

基于深度纹理特征的伪装目标边缘细化检测

袁昊, 葛海波, 辛世澳, 胥冬梅, 杨雨迪

(西安邮电大学电子工程学院, 陕西 西安 710121)

摘要: 为解决传统伪装目标检测(COD)出现的空间信息不完整和目标边界模糊的问题, 提出一种基于深度纹理特征的伪装目标边缘细化检测算法。该算法针对目标的纹理差异和边缘细节设计上下文纹理差异放大模块(CTDAM)、特征边界搜寻模块(FBSM)和边界推理模块(BIM)。CTDAM 利用全局感受野覆盖和并行多分支混合卷积方式突出被遮挡的伪装目标的纹理差异; 在注意力特征融合模块(AFFM)中引入局部注意力和位置通道感知并行注意力指导特征跨层融合, 达到平衡局部信息和增强全局上下文语义信息的效果; FBSM 利用自注意力机制将低层与高层特征相结合, 处理不同边界像素点之间的依赖关系; BIM 利用 FBSM 所提供的边界指导因子, 指导融合后的特征推断出真实目标并细化边缘细节。在 CAMO、CHAMELEON 和 COD 10K 数据集上利用 4 个客观评估指标进行定量和定性实验, 结果表明, 该算法的检测性能优于对比的 8 种先进算法, 在 COD 10K 数据集上, 其平均绝对误差(MAE)达到了 0.034。

关键词: 伪装目标检测; 特征边界搜寻; 注意力特征融合; 上下文信息; 纹理差异

源代码链接: <https://github.com/erfree/BTNet>

中图分类号: TP391

文献标志码: A

DOI: 10.19678/j.issn.1000-3428.0068431

Edge Refinement Detection of Camouflage Targets Based on Deep Texture Features

YUAN Hao, GE Haibo, XIN Shiao, XU Dongmei, YANG Yudi

(School of Electronic Engineering, Xi'an University of Posts and Telecommunications, Xi'an 710121, Shaanxi, China)

【Abstract】 To solve the problem of incomplete spatial information and fuzzy target boundaries in conventional Camouflage Object Detection (COD), a COD algorithm based on depth texture features and edge thinning is proposed. Based on the texture difference and edge details of the target, the algorithm designs Context Texture Difference Amplification Module (CTDAM), Feature Boundary Search Module (FBSM), and Boundary Inference Module (BIM). CTDAM uses global receptive field coverage and parallel multi-branch hybrid convolution to highlight the texture differences of occluded camouflage targets. Additionally, it introduces local attention and position channel perception to guide feature cross-layer fusion in Attention Feature Fusion Module (AFFM). Therefore, it achieves the effect of balancing local information and enhancing semantic information of global context. FBSM uses a self-attention mechanism to combine low- and high-level features to deal with the dependence between different boundary pixels; and BIM uses the boundary guidance factor provided by FBSM to guide the fusion features to infer the real target and refine the edge details. Quantitative and qualitative experiments are conducted on the CAMO, CHAMELEON, and COD 10K datasets using four objective evaluation indices. The results demonstrate that the proposed algorithm is superior to other eight algorithms. On the COD 10K dataset, the Mean Absolute Error (MAE) is 0.034.

【Key words】 Camouflage Object Detection (COD); feature boundary search; attention feature fusion; context information; texture difference

0 引言

伪装就是为了掩盖真实身份而进行的装扮。在数个世纪的演变后, 许多物种为了提高种族延续, 开始在自身的生理特征和身体结构上进行伪装, 当发现天敌后, 生物(植物或动物)利用自身的生理结构特征(颜色、形状、气味、激素等)来适应周围环境, 从

而避免天敌的追捕。

早期的伪装目标检测(COD)方法利用低级特征(颜色、纹理、频域、运动等)来区分目标和背景, 然而, 它们在复杂场景中提取高层次语义信息的能力有限, 这使得伪装目标在定位、分割和边缘细化方面受到严重制约。最近, 随着高精度、大规模数据集(COD 10K、CAMO 等)和卷积神经网络(CNN)的

收稿日期: 2023-09-20 修回日期: 2023-12-01

基金项目: 陕西省自然科学基金(2011JM8038); 陕西省重点产业创新链(群)项目(S2019-YF-ZDCXL-0098)。

通信作者 E-mail: 2023049550@qq.com

出现,使得 COD 得到了极大的发展。因为 CNN 在挖掘低级细节外观线索和高级语义知识方面的出色表现,使得大量研究者在此方面进行探索。目前,现有研究主要针对两点问题:在保护细小特征结构信息的情况下,如何提取多尺度特征并扩大感受野;如何使用边界信息作为辅助引导线索来促进学习过程。

针对上述第一个问题:文献[1]设计了 2 个基于池化的模块,即特征聚合模块(FAM)和全局制导模块(GGM),这 2 个基于池化的模块可以逐步细化高级语义特征,从而生成细节丰富的显著性图,进而更准确地定位具有清晰细节的突出目标;文献[2]设计了一个关系交叉注意力模块(RCAM)来实现跨嵌入子空间的双向消息传播,用于从外观和运动分支中提取判别特征,这确保了彼此之间的约束。上述方法主要通过使用具有不同膨胀率的多个卷积层或具有不同内核大小和步幅的池化层来扩大感受野并提取多尺度信息,这些方法已被证明在利用多尺度上下文感知信息方面是有效的,这有助于定位目标。针对上述第二个问题:目前最先进的 COD 方法主要从目标整体方面进行研究,却忽视了被环境所掩盖的边缘;文献[3]提出了一种端到端的边缘引导检测网络模型 EGNNet,实现了互补的显著边缘信息和显著对象信息的建模和融合;文献[4]提出了一种由新的残差优化模块和类似于 U-Net 的深度监督的 encoder-decoder 网络构成的边界感知网络(BASNet)结构,可以有效地提取图像中的层级特征,捕捉伪装目标检测任务中的全局(粗糙)和局部(细致)信息;文献[5]开发了一个边界引导模块(BGM)来充分挖掘边界信息,BGM 引导解码器通过显式地增强所提取的边界特征来学习边界上下文,这种利用底层特征来挖掘边缘信息的方法,可以作为辅助引导线索来促进伪装目标的边缘细化过程。

现有的算法虽然在伪装目标检测领域已经取得了一定的进展,但是由于伪装对象的纹理相似性,仍存在以下两点问题:虽然不同膨胀率的卷积层或不同内核大小的池化层有助于扩大感受野并挖掘丰富的上下文信息,但由于这些膨胀的卷积层中存在大“孔”,可能会导致一些细节信息在此过程中丢失;虽然利用底层特征可以获取大量的边缘信息,但模型整体的运算速度会大幅下降。

针对上述问题,本文提出基于深度纹理特征的伪装目标边缘细化检测算法。本文的主要贡献有 5 个方面:

1)提出一种新的 BTNet 用于 COD 任务,它可以有效集成不同层次的特征并将有价值的边界信息和上下文信息从编码器馈送到解码器,使得伪装检测效果更加精准。

2)设计特征边界搜寻模块(FBSM),利用自注意力机制将低层特征与高层特征相结合,提取出边界指导因子,再通过逐层的方式将其纳入解码器网络,以帮助模型检测伪装对象的边界。

3)为了进一步放大对象与背景之间的纹理差异和完善伪装边界不完整的情况,设计一个上下文纹理差异放大模块(CTDAM),采用感受野覆盖整个输入特征图的方式实现对视觉场景的全局理解,再引入并行多分支混合卷积模式来增强感受野,进一步放大目标与背景的纹理差异。

