

多尺度特征融合与交互的伪装目标检测网络^①



张成^{1,2}, 刘研^{1,2}, 宋慧慧^{1,2}

¹(南京信息工程大学江苏省大数据分析技术重点实验室, 南京 210044)

²(江苏省大气环境与装备技术协同创新中心, 南京 210044)

通信作者: 宋慧慧, E-mail: songhuihui@nuist.edu.cn

摘要: 伪装目标检测是一项在复杂场景中定位和识别伪装目标的任务。目前基于神经网络的方法已初步运用,但在复杂场景下遇到干扰时,许多方法无法充分利用目标的多级特征来提取丰富的语义信息,仅依靠固定尺寸特征识别伪装目标。为解决这一问题,本文提出了一种基于多尺度特征融合交互的伪装目标检测网络。该网络包含两个创新设计:多尺度特征感知模块和双阶段邻级交互模块。前者旨在通过结合多尺度特征的方式充分捕获复杂场景中丰富的局部-全局场景对比信息。后者则是整合来自相邻层的特征以利用跨层相关性将有价值的上下文信息从编码器传输到解码器网络。本文在 CHAMELEON、CAMO-Test、COD10K-Test 这 3 个公共数据集上对提出的方法进行了评测并与当前的主流方法对比。实验结果表明,本文方法的性能超越了当前的主流方法,在各项指标上达到了优异的性能水平。

关键词: 伪装目标检测; 多尺度特征提取; 跨级特征融合; 深度学习

引用格式: 张成,刘研,宋慧慧.多尺度特征融合与交互的伪装目标检测网络.计算机系统应用,2024,33(8):90-97. <http://www.c-s-a.org.cn/1003-3254/9595.html>

Camouflaged Object Detection Network Based on Multi-scale Feature Fusion and Interaction

ZHANG Cheng^{1,2}, LIU Yan^{1,2}, SONG Hui-Hui^{1,2}

¹(Jiangsu Key Laboratory of Big Data Analysis Technology, Nanjing University of Information Science and Technology, Nanjing 210044, China)

²(Jiangsu Collaborative Innovation Center of Atmospheric Environment and Equipment Technology, Nanjing 210044, China)

Abstract: The task of camouflaged object detection involves locating and identifying camouflaged objects in complex scenes. While deep neural network-based methods have been applied to this task, many of them struggle to fully utilize multi-level features of the target for extracting rich semantic information in complex scenes with interference, often relying solely on fixed-size features to identify camouflaged objects. To address this challenge, this study proposes a camouflaged object detection network based on multi-scale and neighbor-level feature fusion. This network comprises two innovative designs: the multi-scale feature perception module and the two-stage neighbor-level interaction module. The former aims to capture rich local-global contrast information in complex scenes by combining multi-scale features. The latter integrates features from adjacent layers to exploit cross-layer correlations and transfer valuable contextual information from the encoder to the decoder network. The proposed method has been evaluated on three public datasets: CHAMELEON, CAMO-Test, and COD10K-Test, and compared with the current mainstream methods. The experimental results demonstrate that the proposed method outperforms the current mainstream methods, achieving excellent performance across all metrics.

Key words: camouflaged object detection (COD); multi-scale feature extraction; cross-level feature fusion; deep learning

① 基金项目: 国家自然科学基金 (61872189)

收稿时间: 2024-02-22; 修改时间: 2024-03-28; 采用时间: 2024-04-03; csa 在线出版时间: 2024-06-28

CNKI 网络首发时间: 2024-07-02

伪装目标检测 (camouflaged object detection, COD)^[1]旨在识别与周围环境具有相似纹理的目标. 伪装的物体大致可以分为两类, 即自然伪装物体和人工伪装物体. 自然伪装对象隐藏在背景环境中, 具有自身优势 (如颜色、形状等) 来适应环境^[2,3], 人工伪装对象往往出现在现实场景中. 它有多种应用, 如安全和监视 (如搜索和救援工作^[4])、农业 (如检测农业) 和医学成像分析^[5] (如肺部感染分割^[6]和息肉分割^[7]). 因此, COD 由于其应用和科学价值, 在医疗与工业领域备受关注, 成为计算机视觉领域的研究热点之一.

COD 与传统的显著物体检测 (salient object detection, SOD)^[8-10]在任务实现的细节上有相似之处. 任务的侧重点都关于分离图像中的前景和背景. 然而, COD 任务中需要分割的目标在视觉上与背景高度相似, 模型在预测时往往会发生边界模糊, 错判对象等问题, 而 SOD 任务的难题则是注重对于显著物体的判断, 因此前者任务在对于前景与背景的区分难度上相比后者要高很多, 因此更具挑战性. 由于伪装目标与背景之间的内在相似性较高, 在早期的机器学习时代甚至更早期的方法^[11-13]主要依赖研究者自己手工设计特征 (如纹理、颜色、强度等), 这些方法受限于早期的硬件水平具有很强的局限性, 即便是在指标上表现很好的方法缺乏泛化性.

