

文章编号:1007-2780(2024)01-0048-11

基于边缘增强和特征融合的伪装目标分割

李明岩^{1,2}, 吴川^{1,2*}, 朱明^{1,2}

(1. 中国科学院长春光学精密机械与物理研究所, 吉林 长春 130033;

2. 中国科学院大学, 北京 100049)

摘要: 伪装目标分割的任务是使用像素级分割掩码将与背景高度相似的目标进行准确分类和定位, 与传统的目标分割任务相比更具挑战性。针对目标与周围环境高度相似、边界模糊、对比度低等问题, 构建了一种基于边缘增强和特征融合的伪装目标分割方法。首先, 设计了一组边缘提取模块, 能够更准确地分割有效的边缘先验。之后, 引入了多尺度特征增强模块和跨层级特征聚合模块, 分别挖掘层内与层间的多尺度上下文信息。提出了一种简单的层间注意力模块, 利用相邻层级间的差异有效滤除融合后存在的干扰信息。最后, 通过将各级特征图与边缘先验逐级结合的方式, 获得准确的预测结果。实验结果表明, 在 4 个伪装目标基准数据集上, 该模型的表现都优于其他算法。其中加权 F 值提升了 2.4%, 平均绝对误差减少了 7.2%, 在 RTX 2080Ti 硬件环境下分割速度达到了 44.2 FPS。与现有方法相比, 该算法能够更准确地分割伪装目标。

关键词: 深度学习; 伪装目标; 图像分割; 边缘特征; 特征融合

中图分类号: TP391.4 **文献标识码:** A **doi:** 10.37188/CJLCD.2023-0064

Camouflaged object segmentation based on edge enhancement and feature fusion

LI Mingyan^{1,2}, WU Chuan^{1,2*}, ZHU Ming^{1,2}

(1. Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China;

2. University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: The task of camouflaged object segmentation is to accurately classify and localize objects that are highly similar to the background using pixel-level segmentation masks, which is more challenging than traditional object segmentation tasks. Aiming at the problems that the target is highly similar to the surrounding environment, the boundary is blurred, and the contrast is low, a camouflaged target segmentation method based on edge enhancement and feature fusion is constructed. First, a set of edge extraction modules is designed, aiming to accurately segment valid edge priors. Afterwards, a multi-scale feature enhancement module and a cross-level feature aggregation module are introduced to mine multi-scale contextual information within and between layers, respectively. In addition, a simple inter-layer attention module is proposed to effectively filter out the interference information existing after fusion by utilizing the difference between adjacent layers. Finally, accurate prediction results are obtained by

收稿日期: 2023-02-20; 修订日期: 2023-03-20.

*通信联系人, E-mail: wuchuan0458@sina.com

combining feature maps of all levels with edge priors step by step. Experimental results show that the model outperforms other algorithms on four camouflaged target benchmark datasets. Among them, the weighted F value increased by 2.4%, the average absolute error decreased by 7.2%, and the segmentation speed reached 44.2 FPS under the RTX 2080Ti hardware environment. Compared with existing methods, this algorithm can segment camouflage targets more accurately.

Key words: deep learning; camouflaged object; image segmentation; edge feature; feature fusion

1 引言

伪装原本是指动物用来隐藏自己或欺骗其他动物的一种方法,而伪装能力通常会影响到动物的生存概率。随着仿生学技术的发展,伪装技术也常被人类应用于某些场合,例如现代军队所使用的迷彩。与显著目标检测(Salient Object Detection, SOD)相比,伪装目标通常与背景具有高度相似性,因此伪装目标检测(Camouflage Object Detection, COD)更加具有难度。

