

融合上下文感知和背景探索的伪装目标检测方法

陈世洁¹ 范李平² 余肖生¹ 王东娟¹

(1. 三峡大学计算机与信息学院 宜昌 443002; 2. 国网宜昌供电公司 宜昌 443000)

摘要:伪装目标检测(camouflaged object detection, COD)旨在检测出与周围环境高度相似的伪装目标。针对目前 COD 方法中检测结果不完整、边缘细节模糊的问题,提出了一种融合上下文感知和背景探索(CABENet)的伪装目标检测模型。首先,该模型利用 Swin-Transformer 模型作为骨干网络,在多个尺度上提取全局上下文信息;其次,利用提出的注意力层级上下文感知模块扩大感受野,并从通道和空间两个维度增强网络的特征提取能力,再通过全连接解码器捕获隐藏对象的粗略位置图;最后,通过融合注意力机制的背景探索模块从背景信息中挖掘目标的边缘线索,加强伪装目标边缘特征的提取。在 CHAMELEON、CAMO 以及 COD10K 数据集上的实验结果表明,该方法在 4 个评估指标上的性能优于其他 10 个具有代表性的模型,在 COD10K 数据集上,平均绝对误差降至了 0.026。

关键词:伪装目标检测;上下文感知;注意力机制;背景探索

中图分类号: TP391 **文献标识码:** A **国家标准学科分类代码:** 510.4050

Camouflage object detection method integrating context awareness and background exploration

Chen Shijie¹ Fan Liping² Yu Xiaosheng¹ Wang Dongjuan¹

(1. College of Computer and Information, China Three Gorges University, Yichang 443000, China;

2. State Grid Yichang Electric Power Supply Company, Yichang 443000, China)

Abstract: Camouflaged object detection (COD) aims to identify objects that are highly similar to their surrounding environment. To address the issues of incomplete detection results and blurred edge details in current COD methods, a camouflaged object detection method that integrates context awareness and background exploration (CABENet) is proposed. Firstly, the model employs the Swin-Transformer as the backbone network to extract global context information at multiple scales. Secondly, it utilizes a proposed attention-based hierarchical context-aware module to enlarge the receptive field, enhancing feature extraction capabilities from both channel and spatial dimensions, followed by a fully connected decoder to capture coarse position maps of hidden objects. Lastly, by integrating an attention mechanism, the background exploration module explores edge clues from background information, enhancing the extraction of edge features of camouflaged objects. Experimental results on the CHAMELEON, CAMO, and COD10K datasets demonstrate that this method outperforms ten representative models on four evaluation metrics. On the COD10K dataset, the mean absolute error (MAE) is reduced to 0.026.

Keywords: camouflaged object detection; context awareness; attention mechanism; background exploration

0 引言

伪装是自然界中一种常见的防御机制,某些物种利用其具有与周边环境相似外观的特点,来保护自己免受捕食者的伤害。伪装目标检测(camouflaged object detection, COD)旨在检测出隐藏在背景中的伪装目标。近年来,

COD 技术已在医学(如肺部感染分割^[1])、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])等多个领域受到了越来越广泛的关注,有力地推动了这些领域的智能化发展,展现出良好的发展前景。

传统的 COD 方法主要依赖于基于手工设计的低级特征,如颜色^[4]、纹理^[5]、光流^[6]等直接视觉特征。当面临前

景与背景对比度极低的复杂场景时,通常存在手工提取特征耗时、迁移性较差、检测性能较低等问题。为了解决这些问题,近年来研究人员提出了多种基于深度学习的伪装目标检测模型。Fan等^[7]提出的目标搜索识别网络(SINet),通过搜索模块扩大感受野获取更丰富的上下文信息,并利用识别模块消除不相关特征的干扰。此后,他们提出的SINetV2模型^[8],通过分组引导的反转注意力模块挖掘伪装目标的边缘信息,生成精细预测,但在分组引导的过程中可能会存在特征表征分布不均匀的情况。Mei等^[9]构建了由定位模块和聚焦模块组成的COD网络PFNet,该网络首先从全局角度定位潜在目标,然后通过聚焦于模糊区域来逐步预测伪装目标,但在聚焦过程中当前特征与预测先验直接相乘可能会导致特征混淆的问题。Pang等^[10]设计了一种混合比例的三重网络ZoomNet,该网络使用三重结构提取特征,然后利用尺度集成单元来筛选和聚合特定比例的特征,并利用分层混合尺度单元进一步增强混合比例特征,但是这种多元信息策略大幅度增加了网络的计算复杂度。孙备等^[11]提出一种多尺度互交叉注意力改进的单无人机对地伪装目标检测定位模型,该模型对多尺度金字塔进行优化,提升对伪装目标的边界区分能力。Zhang等^[12]提出的注意力引导的语义和边界交互网络采用对比学习的方式将伪装对象与背景分离,从而获得伪装目标的确切位置。尽管上述方法在COD领域已经取得了一定进展,但由于伪装对象与背景高度相似,并且在边缘上模糊不清,导致无法检测出更完整的伪装对象和更准确的边缘细节。

针对以上问题,本文提出了一种融合上下文感知和背景探索的伪装目标检测方法,该方法使用Swin-Transformer模型作为骨干网络在多个尺度上提取全局上下文信息,然后设计了注意力联级上下文感知模块,通过并行多分支混合卷积模式来扩大感受野,增强全局上下文信息特征,并利用空间注意力和通道注意力机制提高对目标区域的感知能力,其次,使用全连接解码器模块来定位伪装对象,最后,设计了融合注意力机制的背景探索模块,利用自注意力机制将提取的边缘先验与目标特征逐层融合,通过关注图像中的背景区域来加强对伪装对象边缘特征的提取。在CHAMELEON、CAMO以及COD10K数据集上的实验结果表明,该方法在4个评估指标上的性能优于其他10个具有代表性的模型。