随着深度学习在计算机视觉领域的快速发展, 现代的 COD 方法也得到了巨大进步, 成为领域内的主流. 例如, Anet^[14]利用分类网络来确定图像中是否包含伪装物体, 然后利用全卷积网络处理 COD 任务. SINet^[1]提出了使用搜索模块粗略选择伪装物体候选区域, 然后引入识别模块来进行精确的伪装物体检测. 此外, 还提出了一个用于 COD 的大规模数据集, 推动了该领域的发展并促进了更多的探索. PraNet^[5]使用并行部分解码器在高级层中聚合特征, 然后基于组合特征生成全局映射提供初始指导区域, 同时利用反向注意模块来建立区域与边界线索之间的关系. FAPNet^[15]采用边界引导与特征聚合的方法来提升模型对于复杂场景下的特征提取能力.

尽管目前主流的深度学习方法相较于传统方法在各项评价指标上都有巨大的提升, 但由于任务中最大的难题, 即前景与背景在视觉上的高度相似性, 现有方法仍容易误判目标的轮廓. 此外, 现实任务中还存在许多挑战, 例如目标大小尺度差距过大, 目标外形极度不

规则, 模型分割时出现语义分歧等. 即使目前主流的方法也难以在复杂场景中准确识别伪装物体, 因此 COD 仍有很大的改进空间. 为了解决上述提到的问题, 本文分析得出如下改进的方向.

首先, 尺度变化是 COD 的主要挑战之一, 因为从骨干网络中提取的浅层特征只能获得较小的感受野, 所以如何有效地从卷积层提取多尺度信息值得进一步探索. 其次, 目前的方法通常将多层次特征整合并输入到解码器网络中, 但忽略了来自不同编码器块的特征表示对结果的贡献. 依据上面分析得出的改进方向, 本文设计了一个基于多尺度特征融合与交互的伪装目标检测网络, 综合考虑了多尺度特征中的各种信息, 使得网络能够兼顾全局语义与局部细节. 具体而言, 主要设计了两个关键模块: 多尺度特征感知模块从伪装物体场景中全面地提取从局部到全局、从小尺度到大尺度的多级特征; 双阶段邻级交互模块则通过跨层融合的方式来有效整合多级特征. 本文的主要创新点如下.

(1) 本文设计了一种多尺度特征感知模块, 提出了一种并行特征提取模块, 通过结合多次下采样操作来获得丰富的多尺度场景信息, 然后再逐步上采样聚合解码来还原特征图的尺寸.

(2) 本文设计了一种双阶段邻级交互模块来有效地融合逐级相邻的两层特征, 并将有用信息从编码器传播到解码器网络, 这使得模型可以自适应地来平衡网络中每个编码器块提供的特征对于解码器网络的贡献.

1 模型结构

受到渐进聚合多尺度场景上下文特征的伪装目标检测^[16]启发, 本文设计了一个全新的网络结构. 图 1 展示了本文所提出的整体网络架构. 网络由两个关键模块组成: 多尺度特征感知模块和双阶段邻级交互模块. 具体而言, 首先将图像输入编码器骨干网络 (以 Res2Net-50^[17]为主干网络) 以提取多层次特征, 分别表示为 f_i ($i = 1, 2, \dots, 5$). 因此, 得到的第 1 层的特征分辨率为 $W/4 \times H/4$, 而当 $i > 1$ 时, 特征的一般分辨率为 $W/2^i \times H/2^i$. 然后, 为了降低计算复杂度, 本文将通道大小设置为 32. 然后在 f_5 上使用空洞空间金字塔池化 (atrous spatial pyramid pooling, ASPP)^[18]来获取含有粗略物体定位信息的全局语义特征. 并且将上面的 f_1, f_2, f_3, f_4 送入多尺度特征感知模块提取局部细节丰富的多尺度特征, 以捕捉伪装对象的尺度变化. 之后, 将聚合后的特征输

入到双阶段邻级交互模块中, 通过融合机制和交叉注意力以有效地整合跨层级特征并将融合后的特征传播

到解码器网络. 技术细节将在第 1.1 节、第 1.2 节、第 1.3 节中分别展示.