传统的方法主要关注图像的底层特征(如颜色、纹理、梯度等)来评估物体与背景的差异,基于三维凸度^[1]、灰度共生矩阵^[2]、Canny边缘检测器^[3]、光流^[4]等方法进行伪装目标检测。这些方法往往只考虑图像的一部分特征,难以适用于所有场景。近年来,深度学习的方法在许多领域展现出优异的性能,为一些传统方法难以处理的问题提供了新的解决方案^[5]。在图像领域,通过卷积神经网络(Convolution Neural Network, CNN)提取出的深度特征的代表能力要远强于图像的底层特征,因此,研究人员开始尝试用深度学习的方法来解决伪装目标分割的问题。Le等人建立了一个新的伪装图像数据集用于基准测试,并提出了一个端到端网络,其中包括一个分割分支和另一个分类分支。分类分支用于预测图像包含伪装对象的概率,随后用于增强分割分支中的分割性能^[6]。Fan等人提出了SINet模型,该模型包含两个模块:搜索模块定位伪装目标,识别模块精确检测伪装目标,同时建立了首个大规模伪装目标数据集^[7]。Mei等人在伪装目标分割任务中引入了干扰的概念,开发了一种新的干扰识别和去除的干扰挖掘方法,帮助对伪装目标的准确检测^[8]。Zhai等人将特征映射解耦为两个特定的任务:一个用于粗略地定位对象,另一个用于准确地预测边缘细节,并通过图迭代推理它们的高

阶关系^[9]。Li等人提出联合训练SOD和COD任务,利用相互矛盾的信息同时提高两个任务的表现^[10]。

以上模型在多个伪装目标数据集上已经取得了较好的效果,但仍存在一些问题:模型的全局建模能力不足,在存在多个伪装目标且与背景高度相似时,模型容易出现漏检和误检的问题。其次,这些模型舍弃了较低层次的特征,而较深层次的特征在进行数次下采样后分辨率迅速下降,虽然减少了计算量,但同时也丢失了大量的边缘等细节信息,这也导致了模型的精度下降。最后,这些模型大多使用了极其复杂的特征融合策略,大幅提高了模型的复杂度与推理时间。

针对以上问题,本文提出了一种基于边缘增强和多级特征融合的伪装目标分割模型。首先,选择ResNet-50作为骨干网络,提取多级深度特征。其次,设计了一个边缘提取模块,选择融合细节信息丰富的两个较低层级特征,通过网络的不断学习,得到精确的边缘先验。同时,通过引入多尺度特征增强模块和跨层级特征聚合模块,分别在层内和层间增强特征表示,弥补多尺度表征能力不足的问题。之后,设计了一种简单但有效的层间注意力模块,利用相邻层的差异性,有选择地筛选出各层级间有用的信息,在保持层内语义完整的同时滤除背景噪声,再与边缘先验引导结合,输出精确的预测图。最后,为了增强网络的学习能力,本文采用加权二元交叉熵损失与加权IOU(Intersection-Over-Union)损失两部分作为损失函数,对3个不同尺度的输出预测同时进行深度监督;采用单独的加权二元交叉熵损失对边缘预测图进行监督,边缘的真值可以使用Canny边缘检测方法从标注图中得到。本文方法在4个伪装目标公开数据集CHAMELEON^[11]、CAMO^[6]、COD10K^[7]、NC4K^[12]上与先进的方法进行对比实验。实验结果表明,本文方法在所有数据集的

4个常用的评价指标上均优于其他方法,同时能够满足实时性的需要,表明本文方法在伪装目标分割任务上具有优异的性能。

2 本文方法

如图1所示。对于一幅输入图像 $I \in \mathbb{R}^{H \times W \times 3}$ (其中 H 为图像高, W 为图像宽,通道数为3),采用ResNet-50作为特征提取网络^[13],获得不同尺度的特征 $f_i, i \in \{1, 2, 3, 4, 5\}$ 。首先,使用一个边

缘提取模块来显式地建模边缘特征,以此指导网络学习并提升检测性能。然后,将获得的多级特征通过多尺度特征增强模块(Multi-scale Feature Enhanced Module, MFEM)筛选出更具辨识度的特征。增强后的特征图每3个层级为1组,同时输入到跨层级特征聚合模块(Cross-level Fusion Module, CFM),有选择地结合不同尺度下的有效特征。通过注意力模块给有价值的信息以更高的权重。最后与边缘信息加以结合,得到最终的伪装目标分割预测图。