1 相关工作

1.1 视觉Transformer网络

在COD领域,常用的主干网络是卷积神经网络(convolutional neural network, CNN)系列网络。与传统的CNN相比,Transformer模型更具有全局感知力,其在处理图像时能够捕捉到更大范围的上下文信息,而不仅仅是局部特征。且由于Transformer的自注意力机制,它能够更轻松地捕捉输入序列中的长距离依赖关系,而不受传统

CNN对于感受野的限制。Carion等^[13]构建了一种新的物体检测框架(detection transformer, DETR),第1次将Transformer应用于目标检测领域。Dosovitskiy等^[14]提出了ViT(visual transformer)模型,ViT只能提取到单一尺度特征,而在语义分割任务上,多尺度的特征非常重要。Wang等^[15]提出了一种能够提取到多尺度特征的金字塔视觉Transformer主干网络PVT,该模型由4个stage组成,用于生成不同尺度的特征,PVT继承了CNN和Transformer的优点,具有更强大的特征提取能力。赵云亮等^[16]提出将CNN的局部特征提取能力与Transformer的全局建模能力有效融合,在提高识别准确率的同时降低计算成本。王琪等^[17]提出基于预过滤注意力模块的Transformer检测模型,该模块以目标点为参照,提取目标点附近部分特征点进行交互,节省训练时长并提高检测精度。Esser等^[18]构建了VQGAN(vector quantized generative adversarial network)模型,其结合了Transformer和CNN的优势,是第1个由语义引导生成百万像素图像的Transformer架构。

1.2 特征增强

现有的大部分COD方法,首先采用卷积神经网络或视觉Transformer网络提取特征,然后采用注意力机制、多尺度特征融合等策略来进一步增强特征,从而提升伪装目标检测性能。Liu等^[19]提出的感受野模块(RFB),通过在Inception架构上加入空洞卷积,从而增强网络的特征提取能力。Wu等^[20]使用搜索注意力和部分解码器组件(partial decoder component, PDC)对粗糙区域进行细化。Wang等^[21]提出的D2CNet,在PDC的基础上引入整体注意力(holistic attention)、残差注意力机制来增强特征。Sun等^[22]提出的C²F-Net模型,利用多尺度通道注意力引导来聚合跨层次特征,同时关注全局和局部信息,从而实现多尺度特征增强提升目标检测性能。Zhang等^[23]提出的BASNet模型,结合了空间注意力机制和通道注意力机制,以增强特征的表征能力。周兴超等^[24]提出一种高分辨率边缘特征增强网络,对多个尺度的特征图分别进行特征增强,有效解决了定位边界区域不准确问题。

2 融合上下文感知和背景探索的伪装目标检测方法

为了解决COD中伪装对象的纹理相似性和边界模糊导致检测结果不佳的问题,本文提出了融合上下文感知和背景探索的伪装目标检测方法,利用上下文信息和背景信息提供的丰富线索区分伪装目标和背景,更好地定位伪装对象。该模型包括Swin-Transformer多尺度特征提取模块,注意力联级上下文感知模块(attention-based hierarchical context-aware module, AHCM)、全连接解码器模块(fully connected decoder, FCD)以及融合注意力机制的背景探索模块(background exploration module, BEM),其总体架构如图1所示。该模型以Swin-Transformer^[25]作为