4)为了解决特征融合后特征多样性缺失的问题,提出一个注意力特征融合模块(AFFM),利用局部注意力模块(LA)增加特征的局部语义信息,再利用位置通道感知并行注意力模块(PPA)平衡局部信息和增强全局上下文的语义信息,确保融合后的特征多样性。

5)利用空间注意力机制设计边缘推理模块(BIM),用以注入边界相关的边缘线索的表示学习,以增强对象结构语义的特征表示,从而使得解码器能够从编码器中获取更多有效的特征,提高 COD 性能。

1 相关工作

1.1 伪装目标检测

随着深度学习的兴起,近年来的伪装目标检测技术大多采用基于深度神经网络的方法,在大数据集(CAMO、COD 10K)上进行训练,并取得了显著的进展。文献[6]设计了一种新的视频伪装目标检测(VCOD)框架,可以利用短期动态和长期时间一致性从视频帧中检测伪装物体。文献[7]提出 MirrorNet,通过引入分割翻转图像的仿生攻击流来增强分割原始图像的主流,并将仿生攻击流的预测翻转后,和主流预测结果进行融合得出最终的伪装目标预测结果。文献[8]对完全标注隐藏物体所导致的不确定性,给出了一个 JCSOD 模式,通过全卷积计算判别器来估算预测结果的置信率,同时也通过对抗训练等策略对置信率的估计进行显式建模。文献[9]通过引入边缘提取作为一项辅助任务,提出 MGL(Mutual Graph Learning)模型来捕获区域和边缘,该模型同时进行伪装物体的分割和边界检测,可以相互指导。文献[10]提供一种混合尺度的三重

网络 ZoomNet, 该模型使用缩放策略, 并使用专门设计的比例集成单元和分层混合尺度单元来学习判别性的混合尺度语义, 以充分发现在候选对象和背景环境间不易发现的线索。文献[11]通过数据增强和改进的 YOLOv5s 算法, 在伪装目标检测中取得了较高的精度和鲁棒性。文献[12]提出了级联 decamouflage 模块, 其由区域增强块和反向注意力挖掘块组成, 以准确地检测伪装对象, 并获得了完整的目标对象。

1.2 特征融合

特征融合是指将主干网络提取出来的特征图进行聚合, 从而达到丰富语义信息的效果。特征融合主要分为两类:

1) 早融合 (Early Fusion), 首先融合不同图层的特征, 然后在融合后的特征上训练预测器, 如文献[13]提出张量融合网络 TFN, 能够端到端地学习模态内和模态间的动态, 采用一种新的多模态融合方法 (张量融合) 对模态间动态进行建模, 模态内动态则通过 3 个模态嵌入子网络进行建模。

2) 晚融合 (Late Fusion), 如特征金字塔网络 (FPN)^[13], 通过在不同尺度上使用多个不同大小的滤波器来检测图像中的特征。文献[14]提出一个分层的注意力特征融合模块 (RAFA), 可以更好地联合相邻尺度之间的特征图^[15]。文献[16]提出了一种用于显著对象检测的通用聚合多级卷积特征框架, 其通过结合粗略语义和精细细节将多级特征图集成到多个分辨率中。

1.3 边界细节增强

在伪装目标检测领域, 现有的大部分技术想要检测出伪装在背景环境中的对象不难, 往往是在处理伪装对象的边界上效果不佳, 经常出现边界模糊不清甚至包含一些背景噪声的情况。目前, 文献[17]提出了一种边界引导模块 BGM, 可以通过增强提取的边界特征来引导解码器学习边界上下文。文献[18]提出了一种边界引导伪标记 (BGPL) 方法, 通过给未标注的数据生成可靠的伪标签来提高检测性能。文献[19]提出了一种边界引导特征交互模块 (BFIM), 融合多层次特征并探究伪装目标与其轮廓之间相辅相成的关系。文献[20]提出一种边缘引导网络, 该网络利用全局边界特征来指导不同层次的区域特征, 以解决边界模糊的问题^[21]。在本文中, 利用低层特征和高层特征的优势设计一个简单但有效的模块 FBSM, 其中高层次特征用于伪装对象的定位, 低层次特征用于提供伪装对象的细节, 使得预测出的边界细化更加精准。

2 算法研究

2.1 算法模型

本文所提出的方法整体架构如图 1 所示 (彩色效果见《计算机工程》官网 HTML 版, 下同), 包含特征边界搜寻模块 FBSM、上下文纹理差异放大模块 CTDAM、注意力特征融合模块 AFFM 和边界推理模块 BIM。首先, 对于输入图像, 本文采用 Res2Net-50 在 5 个不同层提取其特征, 表示为 F_i , $\{i=1, 2, 3, 4, 5\}$; 然后, 利用 CTDAM 模块先增强当前位置特征实现对视觉场景的全局理解, 再利用纹理差异化突出图像中的细节矛盾信息, 输出特征 F_{CTDAM}^i , $\{i=3, 4, 5\}$; 之后, 将特征 F_{CTDAM}^i 输入到 AFFM 模块中进行融合, AFFM 模块利用空间注意力机制确保特征之间的互补互动和整合, 输出融合后的特征 F_{AFFM} ; 最后, FBSM 利用自注意力机制将低层特征与高层特征相结合, 以提供边界指导因子^[22], 将边界指导因子与融合后的特征 F_{AFFM} 注入到 BIM 中, 以增强与对象结构语义的特征表示, 进而推理出伪装目标的精准边界, 同时获得预测图。

2.2 特征边界搜寻模块

当伪装对象存在遮挡、相似干扰时, 往往会对边界的精准检测带来很大的挑战^[23]。现有 COD 算法主要利用低层特征中丰富的边界信息作为辅助引导线索来促进学习过程, 文献[24]利用低层特征中丰富的边缘线索对融合后的高层特征进行表征增强, 但这会导致整个预测的速度过于缓慢, 且检测容易出现偏差。为此, 本文提出了特征边界搜寻模块 FBSM, 如图 1 所示, 采用在传统边界引导模块上引入具有定位信息的高层特征的方式, 来处理多目标伪装时不同边界像素点之间的依赖关系, 这不仅可以有效定位到伪装对象所在的区域, 还省去了较为繁琐的定位功能, 实现了轻量化的定位功能。之后, 为了令边缘信息提取更加细致, 采用自注意力机制捕捉来自不同通道的边界信息, 使得目标边界细化的精准度和定位速度得到提高。根据效率分析实验可知, 该算法在具备定位功能后, 运算效率得到大幅提高, 算法运算速度达到 57.57 帧/s。

在 FBSM 中, 其中 2 个输入 F_1 和 F_2 来自低层特征, 另一个输入 F_3 来自高层特征, 高层次特征用于伪装对象的定位, 低层次特征用于提供伪装对象的细节。为了细化伪装边界细节, 本文引入了自注意力机制, 该机制可以根据输入的不同尺度的特征捕捉丰富的上下文边界信息。然后, 通过从卷积特

