



中国科技期刊卓越行动计划项目入选期刊

控制与决策

CONTROL AND DECISION



基于全局上下文交互融合的伪装目标检测网络

葛斌, 陈宁洁, 夏晨星, 郑海君, 吴涛林

引用本文:

葛斌, 陈宁洁, 夏晨星, 郑海君, 吴涛林. 基于全局上下文交互融合的伪装目标检测网络[J]. 控制与决策, 2024, 39(10): 3347–3356.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2023.1098>

您可能感兴趣的其他文章

Articles you may be interested in

多目标小尺度车辆目标检测方法

Multi-target and small-scale vehicle target detection method

控制与决策. 2021, 36(11): 2707–2712 <https://doi.org/10.13195/j.kzyjc.2020.0635>

基于双分支特征融合的场景文本检测方法

A scene text detection based on dual-path feature fusion

控制与决策. 2021, 36(9): 2179–2186 <https://doi.org/10.13195/j.kzyjc.2020.0002>

基于双边分解与 L_1 暗通道的战场图像增强算法

A battlefield image enhancement algorithm based on bilateral decomposition and L_1 dark channel prior

控制与决策. 2021, 36(5): 1165–1172 <https://doi.org/10.13195/j.kzyjc.2019.1213>

基于FRC框架的红外与可见光图像融合方法

Infrared and visible image fusion based on FRC algorithm

控制与决策. 2021, 36(11): 2690–2698 <https://doi.org/10.13195/j.kzyjc.2020.0669>

尺度自适应的多特征融合相关滤波目标跟踪算法

Scale adaptation and multi-feature fusion correlation filtering object tracking algorithm

控制与决策. 2021, 36(2): 429–435 <https://doi.org/10.13195/j.kzyjc.2019.0445>

基于全局上下文交互融合的伪装目标检测网络

葛斌^{1,2†}, 陈宁洁¹, 夏晨星^{1,2}, 郑海君¹, 吴涛林¹

(1. 安徽理工大学 计算机科学与工程学院, 安徽 淮南 232001;
2. 合肥综合性国家科学中心 能源研究院, 合肥 230031)

摘要: 在伪装目标检测中, 针对以往的特征融合研究大多采用多级特征集成, 而忽略了不同特征间的差异, 提出一种基于全局上下文交互融合网络用于伪装目标检测. 利用改进的金字塔视觉转换器 (PVTv2) 模型作为骨干网络, 在多个尺度上提取全局上下文信息. 首先, 设计边界增强模块来关注伪装目标的结构细节, 并获取物体的边缘特征; 然后, 借鉴动物捕食机制提出特征融合解码模块, 该模块提供位置信息用于潜在目标定位来产生粗略定位图; 最后, 通过所搭建的全局上下文聚合模块进行多层次信息的充分交互, 减少特征聚合过程中的信息丢失. 采用 4 个公开数据集、4 种评价指标进行实验, 实验结果表明, 所提网络性能优于其他 17 个具有代表性的模型.

关键词: 计算机视觉; 图像处理; 伪装目标检测; 上下文信息; 特征融合

中图分类号: TP391 文献标志码: A

DOI: 10.13195/j.kzyjc.2023.1098

引用格式: 葛斌, 陈宁洁, 夏晨星, 等. 基于全局上下文交互融合的伪装目标检测网络[J]. 控制与决策, 2024, 39(10): 3347-3356.

Camouflaged object detection network based on global context interaction fusion

GE Bin^{1,2†}, CHEN Ning-jie¹, XIA Chen-xing^{1,2}, ZHENG Hai-jun¹, WU Tao-lin¹

(1. School of Computer Science and Engineering, Anhui University of Science and Technology, Huainan 232001, China;
2. Institute of Energy, Hefei Comprehensive National Science Center, Hefei 230031, China)

Abstract: In camouflaged object detection, most previous studies on feature fusion have predominantly used multi-level feature integration while neglecting the differences between various features. In this paper, a global context interaction fusion network is proposed for camouflaged object detection, which employs an improved pyramid vision transformer (PVTv2) model as the backbone network to extract global context information at multiple scales. First, a boundary enhancement module is designed to focus on the structural details of camouflaged objects and acquire the edge features of objects. Second, inspired by the hunting mechanisms of animals, a feature fusion decoder module is proposed, which provides position information for potential object localization to produce a rough localization map. Finally, a global context aggregation module is constructed to fully interact with multi-level information and reduce information loss during feature aggregation. Extensive experiments on four publicly available datasets demonstrate that the proposed method surpasses that of 17 other state-of-the-art models under four evaluation metrics.

Keywords: computer vision; image processing; camouflaged object detection; contextual information; feature fusion

0 引言

伪装是一种常见的防御策略, 它能够帮助动物或植物隐蔽在周围的环境中, 以保护自身免受自然天敌的侵害. 伪装物大致可分为自然伪装物和人工伪装物, 如雪地里的白鸽、执行一定任务的军人身穿军用迷彩服等.

伪装目标检测 (camouflaged object detection, COD) 旨在准确、有效地发现“完美”镶嵌在周边环境中的伪装物体, 这使得伪装目标检测相比于传统目标检测更具有挑战性.

近年来, COD 逐渐引起了研究者的兴趣, 除了本身具有的科学研究价值, 在医学 (如肺部感染分割^[1])、

收稿日期: 2023-08-03; 录用日期: 2023-11-03.

基金项目: 国家自然科学基金项目 (62102003); 安徽省自然科学基金项目 (2108085QF258); 安徽省博士后基金项目 (2022B623).

责任编辑: 夏元清.

†通讯作者. E-mail: bge@aust.edu.cn.

工业(如物品表面缺陷的检测^[2])、农业(如蝗虫检测)、艺术(如风景图像检测^[3])等领域均有很多潜在的应用,这表明伪装目标检测技术的研究具有极其重要的意义。

伪装目标检测方法大致可分为两类:一类是传统方法,另一类是基于深度学习的方法。传统方法主要采用手工设计的特征来检测伪装对象^[4-5],如颜色、纹理、光照等。但是在复杂的情况下,基于手工提取的特征往往会失效。

随着深度学习技术的进步,研究者提出了很多基于深度学习的伪装目标检测方法,并取得了显著的效果。Zhuge等^[6]提出了CubeNet框架,通过将 X 连接引入标准的编码器-解码器架构,设计了一种多进多出融合模块,并通过 X 连接将这些融合模块集成至方形融合解码器和子边缘解码器中;Hu等^[7]提出了以迭代反馈的方式通过高分辨率特征来细化低分辨率特征,并设计了一种迭代反馈损失,对每个反馈连接施加更多约束;Ji等^[8]介绍了一种新的利用目标梯度监督进行伪装目标检测的框架,它将任务解耦为两个分支,即上下文编码器和纹理编码器。虽然上述融合方法能够提高伪装目标检测的性能,但是在融合的过程中易忽略与背景相似的部分,导致丢失重要的特征信息以及特征之间能够相互利用的补充信息不足,使得得到的最终预测图不够准确。

