) 以增强空间特征及边缘特征。此外, 在损失函数中引入了图像边界损失, 用于提升显著性图的质量并在学习过程中保留更加清晰的边界。在四个基准数据集上进行实验, 实验结果表明, 文中方法的 F 值较典型方法提升 1.5%、2.7%、1.8% 和 1.6%, 验证了 EGMFNet 网络模型的有效性。

关键词: 显著性目标检测; 多尺度特征融合; 边缘信息引导; 空间注意力模块; 边界损失函数
中图分类号: TP391.4 **文献标志码:** A **DOI:** 10.3788/IRLA20220344

Salient object detection method based on multi-scale feature-fusion guided by edge information

Wang Xiangjun^{1,2}, Li Mingyang^{1,2}, Wang Lin^{1,2}, Liu Feng^{1,2}, Wang Wei^{1,2}

1. State Key Laboratory of Precision Measuring Technology and Instruments, Tianjin University, Tianjin 300072, China;
2. MOEMS Education Ministry Key Laboratory, Tianjin University, Tianjin 300072, China)

Abstract: In this paper, an Edge-information Guided Multi-scale Feature-fusion Network (EGMFNet) is proposed to solve the problems of unclear boundary and incomplete structure of saliency map extracted by deep learning saliency target detection method based on FCN and U-shaped network architecture. EGMFNet uses Residual multi-Channel Fusion Block (RCFBlock) and uses a nested U-shaped network architecture as the backbone model. At the same time, an Edge-information Guided Global Spatial Attention Module (EGSAM) is introduced at the lower level of the network to enhance spatial features and edge features. In addition, image boundary loss is introduced into the loss function, which is used to improve the quality of saliency map and keep clearer boundaries in the learning process. Experiments on four benchmark data sets show that the F values of the proposed method are increased by 1.5%, 2.7%, 1.8% and 1.6% compared with typical methods, which verifies the effectiveness of EGMFNet network model.

Key words: salient object detection; multi-scale feature-fusion; edge-information guidance; spatial attention module; boundary loss function

收稿日期: 2022-05-20; 修订日期: 2022-08-01

作者简介: 王向军, 男, 教授, 博士, 主要从事光电传感与测试、计算机视觉与图像分析、微小型光机电系统及 MEMS 方面的研究。

0 引言

人类在通过视觉系统感知外部环境时,总是有选择性的捕捉场景中关键性的信息和感兴趣的区域,并忽略不感兴趣或是无效的信息^[1-2]。在计算机视觉中,研究人员开启了显著性目标检测方法的研究以尝试复现这种能力。显著性目标检测旨在检测并分割出输入图像中关键的或极具吸引力的对象或区域^[3-5]。目前,显著性目标检测作为一种优秀的预处理手段,在许多计算机视觉任务中发挥着重要作用,例如目标跟踪、目标检测^[6]、图像编辑等。

近年来,U型网络(UNET)依靠其强大的特征提取能力,在图像分割领域取得了不错的成果,特别是在数据量较少时U型网络的性能表现尤为出色。此类模型巧妙地利用了编码器-解码器结构,编码器“由浅入深”的提取特征,并在解码器阶段融合多极特征信息以形成显著性图,这有效的利用了高级特征的语义信息及低级特征的空间信息以引导显著性目标的定位。

Huang等^[3]在网络引入四个具有特定扩张率扩张卷积的并行剩余连通块细化各层级提取的特征以获取上下文信息,并在最终阶段利用上下文信息引导显著性图的预测。Ning等^[7]在特征融合阶段的特定层级引入SNet以在通道层级上指导特征融合的通道信息保留,同时利用高级特中丰富的语义信息指导特征的提取。Li等^[8]利用聚合的不同层次、多尺度特征引导特征的提取及融合,通过这种方式使包含用于定位完整对象的上下文的高级特征将保留到最终输出层,而包含空间结构细节的低级特征将嵌入到每个层中以保留空间结构细节。Jia^[5]等将显着对象检测中的复杂场景转换为信息溶解问题,并利用一个包含正向连接和反向连接的自网络为每一对编解码模块的输出赋予一个权值以指导显著性图的预测。Chen等^[9]在网络中引入U形自细化模块生成多度特征在经过通道压缩后利用U形特征继承模块进行低级、高级特征和全局信息的融合以此增强网络所预测显著性图的细节特征。

然而,显著性目标检测仍然面临着诸多问题:(1)现有显著性目标检测方法在面对复杂的背景信息和目标本身的多样性将影响目标定位的准确性^[10]。网络中引入空间注意力模块、上下文交流模块、多尺

度特征融合模块等增加计算量的同时并不能有效的解决目标定位不准确的问题^[11]。(2)低级特征包含着大量空间结构细节信息以及较为清晰的边缘信息,这些信息将会随着特征层的逐渐深入而丢失,并且仅能在融合阶段的较低层级访问,这使得所提取出的显著性图具有较低质量的目标边界且可能导致图像模糊、结构不完整的情况。

针对边缘不清晰的问题,文中提出了一种利用边缘信息引导空间特征的多级通道特征融合网络(EGMFNet)用于显著性目标检测,其中通过逐级堆叠新设计的通道特征融合模块(RCFBlock)形成三级嵌套的U型结构。提出一种新颖的基于边缘信息引导的全局空间注意力模块(EGSAM),该模块并行地提取边缘信息及全局空间特征并进行融合以强化边缘信息及空间信息,从而保证所预测的显著性图保留更多的细节信息并具有清晰的边缘。同时,在损失函数中引入边界损失函数,使网络能够从显著性真值中学习边界细节。