网络骨架提取图片特征,通过 AHCM 模块得到增强的上下文特征表示,使用 FCD 有效定位隐藏对象,最后采用

BEM 模块使模型关注背景区域,更好地理解目标的边界结构,提高边界的定位准确性。

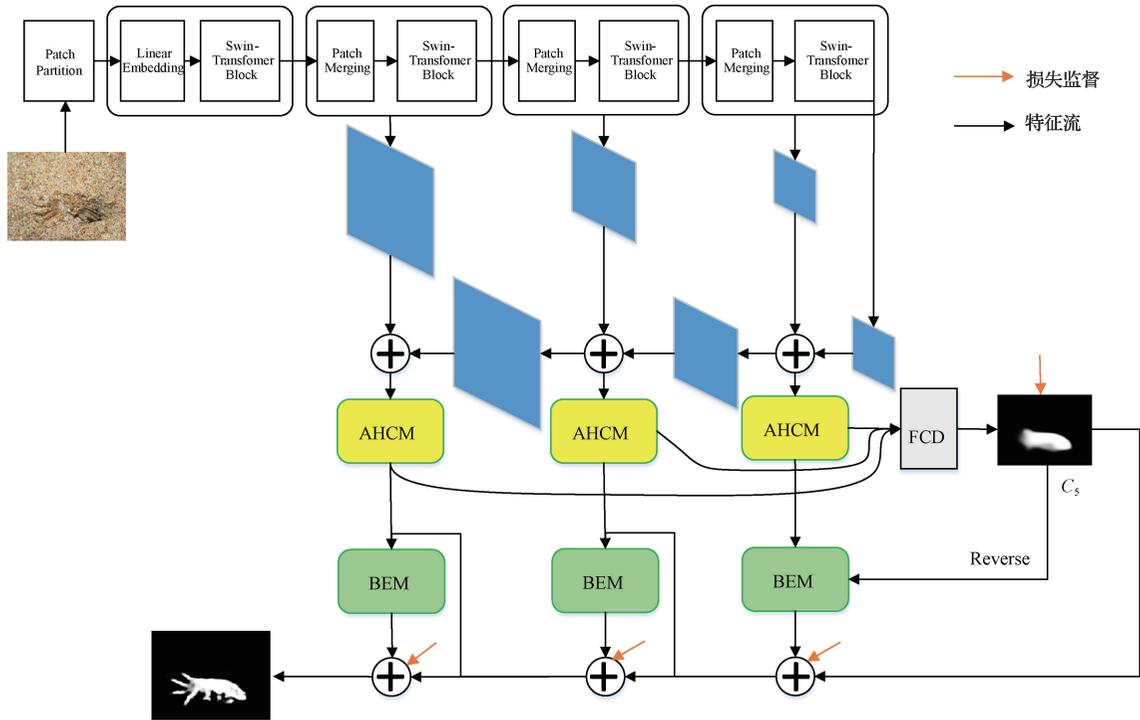


图1 CABENet 框架

Fig.1 CABENet frame diagram

2.1 特征提取模块

Swin-Transformer 是一种特定的 Transformer 变体,通过基于移位窗口的注意力机制捕捉图像中的长距离依赖关系,并使用分层结构减少计算复杂性。Swin-Transformer 一共包含 4 个 stage,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 $O_i, i = \{1, 2, 3, 4\}$ 。低层特征虽然包含目标丰富的局部细节线索,但同时也存在大量的冗余信息,并且 O_i 过于靠近输入图像会占用大量的计算资源对性能的提升有限,于是将 O_2, O_3, O_4 经过特征金字塔^[26]进行特征融合并输入特征增强模块中以生成粗略的预测图。

2.2 AHCM

AHCM 结构如图 2 所示,由 x 个分支组成,每个分支包含不同膨胀率的膨胀卷积。首先,将原始特征输入至不同分支从而扩大感受野,获取目标的多尺度上下文信息;然后,通过双重注意力机制,即空间注意力和通道注意力,提高模型对细节的感知能力,其中空间注意力机制用于捕获图像中不同区域的重要性,而通道注意力机制用于调整特征通道的权重,以提高不同通道的重要性;最后,合并不同分支的结果,增强模型对全局信息的感知能力,加强网络的特征提取能力。

AHCM 具有 4 个平行的残差分支 $\{b_i, i = 1, 2, 3, 4\}$

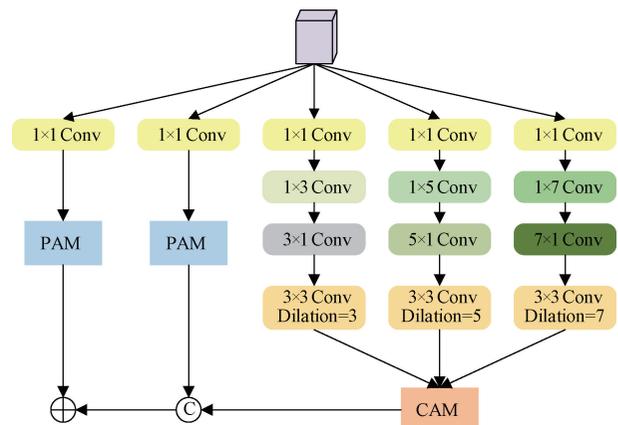


图2 注意力级联上下文感知模块结构

Fig.2 Attention-cascade context-aware module structure diagram

和一个快捷分支 φ 。在每个分支 b_i 中,第 1 卷积层利用 1×1 卷积运算将通道大小减少到 32,第 2 层和第 3 层分别是 $(2i-1) \times (2i-1)$ 的卷积和具有特定的膨胀率 d 的 3×3 卷积(其中 $d = 3, 5, 7; i = 2, 3, 4$)。首先,将后 3 个分支 $\{b_i, i = 2, 3, 4\}$ 拼接,并经过通道注意力模块(CAM)进行特征提取,同时,将残差分支和第 1 个分支输

入空间注意力模块(PAM),然后将4个分支拼接起来,并加上PAM模块输出的残差分支,最后将整个模块的输出馈送到ReLU函数,以获得输出特征 f'_k ,具体过程如下:

$$f_k = \text{PAM}(X_0) + [\text{CAM}(X_2, X_3, X_4), \text{PAM}(X_1)] \quad (1)$$

$$f'_k = \text{ReLU}(f_k) \quad (2)$$

其中, $X_i (i = 0, 1, 2, 3, 4)$ 表示5个分支的输出特征。

1) PAM

PAM结构如图3所示。首先,输入特征 $X \in \mathbf{R}^{C \times H \times W}$, C, H, W 分别表示 X 的通道数、高度和宽度, X 经过3个 1×1 卷积层得到查询特征 Q 、关键特征 K 和值特征 V ,将查询特征和关键特征进行矩阵相乘操作以计算对象之间的相似性,并通过softmax层计算空间注意力图 $S \in \mathbf{R}^{N \times N}$, $N = H \times W$ 表示像素的数量,如式(3)所示。

$$S_{ij} = \frac{\exp(Q_i \cdot K_j)}{\sum_{j=1}^N \exp(Q_i \cdot K_j)} \quad (3)$$

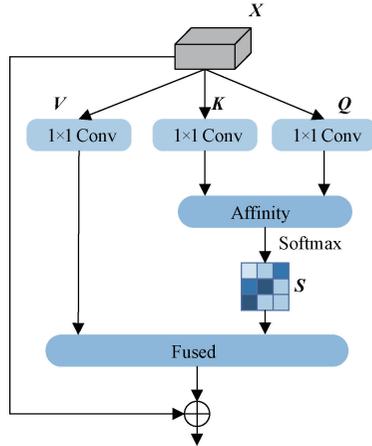


图3 空间注意力结构

Fig. 3 Spatial attention structure diagram

然后,将密集的上下文特征与局部表示特征连接起来,即在 S 的转置和值特征 V 之间执行特征融合操作,并将结果重塑为原始大小。最后,将结果乘以可学习缩放参数 α ,并与输入特征 X_i 执行元素级求和操作,以获得最终输出 E_i :

$$E_i = \alpha \sum_{j=1}^N (V_j S_{ji}) + X_i \quad (4)$$

其中, Q_i 表示矩阵 Q 的第 i 列, K_j 表示矩阵 K 的第 j 列, V_j 表示矩阵 V 的第 j 列, S_{ji} 表示第 i 个位置对 j 个位置的影响。