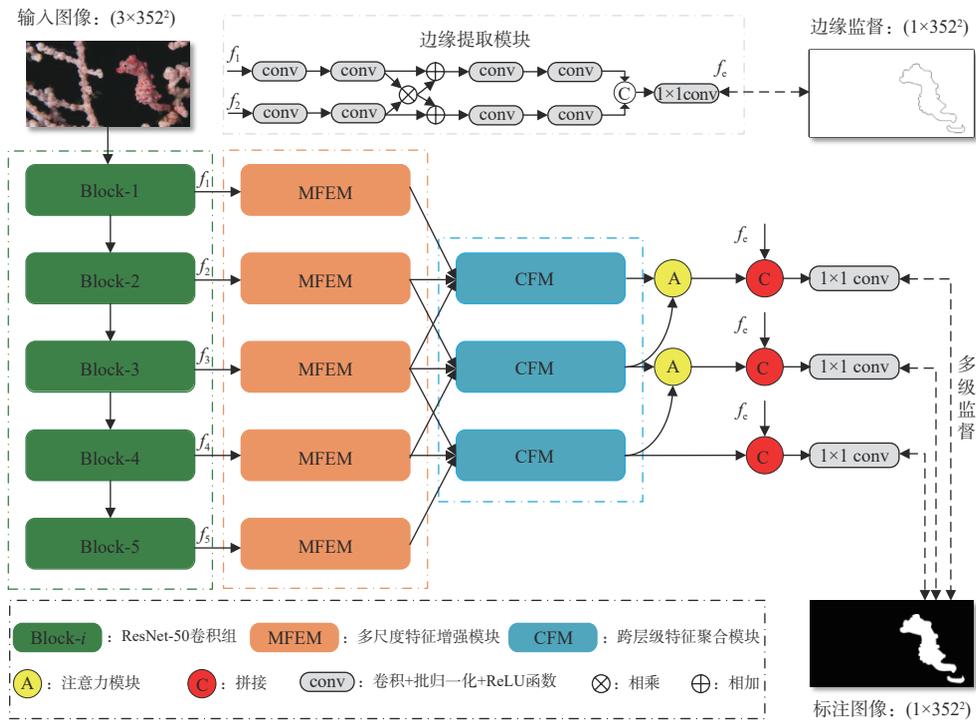


图1 网络结构图

Fig. 1 Network structure diagram

2.1 边缘提取模块

早期的一些相关工作^[14-15]已经证实了边缘信息有助于提升计算机视觉任务的性能。在伪装目标分割任务中,由于目标与背景的高度相似性,难以清晰地辨别伪装对象与周围环境之间的边界。现有的研究表明^[16],通常只有低级特征保留了大量的边缘细节信息,而在经过多个下采样操作后,细节信息会丢失。因此,本文设计了一个边缘提取模块,如图1所示。自ResNet-50中提取出最低两层的特征 $f_1 \in \mathbb{R}^{\frac{H}{4} \times \frac{W}{4} \times 64}$, $f_2 \in \mathbb{R}^{\frac{H}{4} \times \frac{W}{4} \times 256}$, 分别经过两个包含 3×3 卷积、批归一化(Batch

Normalization, BN)及ReLU激活函数的卷积组,用来保证两组特征的通道数一致。然后通过逐元素相乘的方式得到融合的特征表示。将融合后的特征分别与卷积后的特征相加,用来抑制背景噪声和增强感兴趣区域。增强后的特征再次分别输入两个卷积组,通过拼接和一组 1×1 卷积进行降维,得到全局边缘引导特征图 f_e 。最后,将边缘特征图通过双线性上采样的方式恢复到与原图相同的尺寸,用来监督学习边缘表示。目标边缘的真值图可以通过对二值标注图进行Canny边缘检测得到。