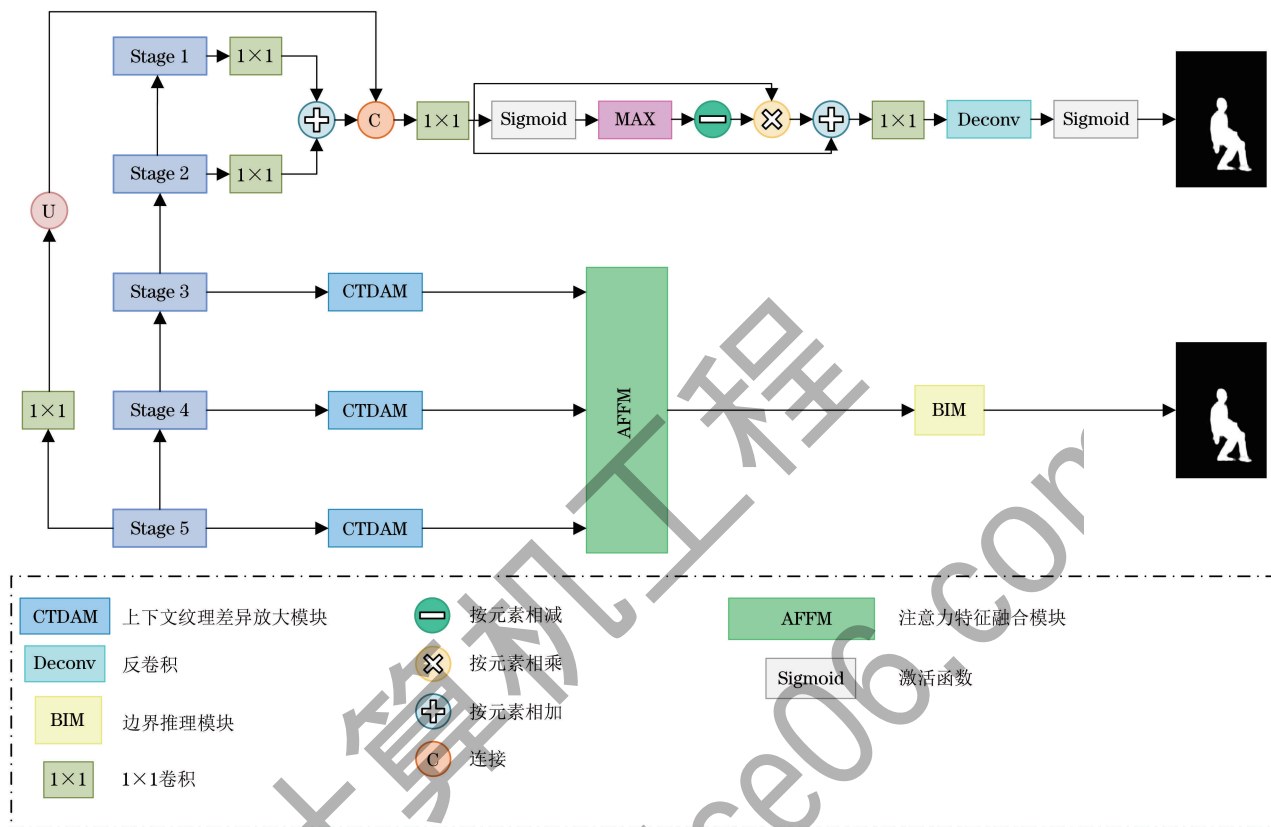


Fig.1 The framework of BTNet

征关联起来并进行特征增强。该模型首先采用感受野覆盖整个输入特征图的方式,通过从其他地图位置上收集信息来增强当前位置的特征,从而实现对视觉场景的全局理解,可以有效弥补固定大小卷积层仅关注局部信息和卷积层堆叠导致特征相对单一的问题,从而增强其全局上下文信息建模和特征多样性。之后,设计并行多分支混合卷积模式,为了扩大感知范围并提取多尺度信息,该部分采用多个具有不同膨胀率和步幅的池化层来增加网络的感受野和提取多尺度信息的能力。考虑到纹理差异在伪装目标检测过程中有着非常重要的作用,在并行多分支混合卷积模式中引入最大池化操作来模拟眼球中的感知细胞,从而突出纹理差异,并随着感受野的增强,纹理差异将进一步放大,从而可以更好地检测出伪装对象的边缘细节。CTDAM 具体结构如图 2 所示,Res2Net-50 提取出的特征 $F_i, \{i = 3, 4, 5\}$ 被馈送到 1×1 卷积后,利用 Softmax 函数来获得关注权重,以获得全局上下文特征 F_c ,其计算公式如下:

为了在保护细小特征结构信息的情况下提取多尺度特征并扩大感受野,设计CTDAM,其可以在扩大感受野和建立长程依赖性的同时将层级之间的特

然后利用瓶颈变换得到特征,将瓶颈的隐藏表示维数降低到 $Ck/4$,最后采用逐元素相加的方式来增强特征 F_l ,这可以很好地弥补因卷积过度而导

致的特征多样性不足。 F_t 计算公式如下:

$$F_t = \text{Conv}_{1 \times 1}(\text{Conv}_{1 \times 1}(F_c)) + F_i \quad (6)$$

之后,CTDAM 采用并行多分支混合卷积模式,首先利用 1×1 卷积将特征 F_t 的通道数减少到 32,并获得特征 F_r 。每个 CTDAM 模块都包含 4 个并行分支:在第一个分支上加入最大池化操作,可以从局部角度检测目标的差异细节,为了进一步体现纹理差异,采用 2 组 3×3 卷积的纹理特征来学习 2 个参数图 $\gamma \in C \times H \times W$ 和 $\beta \in C \times H \times W$;在第二个分支上采用一个 3×3 CBR 操作与参数图 γ 相乘,然后再进行一次 CBR 操作来提取特征,获得具有边缘信息的 F_{e_2} ;在第三个分支上采用 2 次 3×3 CBR 操作再与参数图 β 相乘,获

得特征 F_{r_3} ,为了获得多维信息特征,将特征 F_{r_3} 与上一层分支的特征 F_{e_2} 进行融合,获得特征 F_X ,这样进一步放大了纹理差异,使得检测结果更加精确;为了使并行多分支混合卷积模式效果更好,在第四个分支上保留原始信息,将原始信息与特征 F_X 进行相加运算,获得最终的 CTDAM 输出特征 F_{CTDAM}^i 。具体地,上下文纹理差异放大过程可以描述如下:

$$F_{e_2} = \text{CBR}(\text{CBR}(F_t) \times \text{Max}(F_t)) \quad (7)$$

$$F_{r_3} = \text{CBR}(\text{CBR}(F_t) \times \text{Max}(F_t)) \quad (8)$$

$$F_{\text{CTDAM}}^i = C(F_{e_2}, F_{r_3}) + F_t \quad (9)$$

式中:CBR 表示基本卷积块(具有批量归一化和 ReLU 的卷积操作)。

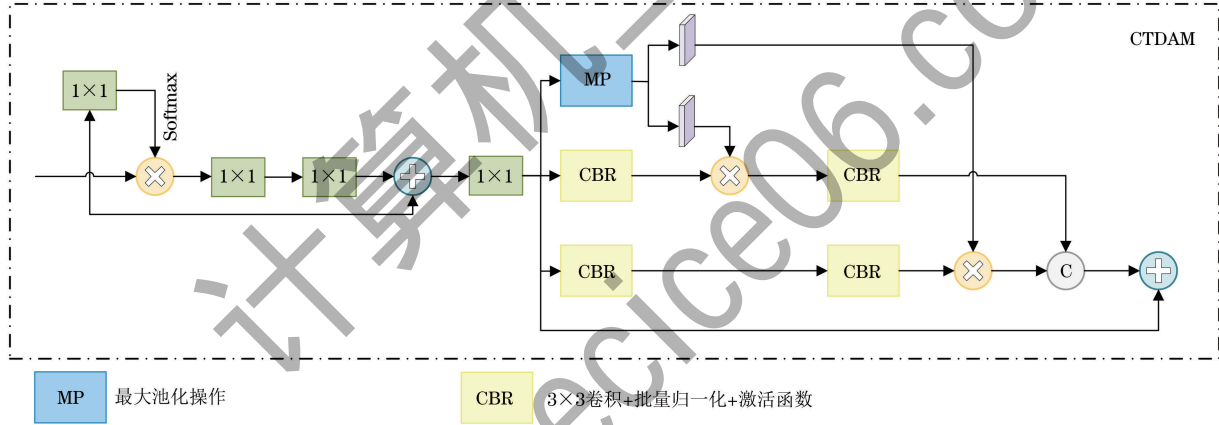


图 2 CTDAM 模块结构

Fig. 2 CTDAM module structure

2.4 注意力特征融合模块

在目标识别领域,特征之间的交互对于增强特征表示和减少特征差异有着至关重要的作用。为了更好地实现特征间的交互,本文设计了一个利用空间注意力机制的特征融合模块 AFFM,实现高层特征和低层特征的融合,确保特征之间的互补互动和整合。在 AFFM 中,提出了 2 种注意力来确保特征融合的可靠性:一个是局部注意力 LA,用于增加特征的局部语义信息;另一个是利用空间注意力和通道注意力构造的一个位置通道感知并行注意力 PPA,主要用于平衡局部信息和增强全局上下文的语义信息。

AFFM 的具体结构如图 3 所示。首先,将 CTDAM 输出的 F_{CTDAM}^i , $\{i = 3, 4, 5\}$ 同时馈送到 3×3 卷积中,以确保通道数相同(通道数设置为 64);然后,将相邻 2 层的特征进行相加运算,进行初步融合,获得初始特征 F_s ;之后,利用 PPA 模块对初始融合的特征进行全局上下文语义信息的增强,分别获得 F_{p_1} 和 F_{p_2} ;接着,利用 CTDAM 输出的第

三层和第五层特征对 F_{p_1} 和 F_{p_2} 进行局部信息加强。CTDAM 中的计算公式如下:

$$F_{p_1} = \text{PPA}(\text{Conv}_{3 \times 3}(F_{\text{CTDAM}}^3) + \text{Conv}_{3 \times 3}(F_{\text{CTDAM}}^4)) \times \text{Conv}_{3 \times 3}(F_{\text{CTDAM}}^3) \quad (10)$$

$$F_{p_2} = \text{PPA}(\text{Conv}_{3 \times 3}(F_{\text{CTDAM}}^5) + \text{Conv}_{3 \times 3}(F_{\text{CTDAM}}^4)) \times \text{Conv}_{3 \times 3}(F_{\text{CTDAM}}^5) \quad (11)$$

式中: F_{CTDAM}^i 表示 CTDAM 纹理差异放大后的特征; $\text{Conv}_{3 \times 3}$ 表示 3×3 卷积层; PPA 表示位置感知通道并行注意力。PPA 模块利用空间注意力(ATs)和通道注意力(ATc),将融合后的局部上下文信息和全局上下文信息输入到 PPA 模块中,其具体表达如下:

$$\text{PPA}(F) = \text{ATs}(F_s) \otimes \text{ATc}(F_s) \quad (12)$$

$$\text{ATs}(F_s) = \sigma(\text{Conv}_{1 \times 1}([P_{\text{avg}}(F); P_{\text{max}}(F)])) \quad (13)$$

$$\text{ATc}(F_s) = \sigma(\text{FFN}(P_{\text{avg}}(F)) + \text{FFN}(P_{\text{max}}(F))) \quad (14)$$

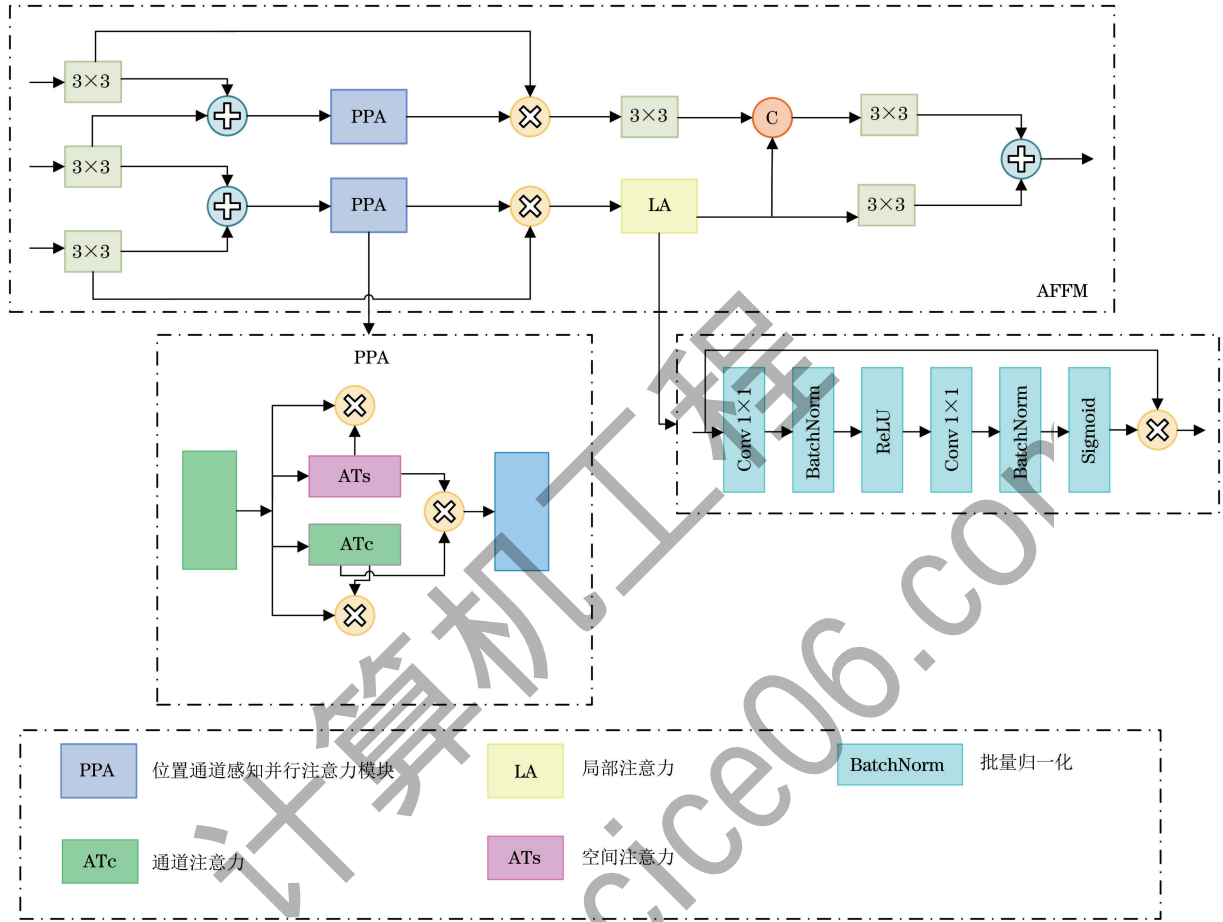


图 3 AFFM 模块结构

Fig.3 AFFM module structure

式中: FFN 表示一个 3 层前馈神经网络, 隐层维数为 1, 输入层和输出层维数相同; P_{avg} 和 P_{max} 分别表示平均池化和最大池化操作。

这时有 2 个分支, 第一个分支通过使用 3×3 卷积使得感受野机制有效地放大了背景与目标之间的差异, 第二个分支则是利用 LA 模块, 从局部的角度增加高层特征的语义信息, 获得丰富的局部特征 F_{p22} 。然后, 将 F_{p22} 与第一个分支进行融合, 但是这样融合后的特征依然会有噪声干扰, 因此, 让每个分支再通过一次 3×3 卷积来去除噪声, 最终得到融合后的特征图 F_{AFFM} :

$$F_{AFFM} = \text{Conv}_{3 \times 3} (C(\text{Conv}_{3 \times 3} (F_{p1}), \text{LA}(F_{p2})) + \text{Conv}_{3 \times 3} (\text{LA}(F_{p2}))) \quad (15)$$