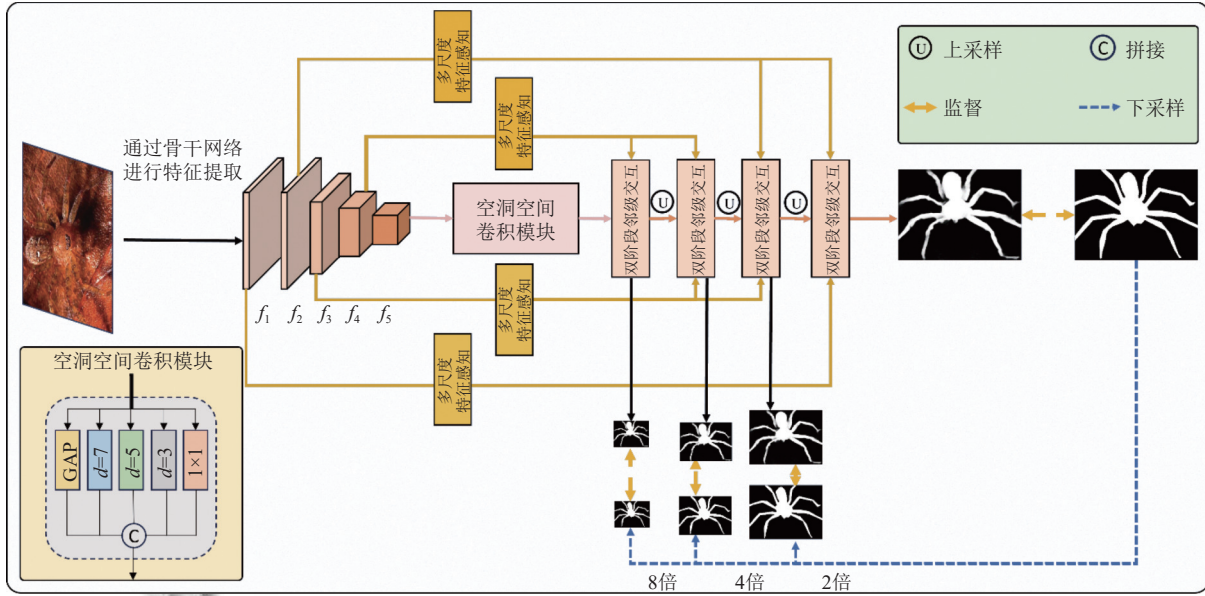


图1 本文方法的总体网络结构图

1.1 空洞空间金字塔池化 (ASPP)

经过 5 层残差块编码后, 会得到具有强语义表征能力的特征 f_5 , 该特征主要反映了潜在伪装物体在图像中的大致位置信息. 为了充分利用这种特性, 本文采用了 ASPP 结合骨干网络提取的最深层特征来处理目标的全局多尺度语义信息. 具体而言, 如图 1 所示, 本文使用 ASPP 处理经过通道压缩处理后的特征 $f_5 \in \mathbb{R}^{32 \times H/32 \times W/32}$, 在 ASPP 中, f_5 特征会同时被 5 个并行分支 $l_i, i = 1, 2, \dots, 5$ 处理. 其中, l_1 是 1×1 的卷积操作, l_5 是全局平均池化操作, 而中间的 3 个分支则分别是空洞率为 $d = 3, 5, 7$ 的 3×3 的卷积操作.

最后, 将各个分支的输出进行通道维度拼接, 再通过 1×1 的卷积操作将拼接后的特征通道降维到与输入特征相同的大小.

1.2 多尺度特征感知模块

因为每个卷积层只能处理一个特殊的尺度, 所以尺度变化是 COD 任务中的主要挑战之一. 受到 U-Net^[19] 在复杂医疗场景下的成功运用的启发, 本文设计了多尺度特征感知模块来实现从单层捕获多尺度信息来表征对象尺度的变化.

将除 f_5 外每一层特征, 即 f_1, f_2, f_3, f_4 , 在进行通道压缩后分别送入多尺度特征感知模块, 下面以特征

f_2 为例. 首先, 将输入特征分别经过 3 次特征生成模块和最大池化重复进行多尺度特征提取和下采样操作得到特征 $f_2^3 \in \mathbb{R}^{32 \times H/32 \times W/32}$. 再通过双线性插值使得特征大小恢复到和输入相同. 然后通过残差连接的方式聚合之前下采样的多尺度特征, 并使用单个 3×3 卷积操作细化提取的特征. 最后得到解码后的输出特征 $H_2 \in \mathbb{R}^{32 \times H/4 \times W/4}$. 详细的流程公式如下:

$$\begin{cases} f_2^1 = \text{Harr}(\text{FG}(f_2)) \\ f_2^2 = \text{Harr}(\text{FG}(f_2^1)) \\ f_2^3 = \text{Harr}(\text{FG}(f_2^2)) \end{cases} \quad (1)$$

式 (1) 介绍了多尺度特征感知模块中的下采样流程, 其中 $\text{Harr}(\cdot)$ 表示 Harr 小波下采样操作^[20], 用来下采样并压缩特征减小计算量. 该下采样方法利用 Harr 小波变换在尽可能多地保留特征图信息的同时降低特征图的空间分辨率. $\text{FG}(\cdot)$ 表示特征生成操作, 即将特征输入送进特征生成模块.

$$\begin{cases} H_2^1 = \text{Conv}_3(\text{FG}(f_2^2) + \text{up}(\text{Conv}_3(f_2^3))) \\ H_2^2 = \text{Conv}_3(\text{FG}(f_2^1) + \text{up}(H_2^1)) \\ H_2 = \text{Conv}_3(\text{FG}(f_2) + \text{up}(H_2^2)) \end{cases} \quad (2)$$

式 (2) 介绍了多尺度特征感知模块中的上采样流程, 其中 $\text{Conv}_3(\cdot)$ 表示将特征送入单个 3×3 卷积, 通过

运用填充的方式使得通过卷积的输出在大小上和输入相同. $up(\cdot)$ 表示上采样操作,用于将下采样过程中得到的特征逐步恢复到模块的输入大小,方便进行下一步的操作.