2) CAM

通道注意力机制通过学习每个通道的重要性权重,使得模型能够更加关注对目标检测任务具有重要影响的通道,其结构如图4所示。首先,将被表示为 $X \in \mathbf{R}^{C \times H \times W}$ 的原始特征重塑为 $X \in \mathbf{R}^{C \times N}$, $N = H \times W$,得到值特征 V' ,

关键特征 K' 和查询特征 Q' ,与空间注意力进行相同的操作获得通道注意力图 $Z \in \mathbf{R}^{C \times C}$,如式(5)所示。

$$Z_{ij} = \frac{\exp(Q'_i \cdot K'_j)}{\sum_{j=1}^C \exp(Q'_i \cdot K'_j)} \quad (5)$$

然后,在 Z 的转置和值特征之间执行矩阵乘法,并将结果重塑为 $\mathbf{R}^{C \times H \times W}$ 。最后,将结果乘以可学习缩放参数 β ,并与 X'_i 执行元素级求和操作,得到最终输出 E_i :

$$E_i = \beta \sum_{j=1}^C (Z_{ij} V'_j) + X'_i \quad (6)$$

其中, Z_{ij} 表示第 j 个通道对第 i 个通道的影响, Q'_i 代表矩阵 Q' 的第 i 行, K'_j 表示矩阵 K' 的第 j 行, V'_j 表示矩阵 V' 的第 j 行。在通道注意力模块提取通道特征的过程中,每个通道的最终特征都是所有通道和原始特征的加权求和。

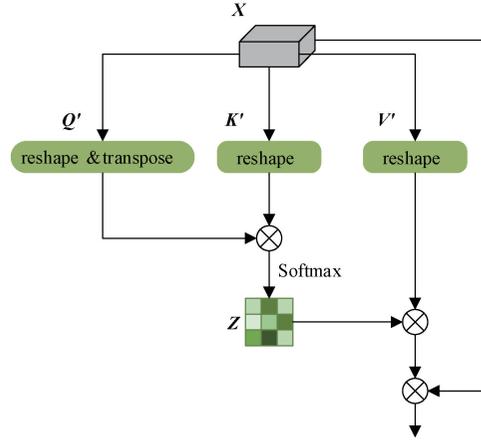


图4 通道注意力结构

Fig. 4 Channel attention structure diagram

2.3 全连接解码器(FCD)

FCD用于聚合AHCM输出的3层特征 $f'_k, k \in \{2, 3, 4\}$,其结构如图5所示。为了确保特征之间的形状匹配,将特征上采样到相同大小,并通过 3×3 卷积改变通道数,然后与另一个特征逐元素相乘,得到了3个细化的特征 $f_k^{mc}, k \in \{2, 3, 4\}$,如式(7)所示。最后将特征 $f_k^{mc}, k \in \{2, 3, 4\}$ 拼接并进行卷积操作以生成粗略位置图 C_5 。

$$\begin{cases} f_4^{mc} = f'_4 \\ f_3^{mc} = f'_3 \otimes \text{BN}[\text{Conv}(\text{U}(f'_4))] \\ f_2^{mc} = f'_2 \otimes \text{BN}[\text{Conv}(\text{U}(f_3^{mc}))] \otimes \text{BN}[\text{Conv}(\text{U}(f'_3))] \otimes \text{BN}[\text{Conv}(\text{U}(f'_4))] \end{cases} \quad (7)$$

式中:BN表示批量归一化操作;U表示上采样操作。

2.4 BEM模块

BEM模块结构如图6所示。首先,对粗略位置图 C_5 执行sigmoid操作和反向操作,获得边缘先验 r_1^k :