1 文中方法

详细介绍提出的基于边缘信息引导的多通道融合网络。网络主要包括两个重要部分:一个三级嵌套的U型主体网络;在较低层级为强化空间细节特征而引入的由边缘信息引导的扩张空间注意力模块。首先,将介绍基础的通道融合残差块以及由RCFBlock堆叠而成的次级U型块——MCFUBlock。其次,将介绍基于边缘信息引导的扩张空间注意力模块。最后,详细描述所提出的EGMFNet网络完整结构以及训练中使用的损失函数。

1.1 通道融合残差块及堆叠方式

高级特征的语义信息能够指导显著性目标的定位,而低级特征所包含的空间特征是保证预测所得显著性图结构完整的关键。随着网络加深,部分空间特征在传播过程中随着池化操作而丢失,并且网络内部缺乏多级特征信息的交流限制了网络获取上下文信息的能力。为了解决上述问题,PoolNet使用不同内核的平均池化层生成不同感受野大小的特征图并进行融合,然而分辨率较小的特征在经过大采样率的上采样操作后可能会出现特征退化的情况。此外,众多方法尝试利用扩张卷积以扩大感受野,但在输入特征

上使用多个扩张卷积进行多尺度特征提取会导致内存和计算成本增加。

为了充分发掘多尺度特征信息,提出了以通道融合残差块为基础并堆叠形成的次级 U 型块。图 1 为通道融合残差块 (RCFBlock) 的结构细节。对于输入图像 $I_0 \in R^{C \times H \times W}$, 首先使用两个卷积层 $conv_{in}$ 和 $conv_1$ 以获得特征图 I_{in} 和 I_1 , 随后使用最大池化层进行尺度变换操作并输入 $conv_2$ 以获得特征图 I_2 。

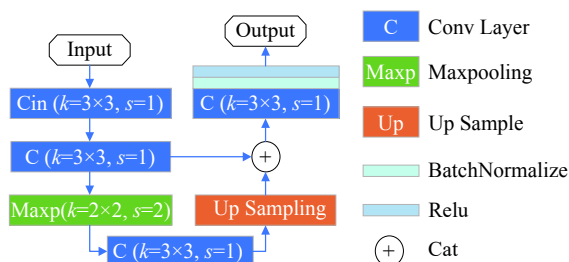


图 1 通道融合残差块 (RCFBlock) 结构示意图

Fig.1 Structural diagram of channel fusion residual block (RCFBlock)

小尺寸特征图 I_2 经上采样操作后与 I_1 进行通道维度上的拼接并使用 $conv_3$ 融合, 最后经过激活函数和批标准化操作得到 I_{out} 。过程可描述为公式 (1)~(2):

$$I_F = conv_3(I_1 \oplus Up(I_2 : I_1)) \quad (1)$$

$$I_{out} = f_{relu}(f_{BN}(I_F)) \quad (2)$$

式中: 卷积核大小均为 3×3 , 步长设置为 1; \oplus 表示通道维度的拼接操作; $Up(I_* : I)$ 表示利用双线性插值方法进行上采样操作。最终, RCFBlock 的输出结果为融合后的特征图 I_F 经过批标准化和激活操作后获得的 I_{out} 。

图 2 为 RCFBlock 堆叠形成的次级 U 型块 MCFU3 Block 的详细结构。其中, 3 表示堆叠的层级, 因此 MCFU2 表示堆叠两级 RCFBlock, MCFU1 表示仅一级 RCFBlock 且不包含池化操作。以 MCFU3 为例叙述 MCFUBlock 的详细计算流程。MCFU3 左侧为编码模块, 右侧为解码模块。对于输入图像 $I_0 \in R^{C \times H \times W}$, MCFU3 渐进的提取三种尺度的特征并在解码阶段分别融合以获得最终的输出结果。过程如公式 (3)~(5) 所示:

$$I_{i+1} = F_{max}(F_{RCF}(I_i)), i = 0, 1 \quad (3)$$

$$I_3 = f_{relu}(f_{BN}(conv(I_2))) \quad (4)$$

$$I_{out} = F_{RCF}(I_1 \oplus Up(\tilde{I} : I_1)) \quad (5)$$

式中: F_{RCF} 表示通过 RCFBlock 进行处理; F_{max} 表示最大池化操作。

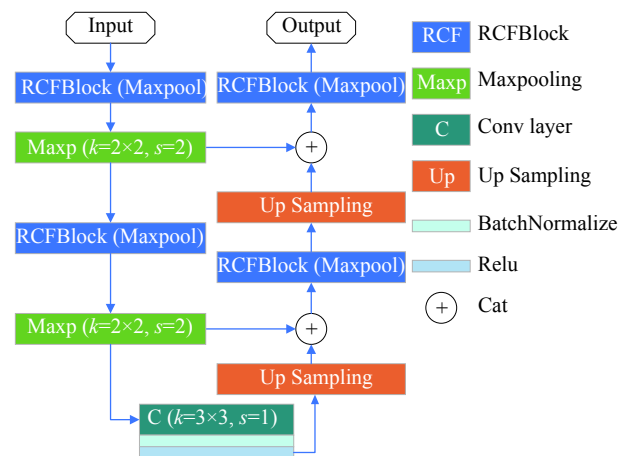


图 2 MCFUBlock 的结构示意图

Fig.2 Structural diagram of MCFUBlock

对于 RCFBlock, 池化层、上采样层以及通道维度拼接操作使其能够提取并整合两个尺度的特征。整体上, 由 RCFBlock 堆叠而成的 MCFU-Block 能够进一步的提取并整合多尺度特征信息, 这样递进式的提取及整合多尺度特征信息能够帮助网络保留更多的空间特征信息。对于整个网络而言, MCFU-Block 能够在不改变输入输出尺度的同时有效利用多尺度特征信息, 这在一定程度上保证了网络预测的显著性图具有足够的空间细节。

1.2 基于边缘信息引导的空间注意力模块

空间信息丢失使得网络无法预测出具有清晰边界的显著性图。文中在空间注意力模块中引入空洞卷积, 这使得注意力模块在指数级扩展感受野的情况下不会丢失分辨率, 并且不增加过多的计算量。