研究表明模仿动物捕食机制有助于精确地区分前景和背景信息,即先在复杂环境中粗略地定位目标,随后聚焦于定位的区域,从而准确地将伪装对象与背景区分开。Yue等^[9]受生物学研究中搜索和识别的启发,提出了一个COD检测框架,其借助于区域搜索模块找到目标大概的位置,然后利用区域细化模块来融合特征以识别目标的准确位置;Zheng等^[10]提出了受人类启发的框架,它模拟了人类在图像中从多个角度、多个距离、多个视角寻找模糊物体的行为;Xing等^[11]提出了一种三阶段架构,分别为搜索、放大和识别,搜索和放大部分旨在粗略定位和放大对象,识别部分旨在对对象进行细化。

尽管COD领域已取得了一定的进展,但是现有的方法对于伪装目标检测仍然存在局限性。此外,目标检测任务通常将边界信息视为关键要素,对伪装对象边界进行定位有利于提高伪装目标检测的性能。He等^[12]设计了边缘重建模型,利用可学习的小波将特征分解为不同的频域,专注于信息最丰富的波段,以挖掘区分前景和背景的微妙线索;Jia等^[13]提出了第1个迭代细化框架,设计了一种判别掩模,使得网

络更多地关注与伪装对象相关的重要的纹理和边缘。

为此,本文提出基于全局上下文交互融合的伪装目标检测网络(COD network based on global context interactive fusion, GCIFNet)。具体而言,首先,使用边界增强模块(boundary enhancement module, BEM)对边界增强的表示进行学习;然后,考虑动物捕食过程中的两阶段过程,提出特征融合解码模块(feature fusion decoder module, FFDM),利用骨干网络提取的特征进行粗略定位;最后,采用全局上下文聚合模块(global context aggregation module, GCAM)实现低级特征、高级特征与全局特征间的完全交互,它对检测不同尺寸的对象具有重要的意义。在4个基准数据集上进行大量实验,实验结果表明GCIFNet具有较优性能。

1 方法

在本节中,首先,概述所提出网络模型GCIFNet;然后,详细描述该模型的各组成部分;最后,提出该模型的整体损失函数并验证其有效性。

基于全局上下文交互融合的伪装目标检测网络框架如图1所示。首先,使用预先训练好的PVTv2^[14]作为骨干网络,分别得到输入图像分辨率为 $(1/4, 1/8, 1/16, 1/32)$ 的多尺度特征 $X_i(i = \{1, 2, 3, 4\})$;将特征 X_1 和 X_2 输入BEM模块中来捕获边界信息;将特征 X_2 、 X_3 和 X_4 输入FFDM模块中,生成粗略的定位图。然后,利用DASPP^[15]扩大感受野来捕获更多的细节信息。接着,为了实现多尺度特征的全局交互,利用BEM模块获得的边界信息和FFDM模块获得的位置信息对GCAM模块进行增强。最后,利用GCAM模块获得最终检测结果,同时使用多级监督策略训练网络。下文将对各模块以及损失函数进行详细阐述。

1.1 边界增强模块

现有的COD方法中由于边缘破坏或身体轮廓被遮挡,往往难以有效、完整地识别出物体的结构和细节。针对此问题,本文提出边界增强模块BEM,如图2所示。将特征 X_1 和 X_2 送入BEM模块中,首先,将特征 X_1 经平均池化操作得到与特征 X_2 相同大小的特征;然后,将池化后的特征和 X_2 经卷积后的结果进行逐元素相乘来提取共享信息,并将共享信息添加至原来的特征中,经卷积融合操作得到特征 X_{12}^t, X_{12}^f 通过卷积、批量归一化和激活函数生成权重值 X_w ;最后,将 X_w 与 X_{12}^t 进行逐元素相乘和相加得到边界信息 B_{out} 。此过程可描述如下:

$$X_1^t = \text{AVG}(X_1) \otimes \text{Conv}_{3 \times 3}(X_2) \oplus \text{AVG}(X_1), \quad (1)$$

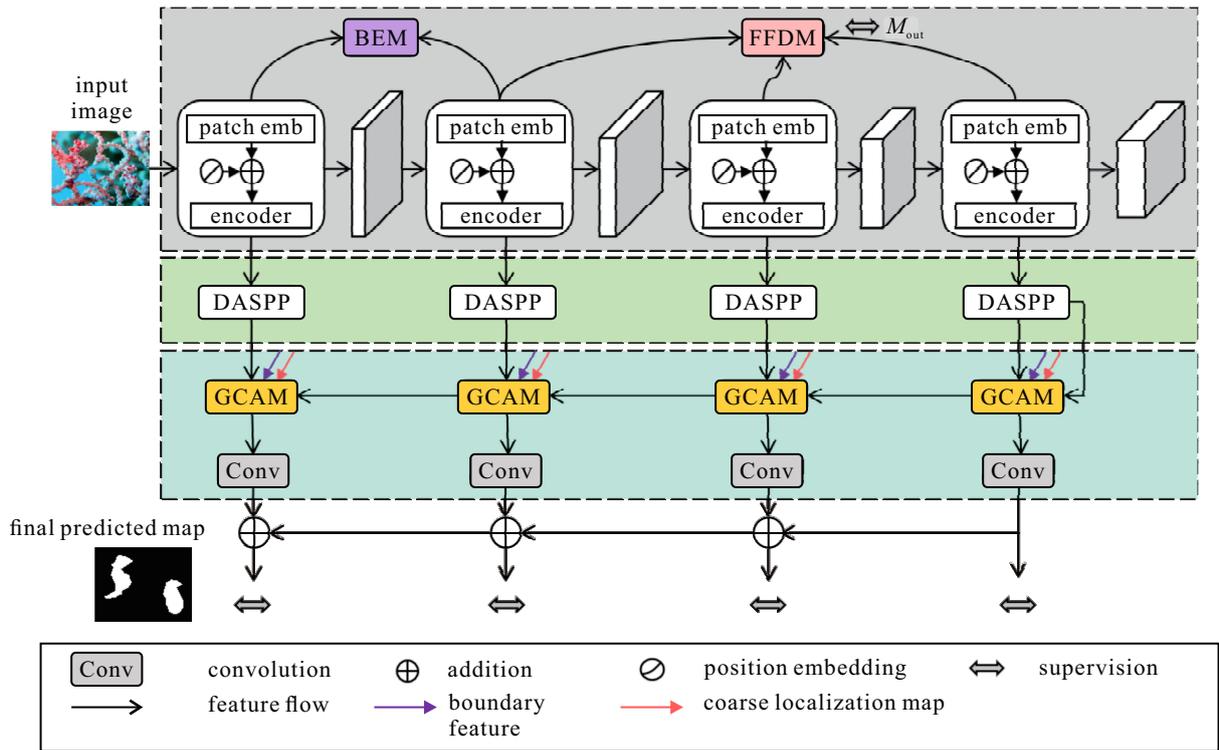


图1 GCIFNet框架

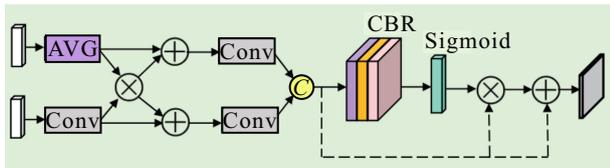


图2 BEM结构

$$X_2^t = \text{Conv}_{3 \times 3}(X_2) \otimes \text{AVG}(X_1) \oplus \text{Conv}_{3 \times 3}(X_2), \quad (2)$$

$$X_{12}^t = \text{Cat}[\text{Conv}_{3 \times 3}(X_1^t), \text{Conv}_{3 \times 3}(X_2^t)], \quad (3)$$

$$X_w = \delta(\text{CBR}(X_{12}^t)), \quad (4)$$

$$B_{\text{out}} = X_w \otimes X_{12}^t \oplus X_{12}^t. \quad (5)$$

其中: $\text{Conv}_{3 \times 3}$ 为 3×3 卷积操作, AVG 为全局平均池化, δ 为 Sigmoid 激活函数, CBR 为卷积、批量归一化和 Relu 激活函数, Cat 表示在通道上拼接操作, \otimes 表示逐元素相乘, \oplus 表示逐元素相加。