图2中的特征生成模块通过在输入特征 f_n 上进行3次独立的 1×1 卷积操作和一次平均池化后的 1×1 卷积操作,得到了3个大小相同的特征 f_n^1, f_n^2 和 f_n^3 .对 f_n 先进行全局平均池化然后再进行一次 1×1 卷积操作,得到特征 f_n^4 .然后在 f_n^2 和 f_n^3 上使用 3×3 大小的卷积操作,得到 $Conv_3(f_n^2)$ 和 $Conv_3(f_n^3)$,再把 $Conv_3(f_n^3)$ 送入 5×5

卷积,得到输出结果 $Conv_5(Conv_3(f_n^3))$,把上述各个分支的结果进行拼接,即将 $f_n^1, Conv_3(f_n^2), Conv_5(Conv_3(f_n^3)), f_n^4$ 进行拼接,得到了最终结果.

1.3 双阶段邻级交互模块

粗暴地将所有特征传递给解码器网络而不去关注不同层次的特征的贡献是不应该的.于是本文设计了这个双阶段邻级交互模块.

如图3所示,本文以两个多尺度特征感知模块的输出特征 f_n^{aware} 与 f_{n+1}^{aware} 为例(f_{n+1}^{aware} 经过上采样操作).

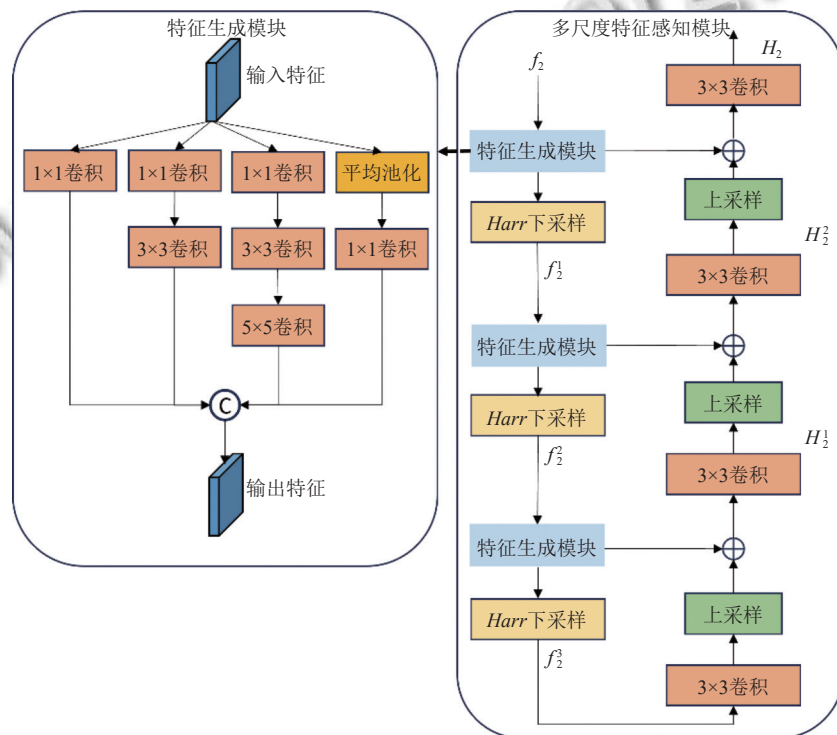


图2 多尺度特征感知模块结构图

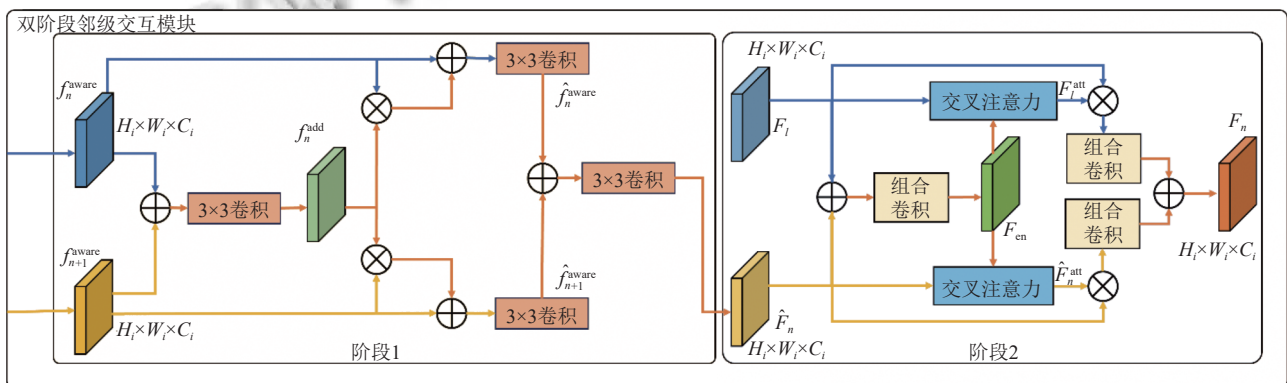


图3 双阶段邻级交互模块结构图

首先通过加法操作融合这两个特征,然后将结果输入带有 Sigmoid 激活函数的 3×3 卷积层,可以得到融合后的特征映射,且有:

$$f_n^{\text{add}} = \sigma(\text{Conv}_3(f_n^{\text{aware}} \oplus f_{n+1}^{\text{aware}})) \quad (3)$$