$$r_1^k = \mathbf{E} - [\sigma((C_{k+1}))], k = 2, 3, 4 \quad (8)$$

式中: $\sigma(x)$ 是sigmoid函数; \mathbf{E} 表示所有元素为1的矩阵。

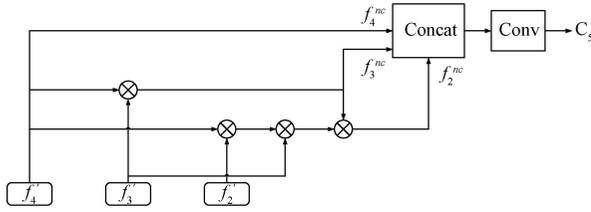


图5 全连接解码器结构

Fig. 5 Fully connected decoder structure

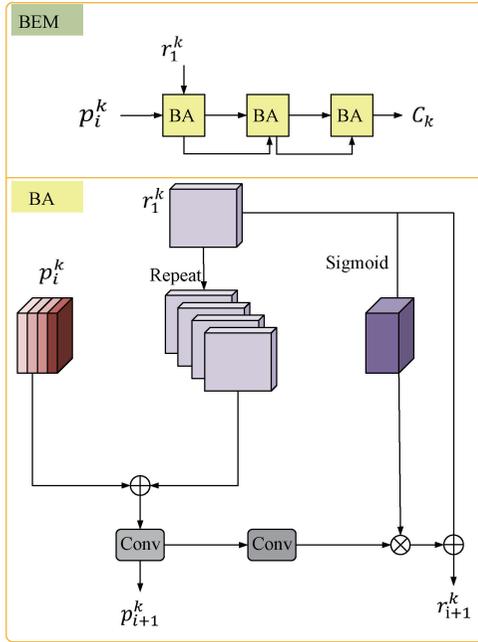


图6 背景探索模块

Fig. 6 Background exploration module

然后,将 AHCM 模块输出的候选特征 p_i^k 与边缘先验 r_1^k 送入 BA (background exploration of attention mechanisms) 模块,进行前景特征与背景特征的交互融合。即将边缘先验按通道数复制,与 AHCM 模块输出的候选特征 p_i^k 进行像素级相加,并通过一个 3×3 卷积层进行通道间的特征融合,得到更新后的细化特征 p_{i+1}^k ,如式(9)所示。然后,通过 3×3 的卷积核,使细化特征的通道数为1,并与经过 sigmoid 操作后的边缘先验相乘,最后,加上原始的边缘先验得到单通道残差先验 r_{i+1}^k ,如式(10)所示。将经过3次 BA 模块后输出的引导作为残差预测 C_k ,如式(11)所示。

$$p_{i+1}^k = p_i^k + \text{Repeat}(r_1^k) \quad (9)$$

$$r_{i+1}^k = r_1^k + \text{conv}(p_{i+1}^k) \otimes \sigma(r_1^k) \quad (10)$$

$$C_k = r_{i+1}^k + C_{k+1} \quad (11)$$

2.5 损失函数

本文采用了损失函数 $L = L_{\text{IoU}}^w + L_{\text{BCE}}^w$,其中, L_{IoU}^w 和 L_{BCE}^w 分别表示对于全局限制和局部(像素级)限制的加权

交叉熵联合(IoU)损失和二进制交叉熵(BCE)损失。与分割任务中广泛采用的标准 IoU 损失不同,本文在标准的 IOU 和 BCE 的基础上,增加了权重因子^[27],使得边缘像素获得更大的权重,总损失如式(12)所示。

$$L_{\text{total}} = L(C_5^{up}, G) + \sum_{i=2}^{i=4} L(\tilde{C}_i^{up}, G) \quad (12)$$

式中: G 为地面真值; C_5^{up} 表示伪装目标的预测图像; \tilde{C}_i^{up} 表示伪装目标边缘的预测图像。

3 实验与分析

3.1 实验数据集

本文在公开的伪装目标检测数据集 CAMO^[28]、COD10K^[7]、CHAMELEON^[29]中进行实验。CAMO 包含 1 250 张图像,其中 1 000 张用于训练,250 张用于测试。CAMO 中涵盖了自然伪装与人工伪装场景下的多种伪装形式(如动物伪装、迷彩伪装和人体彩绘等)。CHAMELEON 包含 76 张手工标注的图像。COD10K 包括 5 066 张伪装图像,其中 3 040 张用于训练,2 026 张用于测试,是目前最大且具有高质量像素级注释的伪装对象数据集。该数据集主要基于自然伪装场景,涵盖了包括飞行、水生、陆地和两栖动物等共 69 个类别的伪装对象。

3.2 实验设置

实验环境使用 Windows 10 操作系统,使用 NVIDIA GeForce 1080Ti 显卡进行模型训练和测试,CPU 配置为 Intel(R) Core(TW) i7-10700CPU,CUDA 版本为 10.1,Pytorch 版本为 1.11.0,Python 语言环境为 3.8,实验设置输入图像大小为 352×352 。在训练过程中,epoch 为 100,批处理数为 30,学习速率为 1×10^{-4} 。

3.3 评价指标

为了将本文模型与其他先进方法进行比较,本文采用结构度量 S_s ^[30]、平均度量 E_ϕ ^[31]、加权度量 F_β^w ^[32]、平均绝对误差(mean absolute error, MAE)^[33] 4 个广泛使用的指标来评估 COD 的性能。 S_s 用来评估预测图和真值图之间的结构相似性; E_ϕ 用来评估预测图的整体和局部精度; F_β^w 用来评价加权精度和加权召回率的综合测度; MAE 用于评估预测图与真值图之间的像素级平均误差。其中, S_s 、 E_ϕ 、 F_β^w 的值越大、MAE 值越小代表模型性能越好。

3.4 实验结果

1) 定量对比

本文分别提供了 CHAMELEON、CAMO 和 COD10K 数据集的定量评估结果。本文将所提出的方法与相关领域中的 10 种先进方法在 3 个基准数据集上进行了比较,结果如表 1 所示。从表 1 可以看出,与先进的 COD 方法相比,本文方法检测精度明显优于其他模型检测精度。在 CAMO 数据集上,与 SINetV2 方法相比,本文方法在结构度量 S_s 、平均增强度量 E_ϕ 、加权 F 度量 F_β^w 上分别提高了 5%、4.6%、7.4%,MAE 下降了 2.5%。