1.2 特征融合解码模块

特征融合策略是提升网络性能的一个重要因素。以往的研究中,如采用跳跃连接^[16]、U型连接^[17]等策略实现特征的有效融合,但是它们易引入大量的噪声,使得网络的整体性能降低。为了解决这一问题,本文提出特征融合解码模块 FFDM, FFDM 模块包括图卷积(graph convolution, GC)模块和特征细化(feature refinement, FR)模块,如图3所示。将特征 X_2 、 X_3 和 X_4 送入 FFDM 模块中,首先,将 $X_i (i = \{2, 3, 4\})$ 并行输入 GC 模块,对伪装目标区域的特征进行增强以

充分利用有效的信息,提高特征的表达能;然后,将生成的结果经 FR 模块得到进一步细化的特征 $X_{i1} (i = \{2, 3, 4\})$,且各层之间进行相互融合得到最终的粗略定位图 M_{out} 。此过程可描述如下:

$$X_{41} = \text{FR}(\text{GC}(X_4)), \quad (6)$$

$$X_{31} = \text{FR}(\text{GC}([\zeta(X_{41}) \oplus X_3] \otimes \zeta(X_{41}))), \quad (7)$$

$$X_{21} = \text{FR}(\text{GC}([\zeta(X_{41}) \oplus X_2 \oplus \zeta(X_{31}) \oplus \zeta(X_3)] \otimes \zeta(X_{31}) \otimes \zeta(X_{41}))), \quad (8)$$

$$M_{\text{out}} = \text{Cat}[\zeta(\text{Cat}(\zeta(X_{41}), X_{31})), \text{Cat}(\zeta(X_{31}), X_{21})]. \quad (9)$$

其中: ζ 为上采样操作, GC 和 FR 模块的操作将在下文详细介绍。

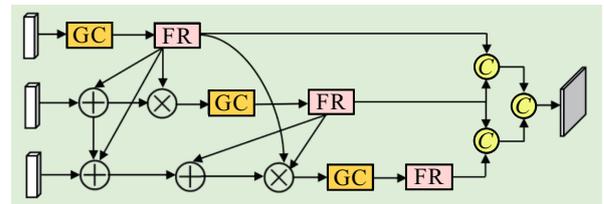


图3 FFDM结构

1.2.1 图卷积(GC)模块

目前,图卷积的主流形式大致可归为两大类:一类是基于频谱的图卷积,另一类则是基于空间的图卷积。为了更好地利用图卷积域下的非局部信息,本文引入了一种两分支图卷积结构^[18]。如图4所示,改进

后的特征所蕴含的全局上下文信息十分丰富. 第1个分支对图像内像素间的空间关系进行建模, 首先, 将输入特征 F 投影至一个坐标空间 Ω 中; 然后, 进行下采样操作将输入特征变换为新坐标空间内的特征 F^* 来实现全局推理, 在特征 F^* 上采用3种可学习的线性变换 k 、 q 和 v 来生成所需特征 G_1 ; 最后, 将特征 G_1 经上采样映射至原始的坐标空间, 得到最终输出特征 G_2 . 此过程可描述如下:

$$F^* = \text{AVG}(F), \tag{10}$$

$$k = \text{Conv}_{1 \times 1}^k(F^*), \tag{11}$$

$$q = \text{Conv}_{1 \times 1}^q(F^*), \tag{12}$$

$$v = \text{Conv}_{1 \times 1}^v(F^*), \tag{13}$$

$$G_1 = \text{Conv}_{1 \times 1}[\delta(k \otimes q^T) \otimes v], \tag{14}$$

$$G_2 = \text{Conv}_{1 \times 1}(\zeta(G_1)), \tag{15}$$

其中 $\text{Conv}_{1 \times 1}$ 为 1×1 卷积操作.

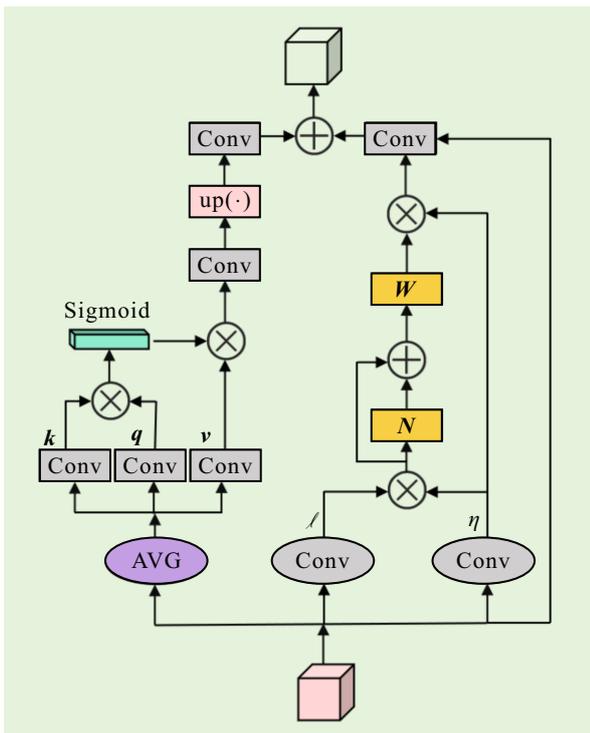


图4 GC结构

第2个分支在网络特征图通道维度上进行相互依赖关系的建模. 首先, 将输入特征 F 投影至特征空间 Θ , 使用函数 ℓ 减小输入特征的维数, 投影函数 η 作为输入特征的线性组合 (其中函数 ℓ 和 η 为通过 1×1 卷积层), 使得得到的新特征 G_{12} 可聚合来自多个区域的信息; 然后, 通过构建邻接矩阵得到新的特征 G_{13} , 将新特征 G_{13} 映射至具有输出的原始坐标空间得到最终特征 G_3 ; 最后, 将两个分支的结果和输入特征进行相加得到特征 G_o . 此过程可描述如下:

$$G_{12} = \ell(F) \otimes \eta(F), \tag{16}$$

$$G_{13} = (1 - N) \otimes G_{12} \otimes W, \tag{17}$$

$$G_3 = \text{Conv}_{1 \times 1}(G_{13} \otimes \eta(F)), \tag{18}$$

$$G_o = G_3 \oplus G_2 \oplus F. \tag{19}$$

其中: $\text{Conv}_{1 \times 1}$ 为 1×1 卷积, N 为通过拉普拉斯平滑构建的邻接矩阵, W 为可训练的权值.