其中, $\sigma(\cdot)$ 表示 Sigmoid 激活函数. 此外,为了保留每个特征的原始信息,采用残差连接将增强的特征与其原始特征相结合. 因此,得到的特征如下:

$$\begin{cases} \hat{f}_n^{\text{aware}} = \text{Conv}_3(f_n^{\text{aware}} \otimes f_n^{\text{add}} + f_n^{\text{aware}}) \\ \hat{f}_{n+1}^{\text{aware}} = \text{Conv}_3(f_{n+1}^{\text{aware}} \otimes f_n^{\text{add}} + f_{n+1}^{\text{aware}}) \end{cases} \quad (4)$$

$$\hat{F}_n = \text{Conv}_3(\hat{f}_n^{\text{aware}} \oplus \hat{f}_{n+1}^{\text{aware}}) \quad (5)$$

\hat{F}_n 融合了两个相邻层级的多尺度特征感知模块的输出,不仅可以增强邻层级特征表示来提升模型对于上下文相关性的理解,同时使用了残差连接的融合方式使得 \hat{F}_n 保留了每个特征的原始信息.

在之后的流程中,将 \hat{F}_n 与相邻层级的双阶段邻级交互模块的低级输出特征 F_l 送入下一步操作(此处 F_l 已经经过上采样). 首先将这两个特征相加,然后通过组合卷积模块(此处为 3×3 卷积和批归一化(BN)和 ReLU 激活函数的组合),得到增强后的特征 F_{en} 如下:

$$F_{\text{en}} = \text{ConvCBR}(\hat{F}_n \oplus F_l) \quad (6)$$

接着将 F_{en} 分别与两个输入做交叉注意力操作^[17],得到 F_l^{att} 与 \hat{F}_n^{att} ,公式表达如下:

$$\begin{cases} F_l^{\text{att}} = \text{CA}(F_l, F_{\text{en}}) \\ \hat{F}_n^{\text{att}} = \text{CA}(\hat{F}_n, F_{\text{en}}) \end{cases} \quad (7)$$

其中, $\text{CA}(\cdot)$ 表示交叉注意力操作. 之后再与两个原输入分别相乘,得到的结果分别通过组合卷积再次相加,最后得到模块的最终输出 F_n 如下:

$$F_n = \text{ConvCBR}(F_l^{\text{att}} \otimes F_l) \oplus \text{ConvCBR}(\hat{F}_n^{\text{att}} \otimes \hat{F}_n) \quad (8)$$

通过交叉注意力操作替代先前直接简单相加或者相乘的操作来聚合邻级多尺度特征,可以在不产生多余噪声的前提下充分获得不同特征之间的聚合信息,而且减少了融合步骤,避免了多次特征融合产生的特征冗余. 经过残差连接和组合卷积块得到的最终输出具有被加强的全局语义的同时还具有相对精细的局部细节.

1.4 损失函数

二元交叉熵是分割任务中广泛应用的损失函数之一,但是其在独立计算时仅考虑了每个像素的损失而忽略了图像的全局结构. 结合标准交并比(IoU)与显著目标

检测中的加权交并比^[21],本文设计了一个如下损失函数:

$$L_{\text{total}} = \sum_{i=1}^4 \lambda_i L_{\text{det}}(G_i^{\text{down}}, S_i) \quad (9)$$

其中, $L_{\text{det}} = L_{\text{IoU}}^w + L_{\text{BCE}}^w$, L_{IoU}^w 与 L_{BCE}^w 分别表示全局和局部限制的加权 IoU 损失和 BCE 损失, λ_i 表示不同损失函数的权重系数. L_{IoU}^w 可以增加硬像素的权重以突出它们的重要性, L_{BCE}^w 更关注硬像素,而不是平等地对待所有像素,两者结合可以从全局和局部角度来进行有效约束. 对真实值进行不同尺度的下采样(即 G_i^{down})使得与 4 个输出图(即 S_i)大小相同,并计算它们之间的交并比,得到最后的结果 L_{total} .

2 实验结果及分析

2.1 实验配置

本文网络结构是端到端的模型,使用两块 NVIDIA 2080Ti GPU,在 PyTorch 1.2.0 框架上进行训练. 使用了用骨干网络(Res2Net-50),该网络已经在 ImageNet^[22]上进行了预训练,在训练图像大小为 352×352 ,使用额外的数据增强策略包括随机翻转、裁剪旋转以进行数据扩充. 测试时保持图像大小为 352×352 . 使用 Adam 算法^[23]对训练过程进行参数优化. 初始学习率设置为 $1\text{E}-4$,批大小设置为 20,训练周期为 200,每训练 50 个周期学习率将按照 0.1 倍率衰减一次.

在测试阶段本文不对模型进行任何调整,仅调整输入图片大小再送入模型进行推理,对于推理的结果使用双线性插值法将模型输出调整至原图大小.

2.2 数据集

本文在伪装目标检测任务的 3 个公共数据集上进行实验. CHAMELEON^[1]是通过 Google 搜索引擎收集的数据集,包含 76 张伪装动物的图像用于测试. CAMO-Test^[14]有 1250 张图像,8 个类别,其中 1000 张图像用于训练,其余 250 张图像用于测试. COD10K-Test^[1]是目前最大的伪装目标数据集,具有高质量的像素级注释. 该数据集共有 10000 张伪装图像,其中 6000 张图像用于训练,4000 张用于测试.