式中: LA 表示局部注意力。

2.5 边缘推理模块

如图 4 所示, 边缘推理模块 BIM 的设计融入了 FBSM 模块提供的边界指导因子 F_ℓ 来表示学习, 以增强对象结构语义的特征表示。由于在特征中每一个特征通道都包含着不同的语义信息, 为了使得融

合后的特征有更好的表达, 本文引入了多尺度通道注意力(MSCA), 该模块可以利用特征图的不同通道维度来收集和整合上下文信息。随后, 为了在 MSCA 的基础上寻找位置信息聚集的区域, 又引入空间注意力模块(SA), 这有助于该模型在多伪装目标的情况下仍具有较高的检测效果。具体地, 首先将下采样的边界指导因子 F_ℓ 与特征 F_{AFFM} 进行逐元素相乘, 然后利用附加的跳跃连接和 3×3 卷积完成初始的特征融合, 获得初始融合特征 F_z , F_z 可以公式化为:

$$F_z = \text{Conv}_{3 \times 3} ((F_{AFFM} \times D(F_\ell)) + F_{AFFM}) \quad (16)$$

接下来, 特征 F_z 被顺序地馈送到 MSCA 和 SA 中以获得重新加权的特征 F_{rw} , F_{rw} 可以公式化为:

$$F_{rw} = \omega_{SA} (\omega_{MS} (F_z) \times F_z) \times (\omega_{MS} (F_z) \times F_z) \quad (17)$$

式中: ω_{SA} 表示 SA 模块获得关注权重的操作。

然后, 将重新加权的特征 F_{rw} 与边界指导因子 F_ℓ 相乘, 得到细化特征 F_y , 这不仅可以抑制背景的干扰, 还可以获得丰富的伪装目标边界信息。细化

特征 F_y 可以公式化为:

$$F_y = (F_{rw} \times F_\ell) + F_{AFFM} \quad (18)$$

最后,针对细化特征 F_y ,通过 1×1 卷积层来获得最终输出 F_{BIM} 。

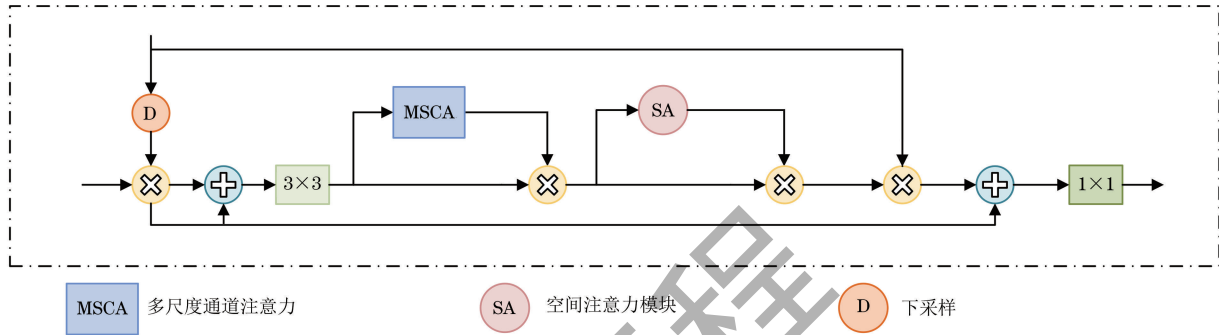


图 4 BIM 模块结构

Fig.4 BIM module structure

3 实验设计与结果分析

3.1 实验细节

3.1.1 实验环境与参数设置

实验环境采用 NVIDIA GeForce RTX3090 显卡, Intel® Core™ i9-12700KF@3.61 GHz 处理器, 使用 PyTorch 框架实现。在训练过程中, 迭代批次为 16, 初始学习率为 1×10^{-4} , 学习率每 30 个 epoch 除以 10。在测试阶段, 输入图像大小为 352×352 , 批量大小设置为 20, 模型在 200 个 epoch 上进行训练, 并传入本文算法中获得最终的预测图。

3.1.2 数据集

本文采用 3 个公共伪装目标数据集, 即 CAMO^[25]、CHAMELEON^[26]、COD 10K^[27]。CAMO 是一个小规模 COD 数据集, 其中包括 1 250 张伪装图像, 分为 8 类, 其中 1 000 张伪装图像用于训练, 其余 250 张图像用于测试^[28]。CHAMELEON 数据集是目前最小的 COD 数据集, 包含 76 张图像。COD 10K 是目前最大、最新的伪装对象数据集, 包含从多个摄影网站下载的 5 066 张伪装图像, 涵盖 5 个超级类别(两栖动物、水生动物、飞行动物、陆地动物以及其他类)和 69 个子类, 其中有 3 040 个训练图像和 2 026 个测试图像。本文根据上述 3 个数据集对网络模型进行评估, 为了实验的公平性, 训练图像由 COD 10K 的 3 040 张训练图像和 CAMO 的 1 000 张训练图像组成。

3.1.3 评价指标

本文采用 4 个广泛使用的指标来比较模型性能。F-measure(F_β) 综合度量查全率和查准率。S-measure(S_α) 评估预测图和真实图之间的结构相似性。E-measure(E_ϕ) 基于人类视觉感知机制评估 COD 的整体和局部精度, 这个指标适用于评估伪装

目标检测的整体和局部精确度。平均绝对误差 (MAE) 评估预测图与地面实况之间的像素级误差。

3.1.4 损失函数

目前, 在图像分割领域中大部分的先进技术普遍采用标准的二值交叉熵损失, 可以独立计算图像的每个像素点的损失, 但没有考虑到图像的全局结构。为此, 本文采用由带权重的交并比 (IoU)^[29] 损失和二值交叉熵 (BCE) 损失^[30] 构成的像素位置感知损失 L_{PPA} , 以解决图像在计算过程中出现全局和局部信息丢失的问题。 L_{IoU}^w 代表带权重的 IoU 损失, 通过增强硬像素的权重来计算全局约束; L_{BCE}^w 代表加权二进制交叉熵损失, 相较于 L_{IoU}^w , 其更加注重复杂的像素而不是针对每个像素使用相同的权重来进行局部约束。在显著性目标检测领域, 这些损失函数被证实是有效的。本文利用输出预测图 F_{BIM} 和边界探索输出图 F_ℓ , 总损失函数定义为:

$$L_{all} = L_{PPA}^l(F_{BIM}, GT) + L_{PPA}^e(F_\ell, GT) \quad (19)$$

$$L_{PPA} = L_{PPA}^w + L_{IoU}^w \quad (20)$$

式中: L_{PPA}^l 代表监督细预测; L_{PPA}^e 表示监督边缘预测。

3.2 实验结果分析

3.2.1 与先进方法的比较

1) 比较方法: 为了证明本文方法的有效性, 将其与 8 种最先进的方法进行比较, 包括 PFNet^[31]、SINetV2^[32]、LSR^[33]、C2F-V2^[34]、MFNet^[35]、BSA-Net^[36]、DGNet^[37]、TPRNet^[38]。为了公平比较, 从公共网站上找到上述所有方法所提供的预测结果, 并通过所提供的开源代码重新训练模型。