1.2.2 特征细化(FR)模块

大多数研究表明, 网络中最终的输出结果通常会带有一些局部细节方面的附加信息影响模型的检测性能. 为此, 本文设计了一个特征细化模块, 如图5所示. 首先, 将特征经 1×1 卷积和 Sigmoid 函数得到注意图 a_1 , 将其与输入的特征进行逐元素相乘; 同时, 将输入特征与反向的注意图进行相乘以进一步挖掘信息; 最后, 将两个结果进行相加和卷积得到最终的输出特征 A_{out} . 此过程可描述如下:

$$a_1 = \delta(\text{Conv}_{1 \times 1}(G_o)), \tag{20}$$

$$A_{out} = \text{CBR}[(a_1 \otimes G_o) \oplus ((1 - a_1) \otimes G_o)], \tag{21}$$

其中 $\text{Conv}_{1 \times 1}$ 为 1×1 卷积.

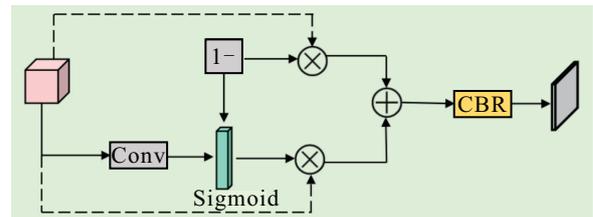


图5 FR结构

1.3 密集空洞空间金字塔池化(dense atrous spatial pyramid pooling, DASPP)

在级联融合操作中, DASPP 将空洞卷积与逐层增加的膨胀率相结合, 使得低层的膨胀率较低, 高层的膨胀率较高. 同时, 空洞卷积通过跳跃连接共享信息, 而扩张率较小的层与扩张率较大的层相互依赖. DASPP 最终呈现的是由多个比率和尺寸的空洞卷积所形成的特征图, 不仅会形成更加密集的特征金字塔, 还会产生更大的感受野, 以挖掘更丰富的上下文信息.

1.4 全局上下文聚合模块

针对缺乏对象不同部分之间或多个对象间的全局语义关系, 简单的高、低特征融合无法精准地捕捉到伪装物体. 为此, 本文设计了一个全局上下文聚合模块 GCAM, 该模块通过挖掘层次间的相关性来有效聚合特征信息, 如图6所示. 首先对于第 k ($k \in \{1, 2, 3\}$) 阶段的 GCAM, 将第 $k + 1$ 阶段 (记为 G_{k+1}) 的

输出与 DASPP 模块当前阶段 (记为 D_k) 获得的特征相融合, 将融合的特征经 3×3 卷积得到特征 GD; 然后, 将经 BEM 模块得到的边界信息 B_{out} 经 3×3 卷积操作, 得到的特征经上采样操作与 GD 进行逐元素相乘再与残差连接得到输出结果 B_{out}^1 ; 接着, 将通过 FFDM 模块得到的粗略定位图 M_{out} 进行与边界信息 B_{out} 相同的操作得到特征 M_{out}^1 ; 最后, 将特征 GD、 B_{out}^1 和 M_{out}^1 进行聚合得到 F_{out} . 此过程可描述如下:

$$GD = \text{Conv}_{3 \times 3}(\text{Cat}(D_k, G_{k+1})), \quad (22)$$

$$B_{out}^1 = \zeta(\text{CB}(B_{out})) \otimes GD \oplus \zeta(\text{CB}(B_{out})), \quad (23)$$

$$M_{out}^1 = \zeta(\text{CB}(M_{out})) \otimes GD \oplus \zeta(\text{CB}(M_{out})), \quad (24)$$

$$F_{out} = \text{Cat}(B_{out}^1, M_{out}^1, GD). \quad (25)$$

其中: $\text{Conv}_{3 \times 3}$ 为 3×3 卷积, CB 为批量归一化的 3×3 卷积.

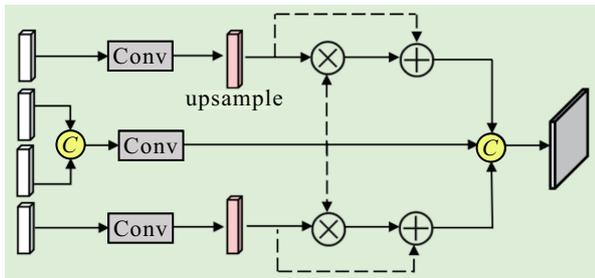


图6 GCAM结构

1.5 损失函数

在 COD 任务中, 二值交叉熵损失 (L_{BCE}) 广泛用于计算每个像素的损失. 然而, 当前景像素的数量远远小于背景像素的数量时, 模型可能会严重偏向背景. 为了解决这个问题, 本文采用了加权二值交叉熵损失 (L_{wBCE})^[19], 根据中心像素与其周围像素的差值为像素分配不同的权值. 基于标准 IOU 损失和加权 IOU 损失在显著目标检测^[19]中的有效性, 本文使用加权 IOU 损失 (L_{wIOU}) 实现全局约束. 在这种情况下, 所提出模型的基本损失函数定义为

$$L = L_{wBCE} + L_{wIOU}, \quad (26)$$

其中 L_{wBCE} 和 L_{wIOU} 分别为加权 BCE 损失和加权 IOU 损失. 该模型有 5 个监督输出, 包括粗略定位映射 M_{out} 和每个阶段得到的 4 个预测映射 (记为 P_1 、 P_2 、 P_3 和 P_4). 模型训练的总损失定义为

$$L_{total} = \sum_{i=1}^4 L(P_i, G) + L_{BCE}(M_{out}, G). \quad (27)$$

这里: G 为地面真值, 对粗略定位图的监督基于标准 BCE 损失, 而对 4 个阶段的预测图采用式 (26) 进行监督.

2 实验和结果

2.1 数据集

本文使用 4 个 COD 数据集 CAMO^[20]、CHAMELEON^[21]、COD 10 K^[22] 和 NC 4 K^[23] 来评估所提出方法. CAMO 由 1 250 张伪装图像和 1 250 张非伪装图像组成; CHAMELEON 包含 76 张手工标注的图像; COD 10 K 包括 5 066 张伪装图像、3 000 张背景图像和 1 934 张非伪装图像; NC 4 K 为另一个大型 COD 测试数据集, 包括 4 121 张来自互联网的图像.

2.2 实验设置

所提出方法是使用 PyTorch 实现的, 利用 Adam 优化器更新网络参数. 所有实验均是在 NVIDIA A 4000 显卡上进行的. 使用双线性插值将输入图像的大小调整至 352×352 . 在训练过程中: 最大 epoch 为 100, 批处理数为 9, 学习速率为 $8e-5$.

2.3 评价指标

本文使用常用的 4 个指标进行定量分析, 分别为结构度量 (S_α)^[24]、平均增强度量 (E_ϑ)^[25]、加权 F 度量 (F_β^w)^[26] 和平均绝对误差 (MAE)^[27]. 其中: S_α 、 E_ϑ 和 F_β^w 值越大、MAE 的值越小, 表明性能越好.