如图 3 所示, EGSAM 由两个部分组成, 包括提取并融合边缘信息的边缘信息引导模块 (EGM) 和具有扩张卷积的空间注意力模块 (ESM)。其中, 边缘信息引导模块利用 Sobel 卷积提取输入图像 $I_0 \in R^{C \times H \times W}$ 的边缘图 I_{Sobel} , 随后进行空间全局最大池化和均值池化, 将生成的二维向量 $E_{Max} \in R^{1 \times H \times W}$ 和 $E_{Avg} \in R^{1 \times H \times W}$ 进行通道维度拼接后进行卷积操作, 经激活得到二维边缘信息引导向量 $A_{edge} \in R^{1 \times H \times W}$ 。边缘信息引导模块的整体过程可以被表示为公式 (6)~(7):

$$I_{Sobel} = conv_{Sobel}(I_0) \quad (6)$$

$$A_{\text{edge}} = F_s(\text{conv}(E_{\text{Max}} \oplus E_{\text{Avg}})) \quad (7)$$

式中: $\text{conv}_{\text{Sobel}}$ 表示 Sobel 卷积; F_s 表示激活函数。

扩张的空间注意力模块 (ESM) 的输入为融合边缘引导信息的原始输入图像 $I_{EG} \in R^{C \times H \times W}$ 。为减小计算量并节约计算资源,对输入图像进行 4 倍下采样操作,获得 $I_{EG_D} \in R^{C \times H/4 \times W/4}$ 。随后,引入卷积核大小为 3×3 ,空洞率分别为 2、4、6 的空洞卷积。将融合后的特征图 $I_{DilatedF} \in R^{C \times H/4 \times W/4}$ 按空间进行全局最大池化和均值池化,将生成的二维向量 $S_{\text{Max}} \in R^{C \times H/4 \times W/4}$ 和 $S_{\text{Avg}} \in R^{C \times H/4 \times W/4}$ 进行通道维度拼接后进行卷积操作,

经激活后得到二维边缘信息引导向量 $A_{\text{Spatial}} \in R^{1 \times H/4 \times W/4}$ 。扩张的空间注意力模块 (ESM) 的过程如公式 (8)~(9):

$$I_{D_i} = \text{conv}_{D_i}(I_{EG_D}), i = 0, 1, 2 \quad (8)$$

$$A_{\text{Spatial}} = F_s(\text{conv}(S_{\text{Max}} \oplus S_{\text{Avg}})) \quad (9)$$

式中: conv_{D_i} 表示空洞卷积。随后,对 A_{Spatial} 进行上采样操作使其恢复为 $A_{\text{Spatial}} \in R^{1 \times H \times W}$,将恢复后的 A_{Spatial} 与原始输入图像 I_0 进行融合获得最终的输入 $I_{EGSAM} \in R^{C \times H \times W}$ 。

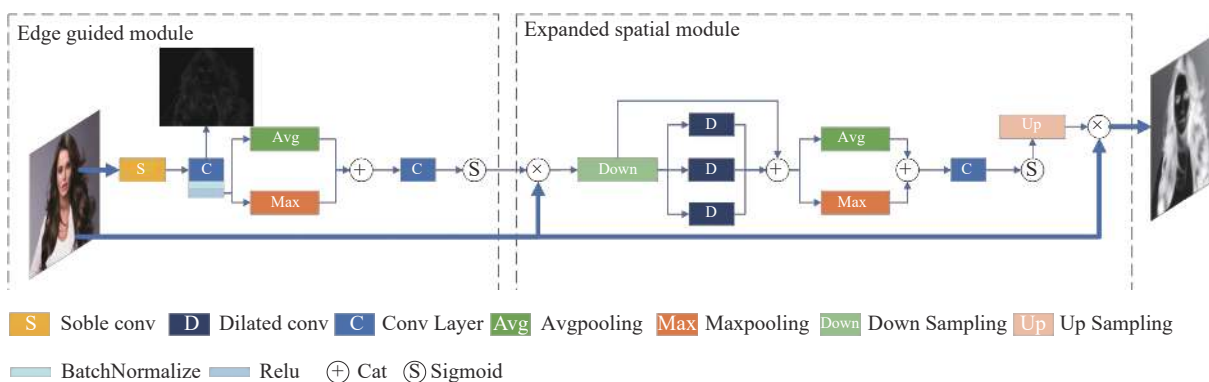


图 3 基于边缘信息引导的扩张空间注意力模块 (EGSAM) 结构示意图

Fig.3 Structural diagram of expanded spatial attention module guided by edge information (EGSAM)

1.3 EGMFNet 网络架构

EGMFNet 的整体结构是一个标准的 U 型结构,如图 4 所示,EGMFNet 共包括六个阶段。在前三个阶段中,每一个阶段包括一级编码器、一级解码器及桥接在中间的 EGSAM,编解码器分别使用 MCFU3、MCFU2、MCFU1。第四阶段和第五阶段由一组编解码器组成(如图 5 所示)。第六阶段仅由一级使用 RUBlock 的解码器组成。在网络的较低阶段(一至四阶段)使用具有两级嵌套的 MCFU 模块进行填充以在尺度较大的前提下获取更多的空间信息。在网络的较高阶段(五至六阶段)选择具有一定深度的具有残差连接的 U 型块 (RUBlock),并且 RCFBlock 的堆叠数量随着层级加深和图像尺度减小而逐级递减以防止过度池化导致空间信息丢失。原因是 RCFBlock 及 MCFUBlock 中包括大量池化层,对小尺度图像做过多池化操作会造成空间特征进一步丢失同时无法获取准确的语义信息。