2.3 实验结果

2.3.1 不同实验方法评估

本文在伪装目标检测的 3 个数据集上进行对比实验,对比实验所采用的评价指标如下:平均绝对误差(MAE)^[24]、平均 E-measure (E_ϕ)^[25]、结构度量 (S_α)^[26]

以及平均 F -measure (F_β)^[27]值作为评价指标。

本文选取了 11 个具有代表性的基于深度学习的方法与本文所提出的方法进行了对比. 在表 1 展示了所有方法的指标结果, 其中黑色加粗代表最好的性能, 下划线代表第 2 名的性能, \uparrow 表示该指标越大越好, \downarrow 表示该指标越小越好. 可见本文的方法在 3 个数据集上都表现优异, 证明了本文方法的有效性.

本文方法在最大数据集 COD10K-Test 上各项指标的性能数据都超越了其他模型, 4 个指标相对于第 2 名分别提升了 0.2%、0.2%、0.1%、0.3%. 在数据量

相对较少的数据集 CHAMELEON 上 E_ϕ 比第 2 名低了 0.1%, 但在另外 3 项指标上分别提升了 0.3%、0.2%、0.2%. 在数据集 CAMO-Test 上平均误差比第 2 名低 0.3%, 但在另外 3 项指标上分别提升了 0.6%、0.7%、0.4%. 本文方法之所以效果较好, 主要由于多尺度特征感知模块捕捉到不同尺度的特征使得模型获得了丰富的局部到全局信息, 同时双阶段邻级交互模块高效地融合了相邻层级之间的多尺度特征使得预测结果更加准确. 不同方法的定性比较如图 4 所示, 可见本文方法相较其他方法拥有更加准确的边界和定位.

表 1 CHAMELEON、CAMO-Test 和 COD10K-Test 测试集上的对比实验

方法	CHAMELEON				CAMO-Test				COD10K-Test			
	$S_\alpha \uparrow$	$E_\phi \uparrow$	$F_\beta \uparrow$	$MAE \downarrow$	$S_\alpha \uparrow$	$E_\phi \uparrow$	$F_\beta \uparrow$	$MAE \downarrow$	$S_\alpha \uparrow$	$E_\phi \uparrow$	$F_\beta \uparrow$	$MAE \downarrow$
FPN ^[28]	0.794	0.783	0.648	0.075	0.684	0.677	0.676	0.131	0.697	0.691	0.481	0.075
PSPNet ^[29]	0.773	0.758	0.630	0.085	0.663	0.659	0.520	0.139	0.678	0.680	0.457	0.080
PiCANet ^[9]	0.769	0.749	0.618	0.085	0.609	0.584	0.573	0.156	0.649	0.643	0.322	0.090
UNet++ ^[30]	0.695	0.762	0.557	0.094	0.599	0.653	0.460	0.149	0.623	0.672	0.408	0.086
BASNet ^[10]	0.687	0.721	0.795	0.118	0.618	0.661	0.503	0.159	0.634	0.678	0.486	0.105
CPD ^[31]	0.853	0.866	0.771	0.052	0.726	0.802	0.618	0.115	0.747	0.770	0.595	0.059
EGNet ^[32]	0.848	0.870	0.702	0.050	0.732	0.768	0.670	0.104	0.737	0.779	0.582	0.056
PraNet ^[5]	0.860	0.898	0.789	0.044	0.769	0.833	0.710	0.094	0.789	0.839	0.671	0.045
SINet ^[11]	0.869	0.891	0.827	0.044	0.751	0.771	0.702	0.100	0.771	0.806	0.679	0.051
LSR ^[33]	<u>0.890</u>	0.935	<u>0.841</u>	<u>0.031</u>	<u>0.787</u>	0.838	0.744	0.080	<u>0.804</u>	0.880	<u>0.715</u>	<u>0.037</u>
PFNet ^[34]	0.882	0.931	0.828	0.033	0.782	<u>0.842</u>	<u>0.746</u>	0.085	0.800	<u>0.877</u>	0.701	0.040
Ours	0.893	<u>0.934</u>	0.843	0.029	0.793	0.849	0.750	<u>0.083</u>	0.806	0.879	0.716	0.034