S 度量 (S_α): 是一种高效的、易于计算的方法, 侧重于结构相似性的测量. 具体可描述为

$$S_\alpha = \alpha \times S_o + (1 - \alpha) \times S_r. \quad (28)$$

其中: S_o 获取目标感知, S_r 获取区域观测特征, α 为参数.

E 度量 (E_ϑ): 将局部的像素值与图像级均值相结合来评估伪装目标检测结果的整体和局部精度. 具体可描述为

$$E_\vartheta = \frac{1}{(W \times H)} \sum_{x=1}^W \sum_{y=1}^H \vartheta(C(x, y) - G(x, y)). \quad (29)$$

其中: ϑ 为增强一致性矩阵, W 和 H 分别为输入的宽度和高度, C 和 G 分别为预测图和真值图.

F 度量 (F_β^w): 是 F -measure 的改进版本, 利用了加权精确度和召回率. 具体可描述为

$$F_\beta^w = \frac{(\beta^2 + 1) \times PR}{\beta^2 P + R}. \quad (30)$$

平均绝对误差 (MAE): 用来计算预测图和真值图间每个像素的平均绝对误差. 具体可描述为

$$MAE = \frac{1}{(H \times W)} \sum_{x=1}^H \sum_{y=1}^W |C(x, y) - G(x, y)|. \quad (31)$$

2.4 与其他 SOTA 方法进行对比

本文将 GCIFNet 与 17 种具有代表性的 COD 方法进行比较, 包括 PoolNet^[28]、EGNet^[29]、SINet^[22]、PraNet^[30]、TINet^[31]、PFNet^[32]、UGTR^[33]、

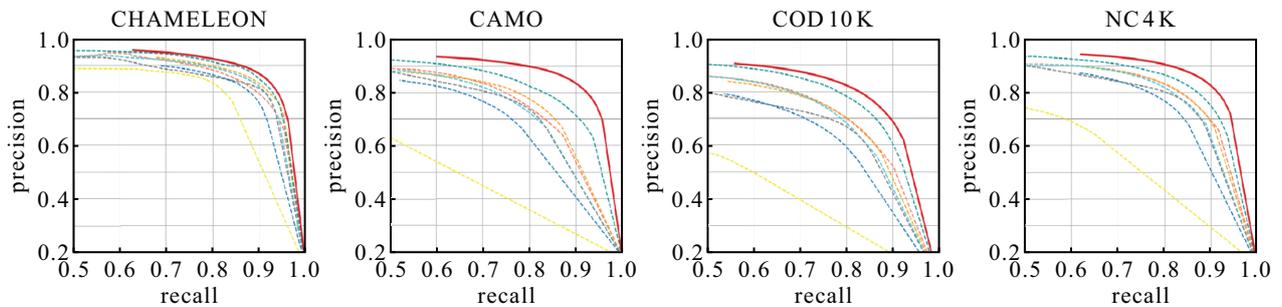
RankNet^[23]、MGL^[34]、C2FNet^[35]、DGNet^[8]、FAPNet^[36]、SINet-2^[37]、ZoomNet^[38]、BSANet^[39]、TPRNet^[40]以及FSPNet^[41].

2.4.1 定量比较

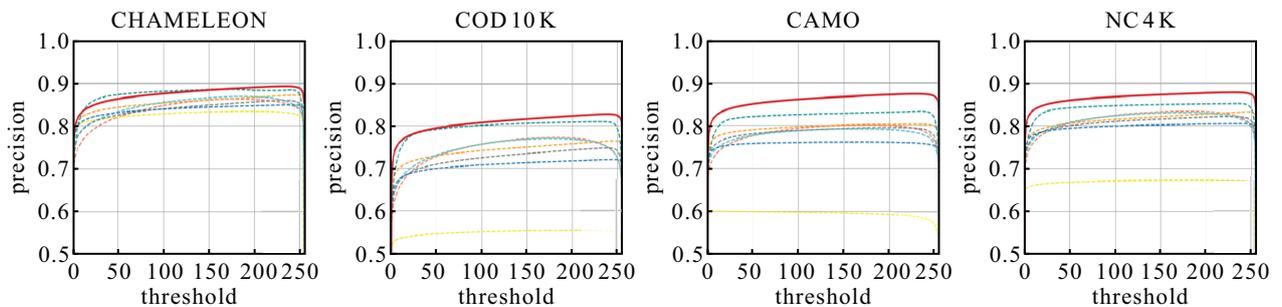
表1为所提出GCIFNet与其他方法在4个COD公开数据集上的定量结果. 对于每个数据集, 可以看

表1 各网络在4个数据集上的指标值对比

model	pub/year	CHAMELEON				COD 10K				CAMO				NC4K			
		$S_\alpha \uparrow$	$E_\theta \uparrow$	$F_\beta^w \uparrow$	MAE \downarrow	$S_\alpha \uparrow$	$E_\theta \uparrow$	$F_\beta^w \uparrow$	MAE \downarrow	$S_\alpha \uparrow$	$E_\theta \uparrow$	$F_\beta^w \uparrow$	MAE \downarrow	$S_\alpha \uparrow$	$E_\theta \uparrow$	$F_\beta^w \uparrow$	MAE \downarrow
Based ResNet 50																	
PoolNet	CVPR/2019	0.854	0.933	0.690	0.054	0.740	0.844	0.506	0.056	0.730	0.819	0.575	0.105	0.785	0.699	0.635	0.073
EGNet	ICCV/2019	0.848	0.831	0.676	0.050	0.737	0.810	0.608	0.056	0.662	0.766	0.612	0.124	0.767	0.850	0.719	0.077
SINet	CVPR/2020	0.872	0.936	0.806	0.034	0.776	0.864	0.631	0.043	0.745	0.804	0.644	0.092	0.808	0.871	0.723	0.058
PraNet	MICCAI/2020	0.868	0.923	0.785	0.037	0.794	0.868	0.645	0.043	0.780	0.835	0.687	0.086	0.829	0.883	0.740	0.055
TINet	TCSVT/2021	0.888	0.911	0.786	0.036	0.803	0.848	0.629	0.041	0.793	0.834	0.690	0.083
PFNet	CVPR/2021	0.882	0.931	0.810	0.033	0.800	0.877	0.660	0.040	0.782	0.842	0.695	0.085	0.829	0.888	0.745	0.053
RankNet	CVPR/2021	0.894	0.942	0.848	0.031	0.809	0.884	0.684	0.035	0.800	0.873	0.728	0.073	0.842	0.898	0.772	0.047
MGL	CVPR/2021	0.893	0.918	0.834	0.030	0.814	0.852	0.711	0.035	0.775	0.812	0.726	0.088	0.833	0.867	0.782	0.052
C2FNet	TCSVT/2022	0.888	0.935	0.828	0.032	0.813	0.890	0.686	0.036	0.796	0.864	0.719	0.080	0.838	0.897	0.762	0.049
FAPNet	TIP/2022	0.893	0.940	0.842	0.028	0.822	0.888	0.731	0.036	0.815	0.865	0.776	0.076
SINet-2	TPAMI/2022	0.888	0.942	0.816	0.030	0.815	0.887	0.680	0.037	0.820	0.882	0.743	0.070	0.847	0.903	0.770	0.048
DGNet	MIR/2022	0.822	0.896	0.693	0.033	0.839	0.899	0.769	0.057	0.857	0.910	0.784	0.042
ZoomNet	CVPR/2022	0.889	0.926	0.835	0.026	0.828	0.923	0.716	0.029	0.793	0.840	0.715	0.072	0.847	0.891	0.782	0.044
BSANet	AAAI/2022	0.895	0.945	0.841	0.027	0.818	0.891	0.698	0.034	0.796	0.851	0.717	0.079
TPRNet	TVCJ/2022	0.897	0.948	0.834	0.028	0.825	0.891	0.704	0.035	0.814	0.868	0.737	0.073	0.825	0.891	0.704	0.035
Based Transformer																	
UGTR	ICCV/2021	0.886	0.940	0.793	0.031	0.818	0.853	0.667	0.035	0.784	0.851	0.684	0.086	0.839	0.874	0.747	0.052
FSPNet	CVPR/2023	0.851	0.895	0.735	0.026	0.856	0.899	0.799	0.050	0.879	0.915	0.816	0.036
ours	2023	0.903	0.947	0.851	0.024	0.858	0.920	0.764	0.026	0.871	0.927	0.823	0.048	0.879	0.926	0.825	0.036