表 1 定量对比
Table 1 Quantitative comparison

模型	会议年份	CHAMELEON				CAMO				COD10K			
		$S_a \uparrow$	$E_\phi \uparrow$	$F_\beta^w \uparrow$	MAE \downarrow	$S_a \uparrow$	$E_\phi \uparrow$	$F_\beta^w \uparrow$	MAE \downarrow	$S_a \uparrow$	$E_\phi \uparrow$	$F_\beta^w \uparrow$	MAE \downarrow
PFNet	CVPR2021	0.882	0.930	0.810	0.033	0.782	0.840	0.695	0.085	0.800	0.868	0.660	0.040
UGTR	ICCV2021	0.888	0.918	0.796	0.031	0.785	0.859	0.686	0.086	0.818	0.850	0.667	0.035
MGL-R	CVPR'2021	0.893	0.923	0.813	0.030	0.775	0.847	0.673	0.088	0.814	0.865	0.666	0.035
SINetV2	TPAMI2022	0.888	0.942	0.816	0.030	0.820	0.882	0.743	0.070	0.815	0.887	0.680	0.037
ZoomNet	CVPR2022	0.889	0.926	0.835	0.026	0.793	0.840	0.715	0.072	0.828	0.923	0.716	0.029
FAPNet	TIP2022	0.893	0.940	0.842	0.028	0.815	0.865	0.776	0.076	0.822	0.888	0.731	0.036
C2FNet	TCSVT/2022	0.888	0.935	0.828	0.032	0.796	0.864	0.719	0.080	0.813	0.890	0.686	0.036
BCNet	NCA2023	0.910	0.944	0.839	0.029	0.830	0.886	0.761	0.068	0.827	0.894	0.704	0.033
PENet	IJCAI'2023	0.902	0.960	0.851	0.024	0.828	0.890	0.771	0.063	0.831	0.908	0.723	0.031
MRR-Net	TNNLS'2023					0.827	0.886	0.760	0.070	0.835	0.889	0.720	0.032
本文	—	0.888	0.960	0.821	0.027	0.876	0.938	0.832	0.041	0.850	0.924	0.732	0.026

2) 定性对比

在视觉检测方面,本文模型与其他6种具有代表性的模型的比较结果,如图7所示。从图7可以看出,本文所提方法能更准确、更完整地分割伪装对象。通过图7A和E可以看出,对于低对比度且边界结构复杂的对象,所提方法相对其他方法得到的伪装对象图像轮廓更明显,清晰度更高,这得益于融合注意力机制的背景探索模块,通过探索背景区域增强边界特征的方式使得目标的边界更加

清晰明确;通过图7B和C可以看出,所提方法能定位具有不同尺度的伪装目标,这得益于注意力级联上下文感知模块,通过有效地整合不同层次的特征信息,使得模型在处理多尺度目标时能够更加灵活和准确;通过图7D可以看出,所提方法能够同时检测出多个伪装对象,这得益于通过Swin-Transformer网络在全局范围内捕获图像特征;通过图7F可以看到,目标被周围环境部分遮挡而本文所提模型仍然能成功地检测出伪装目标。

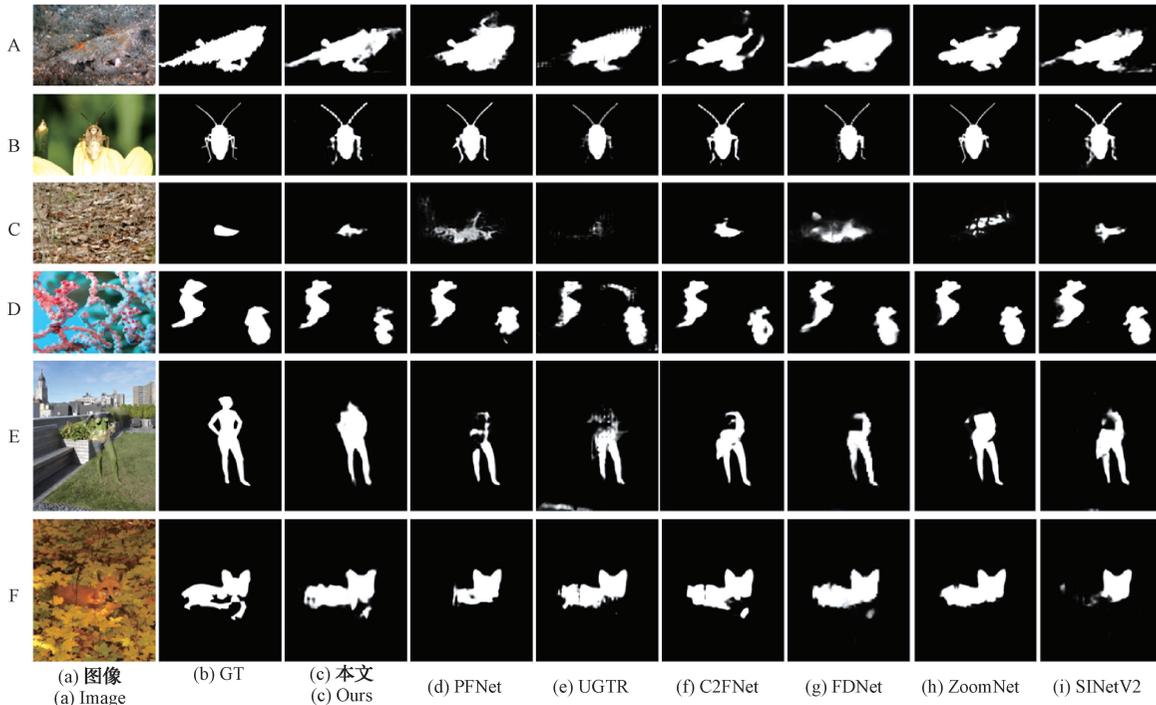


图 7 定性对比

Fig. 7 Qualitative contrast figure

3.5 消融实验

为了验证本文模型中各个模块的有效性,对各个子模

块进行了消融实验,得到的定量结果如表2所示,可视化结果如图8所示。方法1表示以Swin-Transformer和特

征融合部分作为基线模型(Baseline)进行实验。方法2和方法3表示在Baseline的基础上分别加上AHCM和融合注意力机制的BEM进行实验。方法4表示整体模型的实验结果。