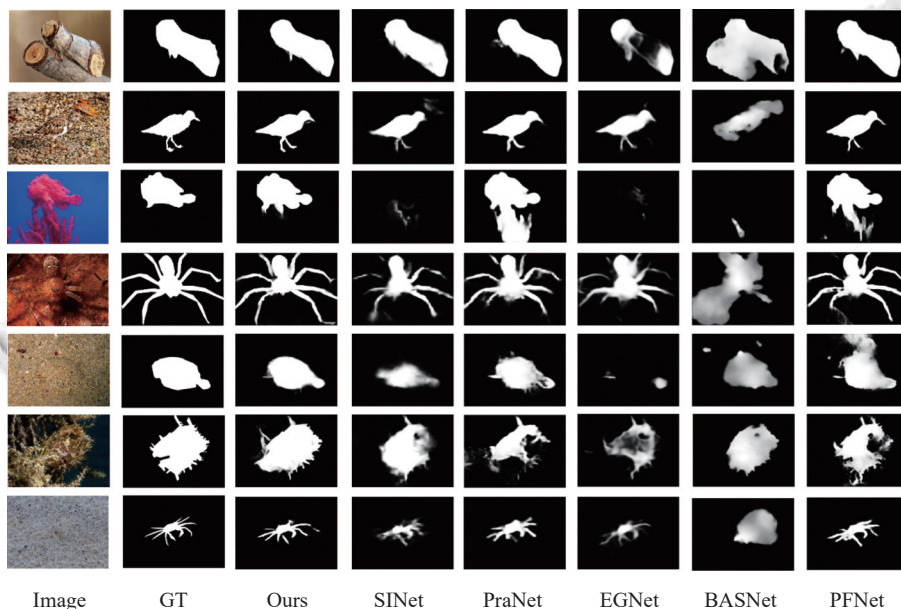


图 4 不同方法的定性比较可视化图

2.3.2 消融实验分析

为了证明本文所提出的方法中的各个模块有效性,

如表 2 所示, 本文在 CHAMELEON 数据集上进行了消融实验.

表2 在 CHAMELEON 数据集上的消融实验

实验	ASPP	多尺度 特征感知	双阶段 邻级交互	CHAMELEON			
				$S_{\alpha} \uparrow$	$E_{\phi} \uparrow$	$F_{\beta} \uparrow$	MAE \downarrow
(1)	—	—	—	0.842	0.866	0.768	0.071
(2)	√	—	—	0.852	0.873	0.775	0.066
(3)	√	√	—	0.870	0.885	0.803	0.045
(4)	—	√	√	0.879	0.923	0.823	0.032
(5)	√	—	√	0.868	0.890	0.806	0.043
(6)	√	√	√	0.893	0.934	0.843	0.029

ASPP 的有效性: ASPP 结合骨干网络提取的最深层特征, 通过全局池化增强了全局语义. 从实验 (1) 和 (2) 中可见, 加入 ASPP 后各项指标获得了提升.

多尺度特征感知模块的有效性: 多尺度特征感知模块用于提取局部细节丰富的多尺度特征. 从实验 (2) 和 (3) 可以看出, 在 ASPP 模块的基础上加入多尺度特征感知模块使得指标获得了一定的提升.

双阶段邻级交互模块的有效性: 双阶段邻级交互模块用于将多尺度特征感知模块提取的多尺度特征进行融合. 从实验 (4) 和 (5) 中可以看出, 在加入了该模块后, 整个模型的性能有了较大的提升, 说明该模块有效地集成了相邻层级的特征.

3 结束语

本文提出了一个基于多尺度特征融合与交互的网络来进行伪装目标检测. 其中, 网络的多尺度特征感知模块使得网络能够从原始图像中获得更加丰富的多尺度场景的全局与局部信息; 双阶段邻级交互模块则有效地融合了相邻层级的特征, 并将它们与编码器中得到的上下文信息一起传递到解码器网络. 本文所提出的网络在伪装目标检测的 3 个公共数据集上进行了全面实验, 结果显示该网络能够显著提升检测的速度与准确度, 相较于现有方法具有更高的性能表现.

参考文献

- Fan DP, Ji GP, Sun GL, *et al.* Camouflaged object detection. Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 2777–2787.
- Li S, Florencio D, Li WQ, *et al.* A fusion framework for camouflaged moving foreground detection in the wavelet domain. IEEE Transactions on Image Processing, 2018, 27(8): 3918–3930. [doi: 10.1109/TIP.2018.2828329]
- Chen G, Liu SJ, Sun YJ, *et al.* Camouflaged object detection via context-aware cross-level fusion. IEEE Transactions on

- Circuits and Systems for Video Technology, 2022, 32(10): 6981–6993. [doi: 10.1109/TCSVT.2022.3178173]
- Mishra B, Garg D, Narang P, *et al.* Drone-surveillance for search and rescue in natural disaster. Computer Communications, 2020, 156: 1–10. [doi: 10.1016/j.comcom.2020.03.012]
- Fan DP, Ji GP, Zhou T, *et al.* PraNet: Parallel reverse attention network for polyp segmentation. Proceedings of the 23rd Medical Image Computing and Computer Assisted Intervention. Lima: Springer, 2020. 263–273.
- Fan DP, Zhou T, Ji GP, *et al.* Inf-Net: Automatic COVID-19 lung infection segmentation from CT images. IEEE Transactions on Medical Imaging, 2020, 39(8): 2626–2637. [doi: 10.1109/TMI.2020.2996645]
- Qiu LT, Wu XJ, Yu ZY. A high-efficiency fully convolutional networks for pixel-wise surface defect detection. IEEE Access, 2019, 7: 15884–15893. [doi: 10.1109/ACCESS.2019.2894420]
- Li TP, Song HH, Zhang KH, *et al.* Recurrent reverse attention guided residual learning for saliency object detection. Neurocomputing, 2020, 389: 170–178. [doi: 10.1016/j.neucom.2019.12.109]
- Liu N, Han JW, Yang MH. PiCANet: Learning pixel-wise contextual attention for saliency detection. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 3089–3098.
- Qin XB, Zhang ZC, Huang CY, *et al.* BASNet: Boundary-aware salient object detection. Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019. 7471–7481.
- Bhajantri NU, Nagabhusan P. Camouflage defect identification: A novel approach. Proceedings of the 9th International Conference on Information Technology. Bhubaneswar: IEEE, 2006. 145–148.
- Zhang X, Zhu C, Wang S, *et al.* A Bayesian approach to camouflaged moving object detection. IEEE Transactions on Circuits and Systems for Video Technology, 2017, 27(9): 2001–2013. [doi: 10.1109/TCSVT.2016.2555719]
- Pan YX, Chen YW, Fu Q, *et al.* Study on the camouflaged target detection method based on 3D convexity. Modern Applied Science, 2011, 5(4): 152–157.
- Le TN, Nguyen TV, Nie ZL, *et al.* Anabran network for camouflaged object segmentation. Computer Vision and Image Understanding, 2019, 184: 45–56. [doi: 10.1016/j.cviu.2019.04.006]
- Zhou T, Zhou Y, Gong C, *et al.* Feature aggregation and