(a) PR curves



(b) F_β curves



图7 Precision-Recall和 F -measure曲线

到 GCIFNet 在 4 个指标上的性能明显优于其他方法. 如与 CAMO 数据集上的 ZoomNet 相比, 所提出方法在结构度量 S_α 、平均增强度量 E_θ 、加权 F 度量 F_β^w 上分别提高了 7.8%、8.7%、10.8%, 平均绝对误差 MAE 下降了 2.4%; 与 CAMO 数据集上的基于 transformer 的 FSPNet 相比, 结构度量 S_α 、平均增强度量 E_θ 、加权 F 度量 F_β^w 分别提高了 1.5%、2.8%、2.4%, 平均绝对误差 MAE 下降了 0.2%. 此外, 如图 7 所示, 本文展示了在 4 个 COD 公开数据集上 GCIFNet 和不同方法的 Precision-Recall 和 F -measure 曲线, 红色的线为 GCIFNet.

2.4.2 定性比较

图 8 为 GCIFNet 与一些具有代表性的 COD 方法的视觉比较结果. 在这组图像中, 它们所呈现的复杂性涵盖了多个方面, 包括大物体、小物体、多个物体、

遮挡、背景干扰以及无法明确定义的边界等. 通过第 2 列、第 4 列、第 5 列和第 8 列, 可以推断出 GCIFNet 相对于其他方法的优势是更加关注伪装目标的边界和真实的位置, 这归功于所提出的边界增强模块和特征融合解码模块. 由图 8 的总体分析可发现, GCIFNet 生成的预测图噪声少, 是因为全局上下文聚合模块将高、低、全局特征进行了融合. GCIFNet 的有效性在这些结果中得到了直观的展现, 即在复杂情况下能够获得较高的准确度.

2.5 消融实验

为了验证所提出的 3 个关键模块, 即 BEM 模块、FFDM 模块以及 GCAM 模块在伪装目标检测中的有效性, 接下来对各个子模块进行一系列消融研究, 所得到的定量结果在表 2 中得以呈现, 可视化比较结果如图 9 所示.