第一至三阶段每一级编解码器间接桥接了 EGSAM

模块,EGSAM 的输出以加权的方式与拼接后的原始解码器输入再次进行拼接,该过程可以表示为:

$$I_{\text{in_}E_i} = \alpha I_{EG_i} + I_{\text{in}}, i = 1, 2, 3 \quad (10)$$

式中: $I_{\text{in_}E_i}$ 表示前三阶段解码器的最终输入; I_{EG_i} 表示 EGSAM 模块的输出; α 表示权值被设定为 0.01; +表示按元素加和操作。

每一级解码器的侧边都有一个输出层用于预测显著性图,六阶段预测所获得的显著性图在上采样至相同尺度后进行第一次拼接。同时,在前三个阶段中 EGSAM 模块的侧边增加了输出层,并将预测的显著性图进行第二次拼接。最后,将两次拼接的结果进行拼接与融合作为网络最终的输出结果。

如图 5 所示, RUBlock 是典型的 U 型结构模块。编码器阶段共包含四层:第一层作为输入层;第二层至第四层使用卷积核大小 3×3 且空洞率分别为 2、4、8 的空洞卷积。解码器阶段共包含三层:第一层与输入层相同;第二、三层使用卷积核大小 3×3 且空洞率分别为 2、4 的空洞卷积。

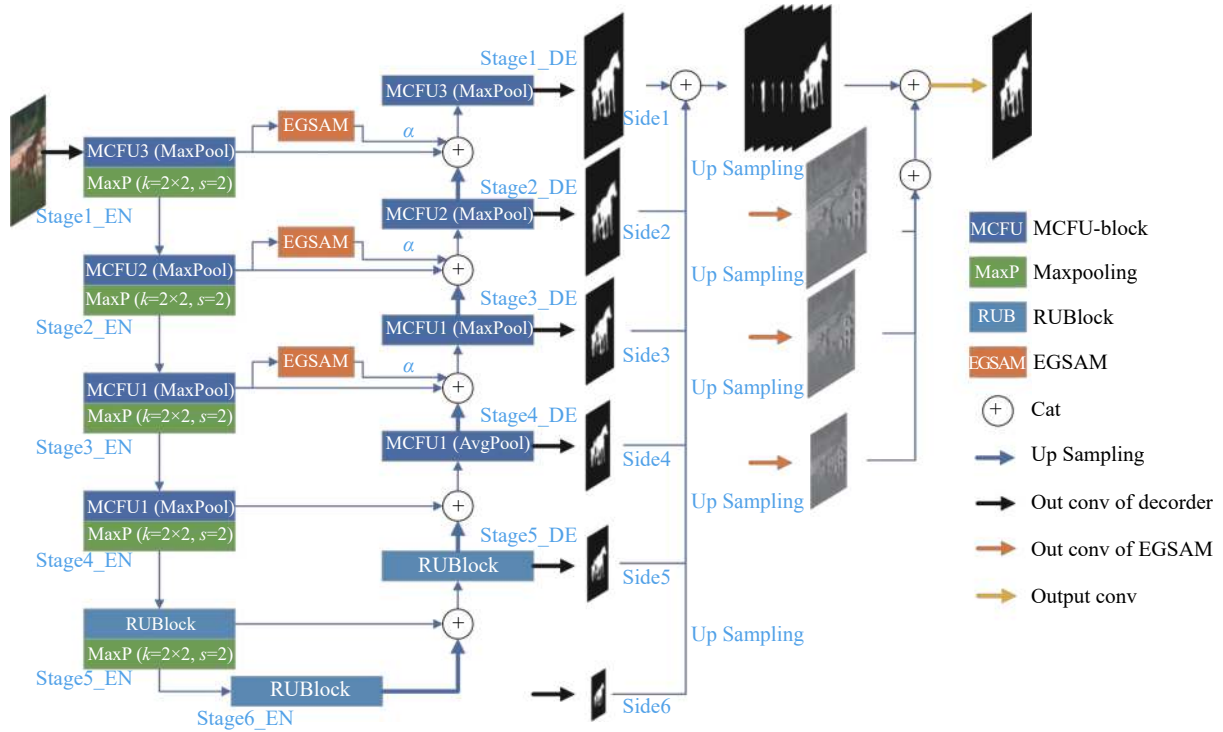


图 4 完整 EGMFNet 的结构示意图

Fig.4 Structure diagram of complete EGMFNet

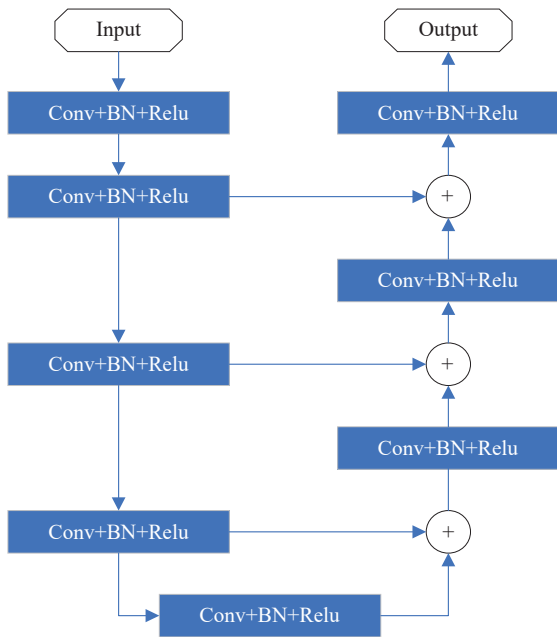


图 5 具有残差连接的 U 型块 (RUBlock) 结构示意图

Fig.5 Structure diagram of U-block with residual connection

1.4 监督方式

在训练阶段,文中使用交叉熵损失 (BCE Loss) 和边界损失 (Boundary Loss)^[12] 计算每一阶段预测获得显著性图的损失。损失函数的定义如公式 (11) 所示:

$$L_s = (1 - \alpha)L_{BCE}^s + \alpha L_{BL}^s, s = 1 \sim 6 \quad (11)$$

式中: L_{BCE}^s 表示交叉熵损失; L_{BL}^s 表示边界损失; s 表示每一个阶段; α 为权值被设定为 0.01 的固定值。边界损失的定义如公式 (12) 所示:

$$L_{BL} = \int_{\Omega} \varphi_G(p) s_G(p) dp \quad (12)$$

式中: $\varphi_G(p)$ 表示预测显著性图的边界质量; $s_G(p)$ 是显著性区域的二元指示函数。当像素点 p 属于标注图像显著性区域时 $s_G(p) = 1$, 否则 $s_G(p) = 0$ 。