- propagation network for camouflaged object detection. *IEEE Transactions on Image Processing*, 2022, 31: 7036–7047. [doi: [10.1109/TIP.2022.3217695](https://doi.org/10.1109/TIP.2022.3217695)]
- 16 刘研, 张开华, 樊佳庆, 等. 渐进聚合多尺度场景上下文特征的伪装物体检测. *计算机学报*, 2022, 45(12): 2637–2651. [doi: [10.11897/SP.J.1016.2022.02637](https://doi.org/10.11897/SP.J.1016.2022.02637)]
- 17 Gao SH, Cheng MM, Zhao K, *et al.* Res2Net: A new multi-scale backbone architecture. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, 43(2): 652–662. [doi: [10.1109/TPAMI.2019.2938758](https://doi.org/10.1109/TPAMI.2019.2938758)]
- 18 Chen LC, Papandreou G, Schroff F, *et al.* Rethinking atrous convolution for semantic image segmentation. *arXiv:1706.05587*, 2017.
- 19 Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. *Proceedings of the 18th Medical Image Computing and Computer-assisted Intervention*. Munich: Springer, 2015. 234–241.
- 20 Xu GP, Liao WT, Zhang X, *et al.* Haar wavelet downsampling: A simple but effective downsampling module for semantic segmentation. *Pattern Recognition*, 2023, 143: 109819. [doi: [10.1016/j.patcog.2023.109819](https://doi.org/10.1016/j.patcog.2023.109819)]
- 21 Wei J, Wang SH, Huang QM. F³Net: Fusion, feedback and focus for salient object detection. *Proceedings of the 34th AAAI Conference on Artificial Intelligence*. New York: AAAI Press, 2020. 12321–12328.
- 22 Russakovsky O, Deng J, Su H, *et al.* ImageNet large scale visual recognition challenge. *International Journal of Computer Vision*, 2015, 115(3): 211–252. [doi: [10.1007/s11263-015-0816-y](https://doi.org/10.1007/s11263-015-0816-y)]
- 23 Kingma D, Ba J. Adam: A method for stochastic optimization. *arXiv:1412.6980*, 2014.
- 24 Perazzi F, Krähenbühl P, Pritch Y, *et al.* Saliency filters: Contrast based filtering for salient region detection. *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition*. Providence: IEEE, 2012. 733–740.
- 25 Fan DP, Gong C, Cao Y, *et al.* Enhanced-alignment measure for binary foreground map evaluation. *Proceedings of the 27th International Joint Conference on Artificial Intelligence*. Stockholm: IJCAI.org, 2018.
- 26 Fan DP, Cheng MM, Liu Y, *et al.* Structure-measure: A new way to evaluate foreground maps. *Proceedings of the 2017 IEEE International Conference on Computer Vision*. Venice: IEEE, 2017. 4548–4557.
- 27 Achanta R, Hemami S, Estrada F, *et al.* Frequency-tuned salient region detection. *Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition*. Miami: IEEE, 2009. 1597–1604.
- 28 Lin TY, Dollár P, Girshick R, *et al.* Feature pyramid networks for object detection. *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu: IEEE, 2017. 2117–2125.
- 29 Zhao HS, Shi JP, Qi XJ, *et al.* Pyramid scene parsing network. *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu: IEEE, 2017. 6230–6239.
- 30 Zhou ZW, Siddiquee MR, Tajbakhsh N, *et al.* UNet++: Redesigning skip connections to exploit multiscale features in image segmentation. *IEEE Transactions on Medical Imaging*, 2020, 39(6): 1856–1867. [doi: [10.1109/TMI.2019.2959609](https://doi.org/10.1109/TMI.2019.2959609)]
- 31 Wu Z, Su L, Huang QM. Cascaded partial decoder for fast and accurate salient object detection. *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach: IEEE, 2019. 3902–3911.
- 32 Zhao JX, Liu JJ, Fan DP, *et al.* EGNet: Edge guidance network for salient object detection. *Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision*. Seoul: IEEE, 2019. 8778–8787.
- 33 Lv YQ, Zhang J, Dai YC, *et al.* Simultaneously localize, segment and rank the camouflaged objects. *Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Nashville: IEEE, 2021. 11591–11601.
- 34 Mei HY, Ji GP, Wei ZQ, *et al.* Camouflaged object segmentation with distraction mining. *Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Nashville: IEEE, 2021. 8772–8781.

(校对责编: 孙君艳)