从方法2可以看出,相比与单一的基础网络,Baseline+AHCM在COD10K数据集上 S_a 、 E_ϕ 、 F_β^w 分别提高了0.3%、0.6%、1.2%,MAE下降了0.2%,表明AHCM模块通过不同分支有效获取了目标的多尺度上下文信息,此外,通过空间注意力和通道注意力,进一步细化上下文特征中的多尺度信息,增强了不同尺度的伪装物体的感知和定位能力。通过图8(c)和(d)的比较,可以看出伪装目标

能被更清晰地检测出来,从视觉效果上进一步证明AHCM对于模型性能的提升是必要的。

从方法3可以看出,相比与单一的基础网络,Baseline+BEM在COD10K数据集上 S_a 、 E_ϕ 、 F_β^w 分别提高了0.9%、1.4%、3.1%,MAE下降了0.3%,这得益于BEM模块为模型提供重要的边缘预测,通过将额外的边界信息嵌入特征中,有效的增强了模型对边界的理解。通过图8(c)和(e)的比较,可以看出BEM模块使伪装目标的边界结构更清晰。从方法4可以看出综合多个模块与使用单个模块相比,性能均得到了提升,从可视图可以看出,完整的模型更有利于识别和定位伪装物体。

表2 消融实验

Table 2 Ablation experiment

方法	CHAMELEON				CAMO				COD10K			
	$S_a \uparrow$	$E_\phi \uparrow$	$F_\beta^w \uparrow$	MAE	$S_a \uparrow$	$E_\phi \uparrow$	$F_\beta^w \uparrow$	MAE↓	$S_a \uparrow$	$E_\phi \uparrow$	$F_\beta^w \uparrow$	MAE↓
1 Baseline	0.877	0.936	0.785	0.033	0.869	0.922	0.801	0.049	0.830	0.899	0.685	0.034
2 Baseline+AHCM	0.880	0.939	0.792	0.032	0.870	0.924	0.805	0.048	0.833	0.905	0.697	0.032
3 Baseline+BEM	0.887	0.942	0.808	0.030	0.870	0.928	0.813	0.046	0.839	0.913	0.716	0.031
4 Baseline+AHCM+BEM	0.888	0.960	0.821	0.027	0.876	0.938	0.832	0.041	0.850	0.924	0.732	0.026

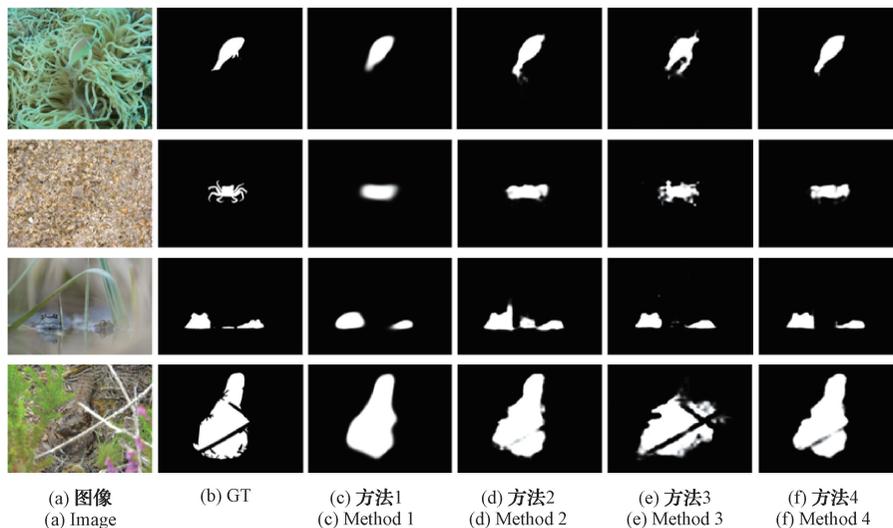


图8 消融实验可视化

Fig. 8 Ablation experiment visualization

4 结论

针对COD中结果不完整、边缘细节模糊的问题,本文提出了一种融合上下文感知和背景探索的伪装目标检测方法。该方法使用Swin-Transformer实现多尺度特征表示,在注意力联级上下文感知模块通过双重注意力机制帮助模型在不同的空间尺度和通道之间进行更细粒度的特征交互和整合,最后通过融合注意力机制的背景探索模块从深层细化粗略预测。在3个公开数据集上的实验结果表明,与其他方法相比,本文方法取得了更好的检测效