交叉熵损失逐像素的计算预测结果相较于标注目标的误差,训练过程尝试减小这种差别,因此交叉熵损失在图像层面引导网络获取更加准确的显著性图。但是,这种逐像素的引导方式并不能给予边界信息和结构信息足够的关注,因此可能导致预测结果边界不清晰和结构不完整的情况。

边界损失函数以积分的方式计算预测显著性目标与标注显著性目标边界的距离,这在反向传播过程中能够引导网络,并最小化预测显著性目标与标注显著性目标中边界和结构上不匹配的部分。因此,文中在交叉熵损失的基础上以加权的方式引入边界损失函数,从理论上,这种混合损失在引导网络准确预测

显著性图的同时进一步关注边界和结构信息,以保证预测显著性图具有清晰的边界和完整的结构。

最终,文中利用深度监督的方式进行训练,用于反向传播的损失函数定义如公式(13)所示:

$$L = \sum_s L_s + L_f \quad (13)$$

式中: L_s 表示每一阶段预测显著性图的混合损失值; L_f 表示最终融合阶段的混合损失值。

2 数据集及评价指标

2.1 数据集

常用的显著性目标检测数据集包括: DUTS^[13]、HKU-IS^[14]、ECSSD^[15]、PASCAL-S^[16]。利用 DUTS 数据集集中的训练集 DUTS-TR 训练 EGMFNet, DUTS-TR 数据集包含 10553 张图像是最大且最常用的显著性目标检测任务训练数据集。同时,文中添加包括含有人物、透明物体、背景复杂、背景阴暗的图像以扩充数据集,扩充后的数据集包含 11552 张图像。此外,对图像进行了水平翻转以增强数据。

利用 DUTS-TE、HKU-IS、ECSSD、PASCAL-S 作为测试集对训练的网络进行测试。

2.2 评价指标

文中使用 F -measure (F_β) 分数、平均绝对值误差 (MAE) 和结构相似性指标 (S -measure) 作为网络性能的评价指标,以衡量网络的性能。

F -measure 是具有非负权重 β^2 的精确率和召回率的加权调和平均值,通常用于综合准确率和召回率, F -measure 可表示为:

$$F_\beta = \frac{(1 + \beta^2) \times Precision \times Recall}{\beta^2 \times Precision + Recall} \quad (14)$$

式中: β^2 设置为 0.3。

平均绝对值误差表示预测显著性图与显著性真值间的平均像素差异,定义如下:

$$MAE = \frac{1}{H \times W} \sum_{x=1}^H \sum_{y=1}^W \|S(x,y) - G(x,y)\| \quad (15)$$

式中: $S(x,y)$ 和 $G(x,y)$ 表示预测显著性图和显著性真值; (H,W) 表示图像尺度即长和宽; (x,y) 表示像素坐标。

结构相似性指标通常被用于评估预测显著性图同显著性真值之间的结构相似性,定义是区域结构相

似性和对象结构相似性的加权求和:

$$S_\alpha = (1 - \alpha)S_r + \alpha S_o \quad (16)$$

式中: S_r 表示区域结构相似性; S_o 表示对象结构相似性; 权值 α 设置为 0.5。

3 实验

3.1 实验细节

EGMFNet 基于 Pytorch 1.10.2 框架搭建,并在一台具有 AMD Ryzen 9 5950X 3.4 GHz CPU (64 GB RAM) 和 RTX 3090 GPU (24 GB 内存) 的 16 核 32 线程的台式机上进行所有的训练和测试过程。在训练的初始阶段,对数据集图像进行了拉伸、翻转和裁剪的预处理操作,批尺寸被设置为 16 并使用 Adam 优化器进行训练,Adam 优化器的超参数设置为: 初始学习率 (learning rate) $l=1 \text{ e-}3$ 、指数衰减率 $\beta=(0.9, 0.999)$ 、 $\epsilon=1 \text{ e-}8$ 、权重衰减=0。经过 48 万次迭代(耗时约 150 h),训练损失有效地收敛。

3.2 对比实验及性能测试

为了对 EGMFNet 的性能进行评估,文中在 DUTS-TE、HKU-IS、ECSSD、PASCAL-S 四个常用的数据集上测试了 EGMFNet 的性能,并选取 MDF、PiCANet、AFNet、R³Net、PoolNet、BANet、BPFINet 等典型网络进行实验对比。

实验结果如表 1 所示,网络在 ECSSD 数据集上取得了最好的结果,相较于第二名性能提升了 1.5%,在 HKU-IS、PASCAL-S 和 DUTS-TE 数据集上文中的网络表现良好, F 值分别提升了 2.7%、1.8% 和 2.6%,证明文中的网络具有很好的鲁棒性。此外,选取了性能较好的 R³Net、PoolNet、BPFINet 对参数量及显著性预测实时性进行了评估,实验选用 20 张 400 pixel×267 pixel 的图像进行测试,结果如表 2 所示。