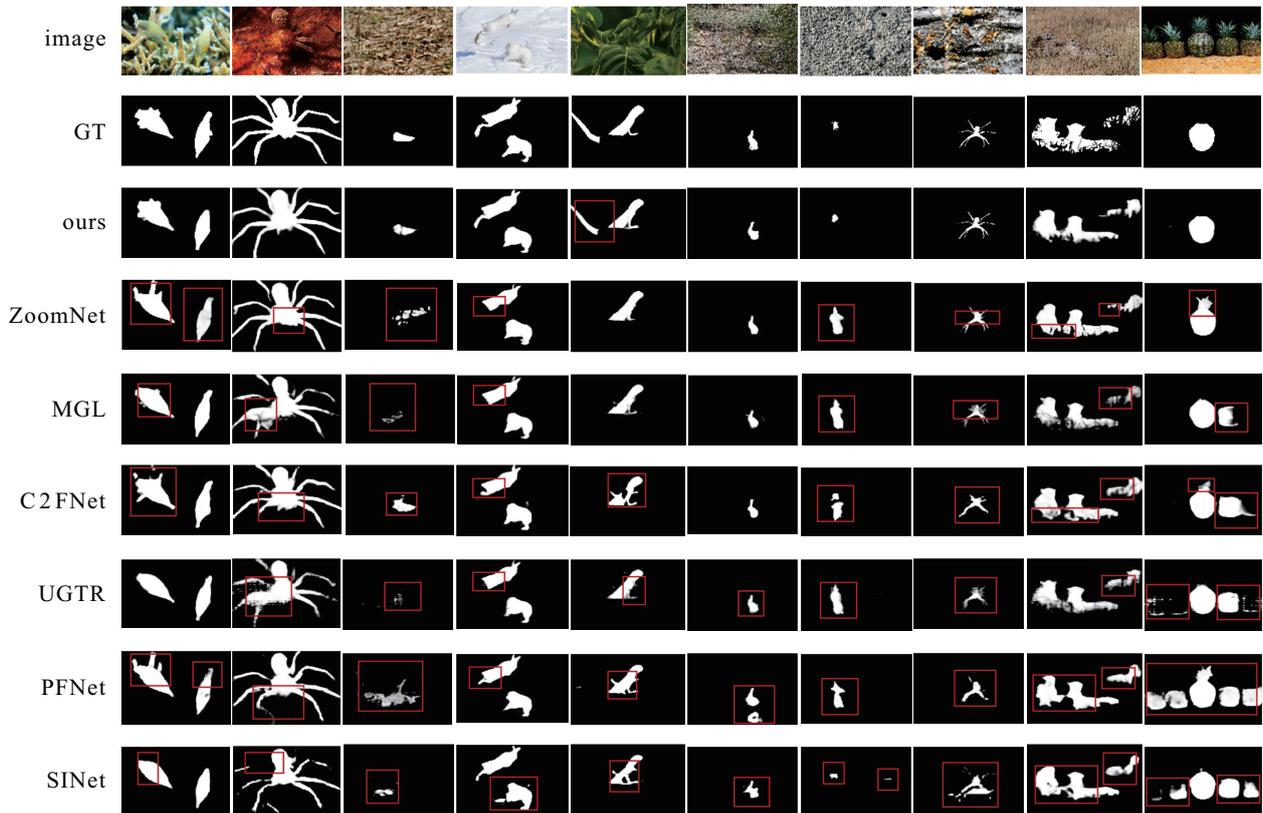


图 8 不同模型的可视化结果

表 2 消融实验分析

model	CHAMELEON				COD 10 K				CAMO				NC 4 K			
	$S_\alpha \uparrow$	$E_\theta \uparrow$	$F_\beta^w \uparrow$	MAE↓	$S_\alpha \uparrow$	$E_\theta \uparrow$	$F_\beta^w \uparrow$	MAE↓	$S_\alpha \uparrow$	$E_\theta \uparrow$	$F_\beta^w \uparrow$	MAE↓	$S_\alpha \uparrow$	$E_\theta \uparrow$	$F_\beta^w \uparrow$	MAE↓
Baseline	0.893	0.941	0.834	0.029	0.844	0.914	0.741	0.029	0.855	0.910	0.799	0.055	0.872	0.922	0.816	0.039
Baseline+BEM	0.898	0.943	0.846	0.027	0.845	0.915	0.743	0.029	0.857	0.914	0.803	0.053	0.874	0.923	0.818	0.038
Baseline+FFDM	0.900	0.944	0.848	0.026	0.847	0.916	0.744	0.028	0.859	0.915	0.805	0.052	0.875	0.924	0.819	0.037
Baseline+BEM+FFDM	0.901	0.945	0.849	0.026	0.850	0.917	0.749	0.028	0.868	0.916	0.810	0.050	0.875	0.924	0.821	0.036
Baseline+FFDM+GCAM	0.902	0.944	0.850	0.025	0.853	0.918	0.750	0.027	0.869	0.919	0.815	0.049	0.876	0.925	0.823	0.037
Baseline+BEM+FFDM+GCAM	0.903	0.947	0.851	0.024	0.858	0.920	0.764	0.026	0.871	0.927	0.823	0.048	0.879	0.926	0.825	0.036

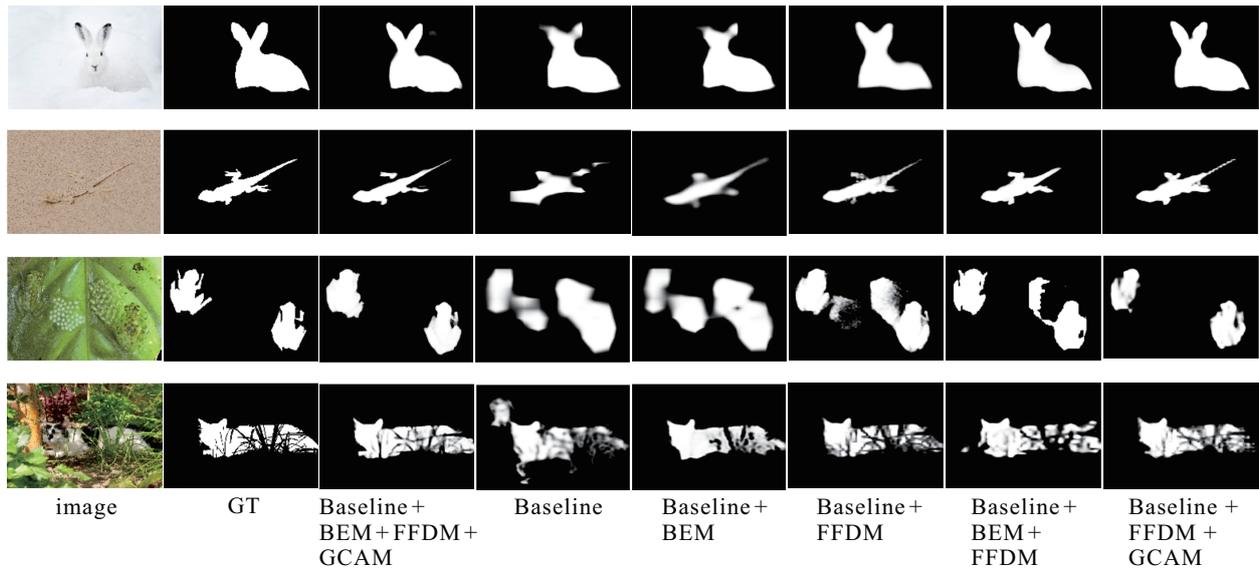


图9 消融研究的可视化结果

通过 Baseline~Baseline+BEM 和 Baseline+FFDM~Baseline+BEM+FFDM 两组模型可验证边界增强模块的有效性. 如表2所示, 相比于单一的基础网络, Baseline+BEM 在 CAMO 数据集上平均增强度量 E_{θ} 和加权 F 度量 F_{β}^w 均提高了 0.4%, 平均绝对误差 MAE 下降了 0.2%; 相比于 Baseline+FFDM, Baseline+BEM+FFDM 在 CAMO 数据集上结构度量 S_{α} 、加权 F 度量 F_{β}^w 分别提高了 0.9%、0.5%, 平均绝对误差 MAE 下降了 0.2%. 此外, 由图9可见, 利用 BEM 模块将额外的边界信息嵌入特征中, 增强了模型对边界的理解.

通过 Baseline~Baseline+FFDM 和 Baseline+BEM~Baseline+BEM+FFDM 两组模型可验证特征融合解码模块的有效性. 如表2所示, 相比于单一的基础网络, Baseline+FFDM 在 CAMO 数据集上结构度量 S_{α} 、平均增强度量 E_{θ} 、加权 F 度量 F_{β}^w 上分别提高了 0.4%、0.5%、0.6%, 平均绝对误差 MAE 下降了 0.3%; 相比于 Baseline+BEM, Baseline+BEM+FFDM 在 CAMO 数据集上结构度量 S_{α} 、加权 F 度量 F_{β}^w 上分别提高了 1.1%、0.7%, 平均绝对误差 MAE 下降了 0.3%. 因此, FFDM 模块能够提高 COD 的性能, 从而使得模型更好地定位伪装对象.

通过 Baseline+FFDM~Baseline+FFDM+GCAM 和 Baseline+BEM+FFDM~Baseline+BEM+FFDM+GCAM 两组模型可验证全局上下文聚合模块的有效性. 如表2所示, 相比于 Baseline+FFDM, Baseline+FFDM+GCAM 在 CAMO 数据集上结构度量 S_{α} 、平均增强度量 E_{θ} 、加权 F 度量 F_{β}^w 上分别提高了 1.0%、0.4%、1.0%, 平均绝对误差 MAE 下降

了 0.3%; 相比于 Baseline+BEM+FFDM, Baseline+BEM+FFDM+GCAM 在 4 个数据集上结构度量 S_{α} 、平均增强度量 E_{θ} 、加权 F 度量 F_{β}^w , 平均绝对误差 MAE 的性能最好. 如图9所示的视觉对比结果也验证了这一点, 各模块的有效性使得网络能够更好地检测伪装对象与背景间的细微差异.

3 结论

本文提出了一种全局上下文交互融合网络 (GCIFNet) 用于伪装目标检测. 首先, 利用所提出边界增强模块对边界特征进行增强; 然后, 设计了特征融合解码模块, 在该模块中利用图卷积以及特征细化模块来增强上下文信息的丰富性, 用于伪装目标定位得到粗略定位图. 此外, 通过全局上下文聚合模块来有效地融合整个网络中的上下文信息. 在 4 个公开数据集上的实验结果表明, 与其他方法相比, 所提出算法取得了较好的检测效果. 在未来的工作中, 将进一步研究 COD 上下文信息的聚合方法, 探究新的网络模型来实现更加精确的伪装目标检测.