果。在未来的工作中,将进一步探究新的网络模型,以使COD方法能实现更加精确的伪装目标检测。

参考文献

- [1] FAN D P, TAO Z, JI G P, et al. Inf-Net: Automatic COVID-19 lung infection segmentation from CT images [J]. IEEE TM I, 2020, 39(8):2626-2637.
- [2] FANG F, LI L, 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 Trans. Graph*, 2010, 29(4):51:1-51:8.
- [4] BOOT W R, NEIDER M B, KRAMER A F. Training and transfer of training in the search for camouflaged targets [J]. *Attention Perception & Psychophysics*, 2009, 71(4):950-963.
- [5] SONG L M, GENG W D. A new camouflage texture evaluation method based on WSSIM and nature image features [C]. 2010 International Conference on Multimedia Technology. IEEE, 2010:1-4.
- [6] HALL J R, CUTHILL I C, BADDELEY R, et al. Camouflage, detection and identification of moving targets[J]. *Proceedings of the Royal Society B: Biological Sciences*, 2013, 280(1758): 20130064.
- [7] FAN D P, JI G P, SUN G, et al. Camouflaged object detection [C]. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020: 2777-2787.
- [8] FAN D P, JI G P, CHENG M M, et al. Concealed object detection[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, 44(10): 6024-6042.
- [9] MEI H, JI G P, WEI Z, et al. Camouflaged object segmentation with distraction mining [C]. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021: 8772-8781.
- [10] PANG Y, ZHAO X, XIANG T Z, et al. Zoom in and out: A mixed-scale triplet network for camouflaged object detection[C]. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022: 2160-2170.
- [11] 孙备,党昭洋,吴鹏,等.多尺度交叉注意力改进的单无人机对地伪装目标检测定位方法[J]. *仪器仪表学报*, 2023, 44(6):54-65.
SUN B, DANG ZH Y, WU P, et al. Multi-scale Intercross attention improved detection and location method for single UAV to ground camouflage target[J]. *Chinese Journal of Scientific Instrument*, 2023, 44(6):54-65.
- [12] ZHANG Q, SUN X, CHEN Y, et al. Attention-induced semantic and boundary interaction network for camouflaged object detection[J]. *Computer Vision and Image Understanding*, 2023, 233: 103719.
- [13] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers [C]. *European Conference on Computer Vision*. Cham: Springer International Publishing, 2020: 213-229.
- [14] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16×16 words: Transformers for image recognition at scale [J]. *ArXiv preprint arXiv:2010.11929*, 2020.
- [15] WANG W, XIE E, LI X, et al. Pyramid vision transformer: A versatile backbone for dense prediction without convolutions[C]. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021: 568-578.
- [16] 赵云亮,唐东林,何媛媛,等.基于CNN融合PGW-Attention的金属表面缺陷识别方法[J/OL]. *电子测量与仪器学报*, 1-11[2024-07-06].
ZHAO Y L, TANG D L, HE Y Y, et al. Metal surface defect recognition method based on CNN fusion PGW-Attention [J/OL]. *Journal of Electronic Measurement and Instrumentation*, 1-11 [2024-07-06].
- [17] 王琪,赵文仓.基于预过滤注意力的Transformer物体检测[J]. *电子测量技术*, 2022, 45(24):145-152.
WANG Q, ZHAO W C. Object detection with Transformer based on pre-filtered attention [J]. *Electronic Measurement Technology*, 2022, 45(24): 145-152.
- [18] ESSER P, ROMBACH R, OMMER B. Taming transformers for high-resolution image synthesis[C]. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021: 12873-12883.
- [19] LIU S, HUANG D. Receptive field block net for accurate and fast object detection[C]. *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018: 385-400.
- [20] WU Z, SU L, HUANG Q. Cascaded partial decoder for fast and accurate salient object detection [C]. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019: 3907-3916.
- [21] WANG K, BI H, ZHANG Y, et al. D2C-Net: A dual-branch, dual-guidance and cross-refine network for camouflaged object detection [J]. *IEEE Transactions on Industrial Electronics*, 2021, 69(5): 5364-5374.
- [22] SUN Y, CHEN G, ZHOU T, et al. Context-aware cross-level fusion network for camouflaged object detection[J]. *ArXiv preprint arXiv: 2105.12555*, 2021.
- [23] ZHANG Z, LIN Z, XU J, et al. Bilateral attention network for RGB-D salient object detection[J]. *IEEE Transactions on Image Processing*, 2021, 30: 1949-

- 1961.
- [24] 周兴超,魏为民,裴仁莹,等. 基于边缘特征增强的多尺度监督图像篡改检测[J]. 国外电子测量技术, 2023,42(9):122-129.
ZHOU X CH, WEI W M, PEI R Y, et al. Multi-scale supervised image tampering detection based on edge feature enhancement [J]. Foreign Electronic Measurement Technology,2023,42(9):122-129.
- [25] LIU Z, LIN Y, CAO Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 10012-10022.
- [26] CHEN T, MO L. Swin-fusion: Swin-transformer with feature fusion for human action recognition[J]. Neural Processing Letters, 2023, 55 (8): 11109-11130.
- [27] QIN X, ZHANG Z, HUANG C, et al. Basnet: Boundary-aware salient object detection [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 7479-7489.
- [28] LE T N, NGUYEN T V, NIE Z, et al. Anabranh network for camouflaged object segmentation [J]. Computer Vision and Image Understanding, 2019, 184: 45-56.
- [29] SKUROWSKI P, ABDULAMEER H, BLASZCZYK J, et al. Animal camouflage analysis: Chameleon database [EB/OL]. <http://kgwisc.aei.polsl.pl/datasets/CamouflageBase>, 2018.
- [30] FAN D P, CHENG M M, LIU Y, et al. Structure-measure: A new way to evaluate foreground maps[C]. Proceedings of the 18th International Conference on Computer Vision (ICCV2017), 2017: 4548-4557.
- [31] FAN D P, GONG C, CAO Y, et al. Enhanced-alignment measure for binary foreground map evaluation[C]. Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI 2018), 2018: 698-704.
- [32] MARGOLIN R, ZELNIK-MANOR L, TAL A. How to evaluate foreground maps? [C]. Proceedings of the 27th International Conference on Computer Vision and Pattern Recognition (CVPR2014). IEEE, 2014: 248-255.
- [33] PERAZZI F, KRAHENBUHL P, PRITCH Y, et al. Saliency filters: Contrast based filtering for salient region detection [C]. Proceedings of the 25th International Conference on Computer Vision and Pattern Recognition (CVPR 2012). IEEE, 2012: 733-740.

作者简介

陈世洁, 硕士研究生, 主要研究方向为图像处理。

E-mail:1459466181@qq.com

范李平, 高级工程师, 主要研究方向为智能电网。

E-mail:184948938@qq.com

余肖生, 副教授, 主要研究方向为计算机视觉、大数据人工与智能。

E-mail:yuxiaosheng@ctgu.edu.cn

王东娟(通信作者), 副教授, 主要研究方向为大数据与人工智能。

E-mail:w dj@ctgu.edu.cn