人眼可识别的帧率阈值为 25 FPS,测试结果表明 EGMFNet 在保证预测显著性图质量的前提下以较少的参数保证了 32 FPS 的预测速度。同时相较于性能较为优秀的 BPFINet 和 PoolNet,EGMFNet 预测速度提升约 6.7%,仅次于 R³Net,表明 EGMFNet 具有较为良好的实时性。

为了更直观地展现 EGMFNet 的性能,展示了 EGMFNet 预测显著性注释图,如图 6 所示。文中所提出网络在复杂环境下具有较好的性能。特别是

表 1 对比实验结果

Tab.1 Comparison of the experimental results

	ECSSD			PASCAL-S			HKU-IS			DUTS-TE		
	F_β	MAE	S_α	F_β	MAE	S_α	F_β	MAE	S_α	F_β	MAE	S_α
MDF	0.832	0.105	0.776	0.768	0.146	0.692	0.861	0.129	0.810	0.730	0.094	0.792
PiCaNet	0.886	0.045	0.917	0.856	0.078	0.848	0.870	0.043	0.904	0.759	0.051	0.869
AFNet	0.908	0.042	0.913	0.821	0.070	0.844	0.888	0.036	0.905	0.792	0.046	0.867
R ³ Net	0.914	0.040	0.910	0.845	0.094	0.800	0.893	0.036	0.895	0.785	0.057	0.834
PoolNet	0.915	0.039	0.921	0.822	0.074	0.845	0.892	0.034	0.911	0.809	0.040	0.883
BPFINet	0.928	0.034	0.926	0.845	0.065	0.857	0.911	0.028	0.918	0.838	0.038	0.882
Proposed	0.943	0.033	0.926	0.868	0.069	0.856	0.928	0.033	0.912	0.852	0.037	0.883

表 2 EGMFNet 参数量及实时性评估

Tab.2 Parameter quantity and real-time evaluation of EGMFNet

	Parameters	Runtime/s	Frame rate/FPS
R ³ Net	56 156 126	0.030	33
PoolNet	71 383 577	0.033	30
BPFINet	68 326 853	0.033	30
Proposed	60 638 928	0.031	32

图 6 第二行及第三行中显著性目标物体颜色与背景颜色一致且具有灰尘及云等干扰物,EGMFNet 仍能十分准确的预测显著性目标并保留了非常清晰的边缘。同时,网络在面对阴影和强光时表现欠佳,如图 6 第一行与第五行所示,树林中的阴影和火焰的强光影响了显著性目标边缘的提取。此外,EGMFNet 在面对小目标和多目标时仍能进行有效处理,如图 6 第六行所示,EGMFNet 能有效的提取出图像中较为明显的载具,但错误的提取了比较靠前的几名士兵。

此外,展示了 ECSSD 数据集上的测试结果,如图 7 所示。面对较为突出的显著性目标,如图 7 第一行所示,EGMFNet 表现出了良好的性能,在保证显著性目标具有完整结构的同时保留了清晰的边缘。但,面对具有争议的显著性目标时(如图 7 第二行),EGMFNet 突出了较为明显的成年鸟类,但并未完全消除幼鸟的干扰并提取了幼鸟的虚影。面对透明物体时(图 7 第三行),EGMFNet 较为有效地提取了完整的蜻蜓翅膀,但翅膀的部分位置出现了虚影及结构不完整的情况。在面对干扰物较多、光线阴暗以及显著性目标不明显的情况时(图 7 第四行),EGMFNet 仍能有

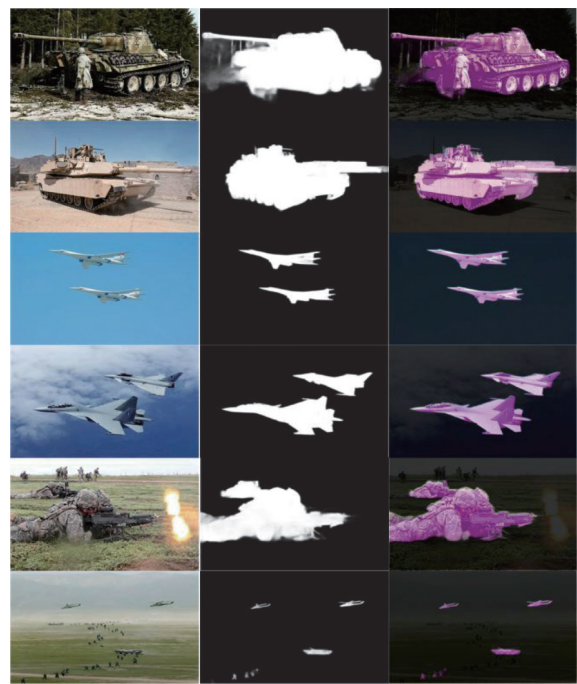


图 6 EGMFNet 预测显著性注释效果图

Fig.6 EGMFNet prediction annotation rendering

效定位显著性目标并保留其完整的结构以及清晰的边缘。

总体上,EGMFNet 能够有效应对复杂场景与小目标场景下的显著性目标提取,但 EGMFNet 仍存在着一些不足。例如,在面对争议性的显著性目标时 EGMFNet 并不能完全消除干扰目标,同时面对透明物体时 EGMFNet 的性能仍有待提升。此外,在面对强光及阴影场景时,EGMFNet 的性能并不理想,可能出现显著性目标结构不完整的情况。