2) 定量比较: 表 1 显示了本文方法和目前最先进的 8 个方法在 CAMO、CHAMELEON、COD 10K 这 3 个数据集上的 4 个评估指标对比结果, 最优结果加粗标注。从结果可以看出, 本文方法在所有评估指标上都优于对比方法。

表 1 各方法在 3 个数据集上的评价指标结果

Table 1 Evaluation index results of each method on three datasets

方法	CHAMELEON				CAMO-Test				COD 10K-Test			
	$S_{\alpha} \uparrow$	$F_{\beta} \uparrow$	$E_{\varphi} \uparrow$	$MAE \downarrow$	$S_{\alpha} \uparrow$	$F_{\beta} \uparrow$	$E_{\varphi} \uparrow$	$MAE \downarrow$	$S_{\alpha} \uparrow$	$F_{\beta} \uparrow$	$E_{\varphi} \uparrow$	$MAE \downarrow$
PFNet	0.794	0.778	0.783	0.051	0.714	0.685	0.777	0.098	0.753	0.662	0.742	0.065
SINetV2	0.751	0.805	0.814	0.048	0.742	0.667	0.803	0.095	0.733	0.672	0.761	0.055
LSR	0.809	0.818	0.808	0.044	0.769	0.682	0.801	0.092	0.774	0.681	0.794	0.056
C2F-V2	0.868	0.817	0.869	0.048	0.773	0.698	0.811	0.091	0.786	0.679	0.803	0.052
MFNet	0.870	0.821	0.907	0.039	0.769	0.710	0.825	0.085	0.789	0.685	0.861	0.045
BSA-Net	0.872	0.833	0.912	0.037	0.783	0.731	0.842	0.083	0.791	0.705	0.869	0.041
DGNet	0.897	0.841	0.935	0.035	0.787	0.744	0.838	0.080	0.804	0.715	0.877	0.037
TPRNet	0.882	0.828	0.931	0.033	0.817	0.746	0.842	0.077	0.801	0.726	0.877	0.036
本文方法	0.894	0.842	0.941	0.027	0.807	0.778	0.861	0.077	0.816	0.732	0.888	0.034

根据表 1 结果所示,在 CHAMELEON 数据集上,本文方法在 3 个评估指标上均优于对比先进方法,其中 SINetV2 和 LSR 这 2 种方法也在目标边界信息上进行了处理,但效果甚微,无法高效精准地捕捉伪装目标,而本文方法则通过捕捉低层特征的边界信息与高层特征的丰富语义信息来处理对象的边界信息,然后再指导跨层次融合后的高层特征,以提高 COD 的性能。相较于 SINetV2,本文方法的 S_{α} 平均提高了 0.143, E_{φ} 平均提高了 0.127, F_{β} 平均

提高了 0.037,平均绝对误差下降了 0.021。在 CAMO 数据集上,将本文方法与目前性能最好的 TPRNet 方法进行比较,本文方法在 F_{β} 上提高了 0.032, E_{φ} 提高了 0.019,而在 S_{α} 上却要降低 0.01。本文方法在 COD 10K 数据集上表现始终优于其他 COD 方法,这是因为它具备边界增强的功能,能够准确地检测伪装对象并定位其边界。

3)定性比较:图 5 显示本文方法和 8 种对比方法的预测结果直观比较。由于人类伪装在背景环境

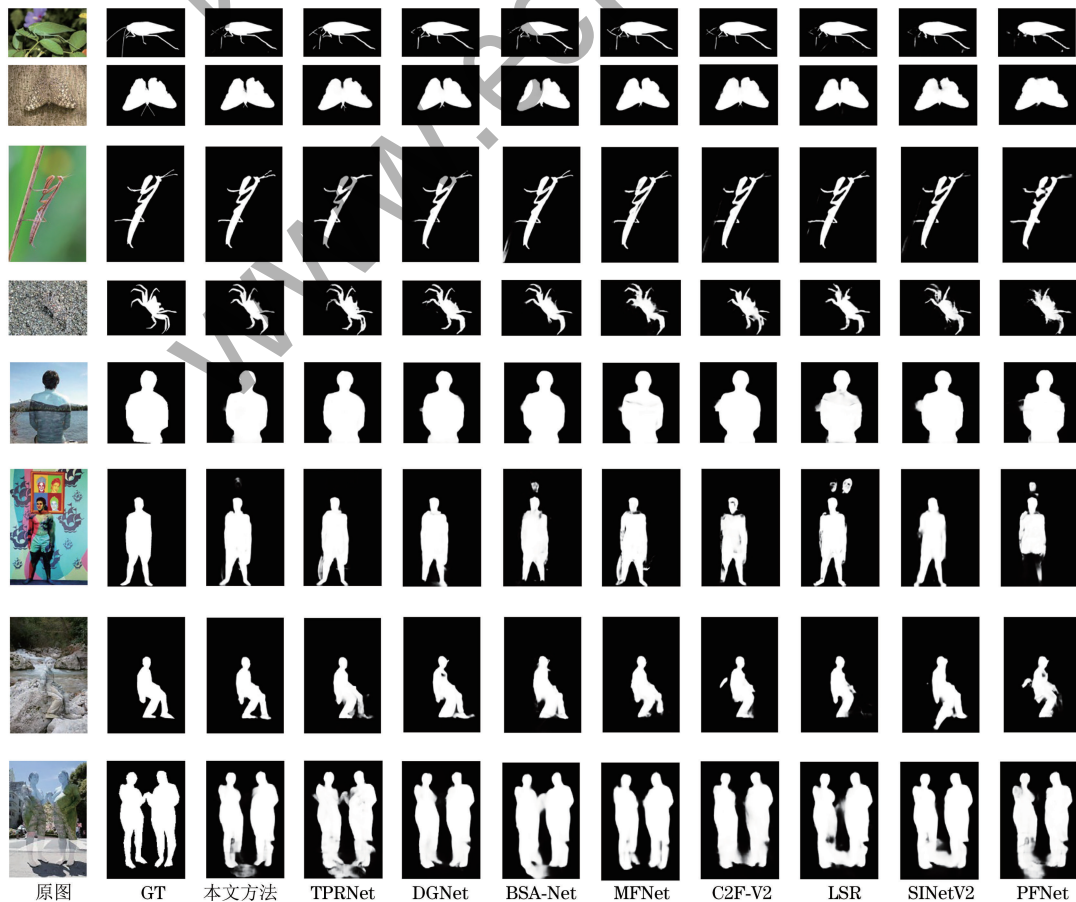


图 5 各方法定性比较结果

Fig. 5 Qualitative comparison results of each method

中往往会更难察觉,因此在预测结果中加入了大量的人类伪装预测效果图,更直观地体现本文方法的检测效果。根据图 5 所示,本文方法的预测结果皆优于对比方法^[39],生成的所有预测图都最接近原始图像。从第 3 行和第 6 行可以看出,在背景去噪和伪装对象定位方面本文方法优于其他对比方法,这说明本文提出的利用最大池化模拟眼球感知细胞来处理伪装对象与边界之间纹理差异的效果显著。从第 4 行和第 5 行可以明显看出,本文方法的边缘细化能力远优于其他方法,这进一步说明本文提出的边界探索模块对高层特征融合的指导效果显著。在第 7 行和第 8 行同时检测多个伪装目标是非常具有挑战性的,然而,本文方法仍然能够有效地检测出所有的伪装对象,并且在处理边缘细节方面也优于其