参考文献 (References)

- [1] Shamim S, Awan M J, Zain A M, et al. Automatic COVID-19 lung infection segmentation through modified unet model[J]. Journal of Healthcare Engineering, 2022, 2022: 6566982.
- [2] Fang F, Li L Y, Gu Y, et al. A novel hybrid approach for crack detection[J]. Pattern Recognition, 2020, 107: 107474.
- [3] Chu H K, Hsu W H, Mitra N J, et al. Camouflage images[J]. ACM Transactions on Graphics, 2010, 29(4): 1-8.

- [4] Zhang X, Zhu C, Wang S, et al. A Bayesian approach to camouflaged moving object detection[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2017, 27(9): 2001-2013.
- [5] Owens A, Barnes C, Flint A, et al. Camouflaging an object from many viewpoints[C]. *IEEE Conference on Computer Vision and Pattern Recognition*. Columbus, 2014: 2782-2789.
- [6] Zhuge M C, Lu X K, Guo Y Y, et al. CubeNet: X-shape connection for camouflaged object detection[J]. *Pattern Recognition*, 2022, 127: 108644.
- [7] Hu X B, Wang S, Qin X B, et al. High-resolution iterative feedback network for camouflaged object detection[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2023, 37(1): 881-889.
- [8] 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.
- [9] Yue G H, Xiao H L, Xie H, et al. Dual-constraint coarse-to-fine network for camouflaged object detection[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2023, PP(99): 1.
- [10] Zheng D H, Zheng X C, Yang L T, et al. MFFN: Multi-view feature fusion network for camouflaged object detection[C]. *IEEE/CVF Winter Conference on Applications of Computer Vision*. Waikoloa, 2023: 6221-6231.
- [11] Xing H Z, Gao S Y, Wang Y, et al. Go closer to see better: Camouflaged object detection via object area amplification and figure-ground conversion[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2023, 33(10): 5444-5457.
- [12] He C M, Li K, Zhang Y C, et al. Camouflaged object detection with feature decomposition and edge reconstruction[C]. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Vancouver, 2023: 22046-22055.
- [13] Jia Q, Yao S L, Liu Y, et al. Segment, magnify and reiterate: Detecting camouflaged objects the hard way[C]. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. New Orleans, 2022: 4703-4712.
- [14] Wang W H, Xie E Z, Li X, et al. PVT v2: Improved baselines with pyramid vision transformer[J]. *Computational Visual Media*, 2022, 8(3): 415-424.
- [15] Yang M K, Yu K, Zhang C, et al. DenseASPP for semantic segmentation in street scenes[C]. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, 2018: 3684-3692.
- [16] 刘建骐, 闫河, 王潇棠, 等. 改进金字塔和跳跃连接的YOLOv5目标检测网络[J]. *控制与决策*, 2023, 38(6): 1730-1736.
- [17] Liu J Q, Yan H, Wang X T, et al. YOLOv5 object detection network with improved pyramid and skip connection[J]. *Control and Decision*, 2023, 38(6): 1730-1736.
- [18] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]. *Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention*. Munich, 2015: 234-241.
- [19] Zhang L, Li X, Arnab A, et al. Dual graph convolutional network for semantic segmentation[J/OL]. 2019, arXiv: 1909.06121.
- [20] Wei J, Wang S, Huang Q. F3Net: Fusion, feedback and focus for salient object detection[J/OL]. 2019, arXiv: 1911.11445.
- [21] Le Trung-Nghia, Nguyen T V, Nie Z L, et al. Anabran network for camouflaged object segmentation[J]. *Computer Vision and Image Understanding*, 2019, 184: 45-56.
- [22] Skurowski P, Abdulameer H, Błaszczuk J, et al. Animal camouflage analysis: Chameleon database[J]. *Unpublished Manuscript*, 2018, 2(6): 7.
- [23] Fan D P, Ji G P, Sun G L, et al. Camouflaged object detection[C]. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle, 2020: 2774-2784.
- [24] Lv Y Q, Zhang J, Dai Y C, et al. Simultaneously localize, segment and rank the camouflaged objects[C]. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Nashville, 2021: 11586-11596.
- [25] Fan D P, Cheng M M, Liu Y, et al. Structure-measure: A new way to evaluate foreground maps[C]. *IEEE International Conference on Computer Vision*. Venice, 2017: 4558-4567.
- [26] Fan D, Gong C, Cao Y, et al. Enhanced-alignment measure for binary foreground map evaluation[J/OL]. 2018, arXiv: 1805.10421.
- [27] Margolin R, Zelnik-Manor L, Tal A. How to evaluate foreground maps[C]. *IEEE Conference on Computer Vision and Pattern Recognition*. Columbus, 2014: 248-255.
- [28] Perazzi F, Krähenbühl P, Pritch Y, et al. Saliency filters: Contrast based filtering for salient region detection[C]. *IEEE Conference on Computer Vision and Pattern Recognition*. Providence, 2012: 733-740.
- [29] Liu J J, Hou Q B, Cheng M M, et al. A simple pooling-based design for real-time salient object detection[C]. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach, 2019:

- 3917-3926.
- [29] Zhao J X, Liu J J, Fan D P, et al. EGNet: Edge guidance network for salient object detection[C]. IEEE/CVF International Conference on Computer Vision. Seoul, 2019: 8779-8788.
- [30] Fan D P, Ji G P, Zhou T, et al. PraNet: Parallel reverse attention network for polyp segmentation[C]. International Conference on Medical Image Computing and Computer-assisted Intervention. Seattle, 2020: 263-273.
- [31] Zhu J C, Zhang X Y, Zhang S, et al. Inferring camouflaged objects by texture-aware interactive guidance network[C]. Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto, 2021: 3599-3607.
- [32] Mei H Y, Ji G P, Wei Z Q, et al. Camouflaged object segmentation with distraction mining[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, 2021: 8768-8777.
- [33] Yang F, Zhai Q, Li X, et al. Uncertainty-guided transformer reasoning for camouflaged object detection[C]. IEEE/CVF International Conference on Computer Vision. Montreal, 2021: 4126-4135.
- [34] Zhai Q, Li X, Yang F, et al. Mutual graph learning for camouflaged object detection[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, 2021: 12992-13002.
- [35] Chen G, 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.
- [36] 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.
- [37] 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.
- [38] 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]. IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, 2022: 2150-2160.
- [39] Zhu H W, Li P, Xie H R, et al. I can find you! Boundary-guided separated attention network for camouflaged object detection[C]. Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto, 2022: 3608-3616.
- [40] 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.
- [41] Huang Z, Dai H, Xiang T Z, et al. Feature shrinkage pyramid for camouflaged object detection with transformers[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, 2023: 5557-5566.

作者简介

葛斌(1973—),男,教授,博士,从事图像处理、信息安全等研究, E-mail: bge@aust.edu.cn;

陈宁洁(1997—),女,硕士生,从事图像处理、计算机视觉等研究, E-mail: chenningjie@163.com;

夏晨星(1991—),男,副教授,博士,从事图像处理、计算机视觉等研究, E-mail: cxxia@aust.edu.cn;

郑海君(1998—),男,硕士生,从事图像处理、计算机视觉等研究, E-mail: navy626@163.com;

吴涛林(1998—),男,硕士生,从事图像处理、计算机视觉等研究, E-mail: wtl19982023@163.com.