3.3 消融实验

为验证 EGMFNet 中基本网络、EGSAM 模块及

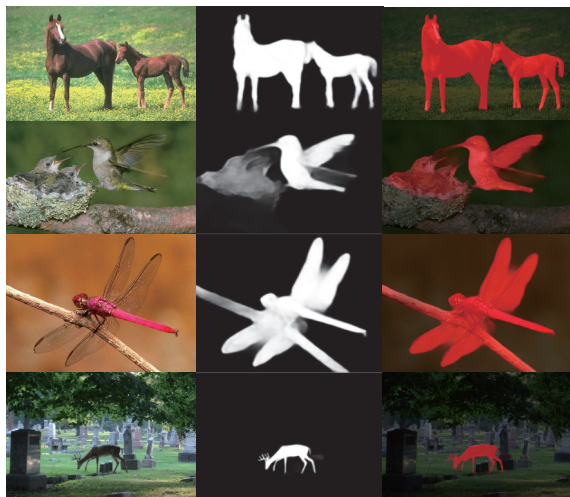


图 7 EGMFNet 预测显著性注释效果图

Fig.7 EGMFNet prediction annotation rendering

边界损失的有效,在保证除模块及损失函数变动外其余参数不变的前提下进行了消融实验并在 ECSSD 数据集上评估了各方案网络的性能。实验评估结果如表 3 所示。

对比实验数据可知,在基础网络中引入 EGSAM 模块以及使用混合损失函数进行训练在一定程度上

提升了网络的性能。对比第一组 (GA) 与第三组 (GC) 可以看出,引入了 EGSAM 的完整 EGMFNet 性能相较于基础网络有了十分明显的提升, F 值提升 1.4%, 结构相似性指标提升 S 值 0.8%, 这表明引入三阶段 EGSAM 模块有助于提升预测显著性图结构的完整性并有效保留更多的空间细节。同时,为验证混合损失函数的有效性,将四组实验分成两组对比实验:第一组 (GA) 与第二组 (GB); 第三组 (GC) 与第四组 (GD)。由对比实验数据可以看出,利用混合损失函数训练的网络性能相较于使用 BCE Loss 训练的网络性能有了一定的提升,第一组 (GA) 与第二组 (GB) 相比 F 值提升 0.3%, 结构相似性指标 S 值提升 0.3%。特别是对于具有更多空间细节的网络性能提升更加明显,第三组 (GC) 与第四组 (GD) 相比 F 值提升 0.7%, 结构相似性指标 S 值提升 1.6%。第三组 (GC) 与第四组 (GD) 的对比实验结果表明在 BCE Loss 的基础上引入边界损失函数能够帮助网络保留清晰的显著性目标边界,这种效果在网络所提取的特征图具有丰富的空间特征时更加有效。

表 3 消融实验结果。其中 GA 为使用 BCE Loss 训练的基础网络, GB 为使用混合损失训练的基础网络, GC 为引入 3 阶段 EGSAM 模块的完整网络并使用 BCE Loss 进行训练, GD 使用混合损失训练 GC 中的网络

Tab.3 Ablation experimental results. GA is the basic network trained with BCE Loss, GB is the basic network trained with mixed loss, GC is the complete network with three-stage EGSAM module and trained with BCE Loss, and GD is the complete network with three-stage EGSAM module and trained with mixed loss

Groups	Structure	Loss	F_{β}	MAE	S_a
GA	Baseline	BCE	0.923	0.041	0.908
GB	Baseline	BCE+BL	0.926	0.040	0.911
GC	Baseline+EGSAM(3 stages)	BCE	0.936	0.036	0.919
GD	Baseline+EGSAM(3 stages)	BCE+BL	0.943	0.033	0.926

3.4 EGSAM 模块层级验证实验

如 1.1 节所示,EGMFNet 主体结构由 RCFBlock 堆叠而成的 MCFUBlock 组成,此节将对 RCFBlock 堆叠数量对网络性能的影响进行实验测试。实验分为三组进行:第一组使用 MCFU4、MCFU3、MCFU2 替换原网络中的 MCFU3、MCFU2、MCFU1;第二组为对照组,即为 EGMFNet;第三组使用 MCFU2、MCFU1、RAFBLOCK 进行替换。实验在 ECSSD 数据集上对三

组实验中的网络进行性能测试。

实验结果如表 4 所示,第一组实验表明过多堆叠 RCFBlock 不仅增加了参数量,同时大量的池化操作导致空间信息丢失严重影响了网络性能。第三组实验相较于第一组实验,网络性能有了明显提升,但少量堆叠 RCFBlock 使得网络较浅无法有效获取空间信息及语义信息,这使得网络性能无法达到最好的效果。因此,采用 MCFU3、MCFU2、MCFU1 的组合形

表 4 RCFBlock 堆叠数量验证实验

Tab.4 RCFBlock stack quantity verification experiment

No.	F_β	MAE	S_α	Parameters
1	0.889	0.089	0.831	64 609 584
2	0.943	0.033	0.926	60 638 928
3	0.925	0.040	0.911	48 277 712

式搭建 EGMFNet。

3.5 EGSAM 模块层级验证实验

EGMFNet 网络中仅前三层添加了 EGSAM 模块, 为验证 EGSAM 添加层级实验, 从上至下依次增加网络添加层级, 并在 ECSSD 数据集上进行了测试, 通过

显著性目标检测性能指标评估及参数量、权重文件大小等方面进行了比较。

实验结果如表 5 所示, 添加 EGSAM 模块有效提升了网络的性能。随着添加 EGSAM 模块的层级逐渐增多直至添加至第三阶段, 网络性能有较大幅度的提升, F 值提升 2.1%, 随之而来的参数量和模型大小也在可控范围内逐渐增大。然而, 在第四阶段添加 EGSAM 模块使得参数量和模型大小急剧增加, 这样的变化在增加 GPU 负担的同时对网络性能的提升并不明显甚至存在一些负面影响, 因此仅在 EGMFNet 的前三个阶段桥接 EGSAM 模块以提升网络性能。

表 5 EGSAM 模块层级验证实验结果

Tab.5 Verify the experimental results at EGSAM module level

No.	Stage with EGSAM	F_β	MAE	S_α	Parameters	Size/MB
1	Baseline	0.923	0.041	0.908	54 437 157	207.67
2	Stage 1	0.936	0.038	0.916	54 734 330	208.79
3	Stage 1+2	0.941	0.035	0.923	55 916 197	213.30
4	Stage 1+2+3	0.943	0.033	0.926	60 638 928	231.32
5	Stage 1+2+3+4	0.942	0.033	0.924	79 521 275	303.35