他对比方法。总体而言,本文方法在面对不同的挑战性因素时都可以取得良好的性能,能够有效地检测出伪装对象。

3.2.2 效率分析

为了研究 BTNet 的效率,将其与其他 8 个先进方法在 MAC 和运行速度方面进行比较,结果如图 6 所示。从图 6 可以看出,TPRNet 的运行速度最快(63.12 帧/s),而本文方法运算速度位列第二(57.57 帧/s),但本文方法的 MAC 较低(14.71 G),这表明本文方法在伪装目标检测过程中运算效率相较于其他方法显著提升,保证了精度和效率的有效平衡,是一种良好的伪装目标检测方案,在自然伪装或军事伪装检测方面有着较高的应用价值。

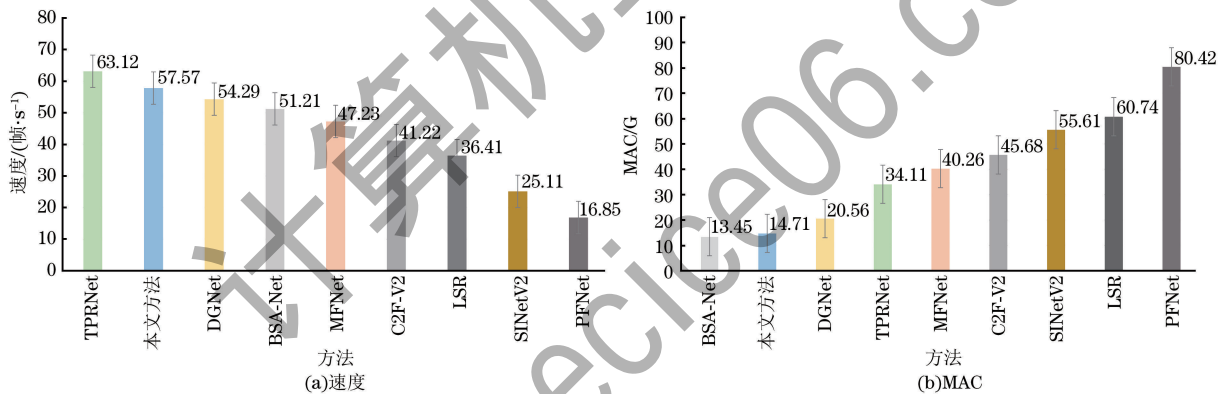


图 6 各方法效率对比结果

Fig.6 Results of efficiency comparison of each method

3.2.3 消融研究

为了验证本文提出的 4 个关键模块,即 FBSM 模块、CTDAM 模块、AFFM 模块以及 BIM 模块在伪装目标检测中的有效性,本节针对各个子模块进行详细的消融实验,结果如表 2 所示。

1) 基线设置。B 表示基线网络,在主干网络 Res2Net-50 提取出的 5 层特征图上使用 1×1 卷积以减少通道数量,然后采用加法和 2 个 3×3 卷积对多层次的特征图进行融合,并作为最终的预测图。

2) FBSM 的有效性。在基线网络中的特征融合阶段引入 FBSM 提供的边界指导因子,从表 2 可知,在 3 个数据集和 4 个评测标准下“B+FBSM”都要比“B”有更好的度量结果,表明 FBSM 可以在特征融合时提供丰富的指导效果,使得预测结果更加精准。

3) CTDAM 的有效性。主干网络提取出的 5 层特征网络中的特征融合预测结果如表 2 所示,其效果比“B+FBSM”稍好,但提升仍然有限。根据以上 2 个实验可以看出,FBSM 提供的边界指导因子

表 2 消融实验结果

Table 2 Results of ablation experiment

实验 分组	方法	CHAMELEON				CAMO-Test				COD 10K-Test			
		S _a ↑	E _φ ↑	F _β ↑	MAE ↓	S _a ↑	E _φ ↑	F _β ↑	MAE ↓	S _a ↑	E _φ ↑	F _β ↑	MAE ↓
1	B	0.811	0.813	0.703	0.060	0.731	0.765	0.611	0.100	0.765	0.786	0.546	0.051
2	B+FBSM	0.823	0.825	0.712	0.055	0.742	0.771	0.632	0.095	0.771	0.791	0.557	0.049
3	B+CTDAM	0.831	0.834	0.725	0.053	0.748	0.781	0.633	0.093	0.774	0.796	0.568	0.048
4	B+FBSM+CTDAM	0.845	0.849	0.742	0.049	0.758	0.793	0.644	0.089	0.782	0.812	0.581	0.046
5	B+FBSM+CTDAM+AFFM	0.867	0.873	0.756	0.041	0.771	0.814	0.668	0.084	0.792	0.833	0.603	0.044
6	本文方法	0.889	0.911	0.788	0.028	0.795	0.836	0.698	0.078	0.812	0.856	0.631	0.039

搭配简单的特征融合时,其检测效果会大打折扣,远没有直接放大纹理差异的 CTDAM 效果好。因此,将两者联合起来后,在 COD 10K 大数据中,相较于“B+CTDAM”,在结构度量 S_α 、自适应度量 E_ϕ 、加权度量 F_β 上分别提高了 0.008、0.016、0.013,平均绝对误差下降了 0.002。

4)AFFM 的有效性。用 AFFM 取代基线网络中简单的特征融合,使得该网络具备整合多元信息的能力。将不同层次被放大的纹理差异区域进行融合,使得预测出的结果在 CHAMELEON 这种小数据中提升非常显著,其 S_α 、 E_ϕ 、 F_β 分别提升了 0.022、0.024、0.014,MAE 下降了 0.008,而在 COD 10K 这种大数据下,其效果仍然要比“B+FBSM+CTDAM”高出很多。

5)BIM 的有效性。在“B+FBSM+CTDAM+AFFM”上引入 BIM,利用 FBSM 所提供的边界指导因子验证边缘线索与伪装目标特征融合运算的有效性。在大数据 COD 10K 下,其效果相较于“B+FBSM+CTDAM+AFFM”在 S_α 、 E_ϕ 、 F_β 方面性能增益分别为 0.02、0.023、0.028,MAE 下降了 0.005。从这可以看出,边界指导因子可以很好地指导和约束特征融合。

4 结束语

本文提出一种基于深度纹理特征的伪装目标边缘细化检测算法。首先,利用自注意力机制设计了一个 FBSM 算法来显式地模拟边界特性,它可以通过增强边界细节来提升 COD 性能;然后,为了更好地利用上下文信息以及目标与背景之间的纹理差异,提出了上下文纹理差异放大模块 CTDAM,其可以在扩大感受野的同时将纹理差异放大;随后,为了捕捉伪装目标的尺度变化,提出了一种注意力特征融合模块 AFFM,利用空间注意力机制将不同层次的特征进行融合,确保特征之间的互补互动和整合;最后,将边界指导因子和融合后的特征注入边缘推理模块 BIM 中以增强对象结构语义的特征表示。经过消融实验评估 BTNet,并将本文方法与目前最先进的 COD 检测方法进行对比,数值和视觉结果均表明,BTNet 优于其他对比方法,验证了其在单目标及多目标情况下对伪装目标检测的可靠性与有效性。

参考文献

[1] LIU J J, HOU Q B, CHENG M M, et al. A simple pooling-based design for real-time salient object detection [C] // Proceedings of IEEE/CVF Conference on Computer Vision

and Pattern Recognition. Washington D. C., USA: IEEE Press, 2019: 3917-3926.

[2] JI G P, FU K R, WU Z, et al. Full-duplex strategy for video object segmentation [C] // Proceedings of IEEE/CVF International Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2021: 4922-4933.