3.6 EGSAM 融合系数设置验证实验

如 3.3 节中所示, EGSAM 模块的输出结果分别经过两次调用, 其中包含一次与多层级融合特征的加权融合。为研究融合系数 α 对网络性能的影响, 设置了验证实验, 并在 ECSSD 数据集上进行评估。

实验结果如表 6 中所示, 当 α 数值较大时, 过多冗余边缘细节影响了网络性能。随着 α 逐渐变小 ($\alpha < 0.1$) 时, 添加 EGSAM 模块的网络性能相较于基础网络有所提升, 当 $\alpha = 0.01$ 时取得最好效果, F 值较基

础网络提升 2.1%。当 α 继续减小 ($\alpha < 0.01$), EGSAM 模块对于网络的影响弱化, 使得网络性能降低, 因此将 EGSAM 融合系数设置为 $\alpha = 0.01$ 以取得最好的效果。

4 结 论

文中提出了新型的 EGMFNet 用于显著性目标检测任务, EGMFNet 的主干网络是一个具有三级嵌套的 U 型网络结构, 网络的较低层级使用由不同数量 RCFBlock 填充的 MCFUBlock 构建。为减少高级语义信息因和全局空间特征因池化操作而丢失或被稀释, 在 EGMFNet 的较高层级使用带有残差连接的 U 型块进行填充。同时, 在网络的较低层级引入边缘信息引导的扩张空间注意力模块 (EGSAM) 进一步提取和利用边缘及空间特征信息。此外, 使用融合边界损失函数的混合损失函数端到端的训练 EGMFNet, 在四个常用数据集上的测试结果表明, EGMFNet 较典型方法 F 值提升了 1.5%、2.7%、1.8% 和 1.6%, 具有良好的性能。

表 6 EGSAM 融合系数设置验证实验

Tab.6 EGSAM fusion coefficient setting experiment

No.	α	F_β	MAE	S_α
1	Baseline	0.923	0.041	0.908
2	1	0.842	0.104	0.832
3	0.1	0.911	0.052	0.894
4	0.05	0.926	0.041	0.913
5	0.01	0.943	0.033	0.926
6	0.005	0.940	0.037	0.920

参考文献:

- [1] Sun Zhaolei, Hui Bin, Qin Mofan, et al. Object detection method based on saliency measure for infrared radiation image [J]. *Infrared and Laser Engineering*, 2015, 44(9): 2633-2637. (in Chinese)
- [2] Huang Mengke, Liu Zhi, Ye Linwei, et al. Saliency detection via multi-level integration and multi-scale fusion neural networks [J]. *Neurocomputing*, 2019, 364(9): 310-321.
- [3] Li Tengpeng, Song Huihui, Zhang Kaihua, et al. Recurrent reverse attention guided residual learning for saliency object detection [J]. *Neurocomputing*, 2020, 389(3): 170-178.
- [4] Jia Fengwei, Wang Xuan, Guan Jian, et al. Bi-connect net for salient object detection [J]. *Neurocomputing*, 2020, 384(1): 142-155.
- [5] Ullah Inam, Jian Muwei, Hussain Sumaira, et al. Global context-aware multi-scale features aggregative network for salient object detection [J]. *Neurocomputing*, 2021, 455(1): 139-153.
- [6] Jiang Guoqing, Wan Lanjun. Detection of dim and small infrared targets based on the most appropriate contrast saliency analysis [J]. *Infrared and Laser Engineering*, 2021, 50(4): 20200377. (in Chinese)
- [7] Li Ning, Huang Jincai, Feng Yanghe. Construction of multi-channel fusion salient object detection network based on gating mechanism and pooling network [J]. *Multimedia Tools and Applications*, 2021, 81: 12111-12126.
- [8] Li Zun, Lang Congyan, Liew Junhao, et al. Cross-layer feature pyramid network for salient object detection [J]. *IEEE Transactions on Image Processing*, 2021, 30: 4587-4598.
- [9] Chen Tianyou, Hu Xiaoguang, Xiao Jin, et al. BPFINet: Boundary-aware progressive feature integration network for salient object detection [J]. *Neurocomputing*, 2021, 451(8): 152-166.
- [10] Yao Zhaojian, Wang Luping. ERBANet: Enhancing region and boundary awareness for salient object detection [J]. *Neurocomputing*, 2021, 448: 152-167.
- [11] Cui Wenzhao, Zhang Qing, Zuo Baochuan. Deep saliency detection via spatial-wise dilated convolutional attention [J]. *Neurocomputing*, 2021, 445(6): 35-49.
- [12] Kervadec Hoel, Bouchtiba Jihene, Desrosiers Christian, et al. Boundary loss for highly unbalanced segmentation [J]. *Medical Image Analysis*, 2021, 67: 101851.
- [13] Wang Lijun, Lu Huchuan, Wang Yifan, et al. Learning to detect salient objects with image-level supervision [C]//IEEE Conference on Computer Vision & Pattern Recognition. IEEE, 2017: 21-26.
- [14] Li Guanbin, Yu Yizhou. Visual saliency detection based on multiscale deep CNN features [J]. *IEEE Transactions on Image Processing*, 2016, 25(11): 5012-5024.
- [15] Yan Qiong, Li Xu, Shi Jianping, et al. Hierarchical saliency detection [C]//IEEE Conference on Computer Vision and Pattern Recognition, 2013: 23-28.
- [16] Yin Li, Hou Xiaodi, Koch Christof, et al. The secrets of salient object segmentation [C]// IEEE Conference on Computer Vision & Pattern Recognition, 2014: 23-28.