[3] ZHAO J X, LIU J J, FAN D P, et al. EGNet: edge guidance network for salient object detection [C] // Proceedings of IEEE/CVF International Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2019: 8779-8788.

[4] QIN X B, ZHANG Z C, HUANG C Y, et al. BASNet: boundary-aware salient object detection [C] // Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2019: 7479-7489.

[5] ZHOU T, ZHOU Y, GONG C, et al. Feature aggregation and propagation network for camouflaged object detection [J]. IEEE Transactions on Image Processing, 2022, 31: 7036-7047.

[6] CHENG X L, XIONG H, FAN D P, et al. Implicit motion handling for video camouflaged object detection [C] // Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2022: 13864-13873.

[7] YAN J N, LE T N, NGUYEN K D, et al. MirrorNet: bio-inspired camouflaged object segmentation [J]. IEEE Access, 2021, 9: 43290-43300.

[8] LI A X, ZHANG J, LÜ Y Q, et al. Uncertainty-aware joint salient object and camouflaged object detection [C] // Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2021: 10071-10081.

[9] ZHAI Q, LI X, YANG F, et al. Mutual graph learning for camouflaged object detection [C] // Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2021: 12997-13007.

[10] PANG Y W, ZHAO X Q, XIANG T Z, et al. Zoom in and out: a mixed-scale triplet network for camouflaged object detection [C] // Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2022: 2160-2170.

[11] PENG Y K, LIU X Q, ZHAO M Z, et al. Research on camouflage object detection based on improved YOLOv5s [C] // Proceedings of the 8th Annual International Conference on Network and Information Systems for Computers. Washington D. C., USA: IEEE Press, 2022: 644-648.

[12] CHOU M C, CHEN H J, SHUAI H H. Finding the Achilles heel: progressive identification network for camouflaged object detection [C] // Proceedings of IEEE International Conference on Multimedia and Expo. Washington D. C., USA: IEEE Press, 2022: 1-6.

[13] ZADEH A, CHEN M, PORIA S, et al. Tensor fusion network for multimodal sentiment analysis [EB/OL]. [2023-08-05]. <https://arxiv.org/abs/1707.07250>.

[14] ZHOU Z W, SIDDIQUEE M M R, TAJBAKHSH N, et al. UNet++: a nested U-Net architecture for medical image segmentation [EB/OL]. [2023-08-05]. https://link.springer.com/chapter/10.1007/978-3-030-00889-5_1.

[15] ZHOU X W, GUO Y, HE W H, et al. Hierarchical attentional feature fusion for surgical instrument segmentation [C] // Proceedings of the 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society. Washington D. C., USA: IEEE Press, 2021: 3061-3065.

[16] ZHANG P P, WANG D, LU H C, et al. Amulet: aggregating multi-level convolutional features for salient object detection [C] // Proceedings of IEEE International Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2017: 202-211.

- [17] HU K, ZHANG X, LEE D J, et al. Boundary-guided and region-aware network with global scale-adaptive for accurate segmentation of breast tumors in ultrasound images [J]. IEEE Journal of Biomedical and Health Informatics, 2023, 27(9): 4421-4432.
- [18] ZHAO C H, SHEN Y, SU N, et al. Gully erosion monitoring based on semi-supervised semantic segmentation with boundary-guided pseudo-label generation strategy and adaptive loss function[J]. Remote Sensing, 2022, 14(20): 5110.
- [19] XIAO J, CHEN T Y, HU X G, et al. Boundary-guided context-aware network for camouflaged object detection[J]. Neural Computing and Applications, 2023, 35(20): 15075-15093.
- [20] ZHENG Q P, ZHENG L, DENG J K, et al. Transformer-based hierarchical dynamic decoders for salient object detection [J]. Knowledge-Based Systems, 2023, 282: 111075.
- [21] 李军侠, 王星驰, 殷梓, 等. 边缘深度挖掘的弱监督显著性目标检测[J]. 计算机工程, 2023, 49(7): 169-178.
- [22] LI J X, WANG X C, YIN Z, et al. Weakly supervised salient object detection via edge depth mining [J]. Computer Engineering, 2023, 49(7): 169-178. (in Chinese)
- [22] WANG X, DING J J, ZHANG Z, et al. IPNet: polarization-based camouflaged object detection via dual-flow network[J]. Engineering Applications of Artificial Intelligence, 2024, 127: 107303.
- [23] LIANG Y H, QIN G H, SUN M H, et al. A systematic review of image-level camouflaged object detection with deep learning[J]. Neurocomputing, 2024, 566: 127050.
- [24] SHI C J, REN B J, CHEN H R, et al. Camouflaged object detection based on context-aware and boundary refinement[J]. Applied Intelligence, 2023, 53(19): 22429-22445.
- [25] PANG Y W, ZHAO X Q, ZHANG L H, et al. Multi-scale interactive network for salient object detection [C]// Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2020: 9413-9422.
- [26] XU X Q, CHEN S H, LÜ X, et al. Guided multi-scale refinement network for camouflaged object detection [J]. Multimedia Tools and Applications, 2023, 82(4): 5785-5801.
- [27] TANKUS A, YESHURUN Y. Convexity-based visual camouflage breaking [J]. Computer Vision and Image Understanding, 2001, 82(3): 208-237.
- [28] 曹健, 陈怡梅, 李海生, 等. 基于深度学习的道路小目标检测综述[J]. 计算机工程, 2023, 49(10): 1-12.
- [29] CAO J, CHEN Y M, LI H S, et al. Survey of small target detection on roads based on deep learning [J]. Computer Engineering, 2023, 49(10): 1-12. (in Chinese)
- [29] ZHANG C, BI H B, XIANG T Z, et al. Collaborative camouflaged object detection: a large-scale dataset and benchmark[J]. IEEE Transactions on Neural Networks and Learning Systems, 2024, 15: 1-15.
- [30] ZHANG C W, LI X D, LI X D. Decode after filtering: a network for camouflage object segmentation [J]. Soft Computing, 2022, 26(4): 2033-2043.
- [31] MEI H Y, JI G P, WEI Z Q, et al. Camouflaged object segmentation with distraction mining [C]// Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2021: 8772-8781.
- [32] FAN D P, JI G P, CHENG M M, et al. Concealed object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(10): 6024-6042.
- [33] LÜ Y Q, ZHANG J, DAI Y C, et al. Simultaneously localize, segment and rank the camouflaged objects [C]// Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2021: 11591-11601.
- [34] CHENG, LIU S J, SUN Y J, et al. Camouflaged object detection via context-aware cross-level fusion [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 32(10): 6981-6993.
- [35] LIU K, QIU T, YU Y, et al. Edge-guided camouflaged object detection via multi-level feature integration [J]. Sensors (Basel), 2023, 23(13): 5789.
- [36] ZHU H W, LI P, XIE H R, et al. I can find you! boundary-guided separated attention network for camouflaged object detection [J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2022, 36(3): 3608-3616.
- [37] JI G P, FAN D P, CHOU Y C, et al. Deep gradient learning for efficient camouflaged object detection [J]. Machine Intelligence Research, 2023, 20(1): 92-108.
- [38] ZHANG Q, GE Y L, ZHANG C, et al. TPRNet: camouflaged object detection via transformer-induced progressive refinement network[J]. The Visual Computer, 2023, 39(10): 4593-4607.
- [39] 戚玲珑, 高建瓴. 基于改进 YOLOv7 的小目标检测[J]. 计算机工程, 2023, 49(1): 41-48.
- [39] QI L L, GAO J L. Small object detection based on improved YOLOv7[J]. Computer Engineering, 2023, 49(1): 41-48. (in Chinese)