2.2 多尺度特征增强模块

伪装目标通常具有较大的尺度范围。每个单独的卷积层只能处理一种特定的尺度,为了从单个层级获取多尺度信息以表征尺度的变化,受Inception-V3^[17]启发,设计了一个多尺度特征融合模块(MFEM),如图2所示。对于每个输入的特征 f_i ,每个MFEM包含4个并联的残差支路和一个主路。在每个残差支路中,首先使用一个 1×1 卷积将特征的通道数降为64。接下来的两层分别为 $(2i-1) \times 1$ 和 $1 \times (2i-1)$ 大小的卷积核, $i \in \{1, 2, 3, 4\}$ 。这两层卷积的串联等效于两个 $(2i-1) \times (2i-1)$ 大小的卷积核的串联,在不影响模块表征能力的同时能够减少计算的损耗。在支路的最后是一个膨胀率为 $(2i-1)$, $i \in \{2, 3, 4\}$ 的 3×3 卷积层,增大感受野以捕捉细粒度特征。最后,将4个支路的特征拼接在一起并通过卷积来保证通道数一致,拼接后的特征与主路相加后,通过ReLU函数来获得输出特征 f'_i 。

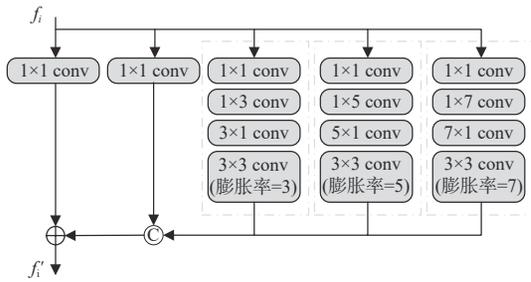


图2 多尺度特征增强模块

Fig. 2 Multi-scale feature enhanced module

2.3 跨层级特征聚合模块

在跨层级特征融合的过程中,如何有效地保持层内语义的一致性及利用层间的相关性是构建伪装目标分割网络的关键。为此,本文提出了一个跨层级特征聚合模块(CFM),如图3所示。给

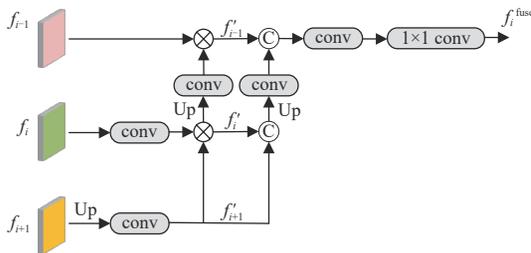


图3 跨层级特征聚合模块

Fig. 3 Cross-level fusion module

定一组特征 f_{i-1}, f_i, f_{i+1} , $i \in \{2, 3, 4\}$, 首先将较高层的特征 f_{i+1} 与中间层特征 f_i 分别经过一个卷积组后进行逐元素相乘,其中 f_{i+1} 在卷积前进行一次双线性上采样操作以保证维度大小相同。将聚合后的特征再次进行上采样与卷积操作并与较低层特征相乘,得到 f'_{i-1}, f'_i, f'_{i+1} , $i \in \{2, 3, 4\}$ 。之后,将初步融合后的特征由较高层到较低层逐级进行拼接操作。最后,将拼接后的特征矩阵分别经过一个 3×3 大小的卷积组和一个 1×1 大小的卷积核,将通道数减少为原始通道数,得到最后的输出特征 f_i^{fuse} , $i \in \{2, 3, 4\}$ 。整个过程定义如式(1)~式(5)所示:

$$f'_{i+1} = \text{Conv}_4(f_{i+1}), \quad (1)$$

$$f'_i = \text{Conv}(f_i) \cdot f'_{i+1}, \quad (2)$$

$$f'_{i-1} = f_{i-1} \cdot \text{Conv}_4(f'_i), \quad (3)$$

$$f_i^c = \text{Cat}(f'_{i-1}, \text{Conv}_4(\text{Cat}(f'_i, f'_{i+1}))), \quad (4)$$

$$f_i^{fuse} = \text{Conv}_1(\text{Conv}(f_i^c)), \quad (5)$$

其中: $\text{Conv}_4(\cdot)$ 表示一个2倍的双线性上采样接一个包含 3×3 卷积、批归一化和ReLU函数的卷积组, $\text{Cat}(\cdot)$ 表示对括号内的两个元素进行拼接操作, $\text{Conv}_1(\cdot)$ 是一个单独的 1×1 卷积核。接着,将输出特征传入注意力模块进行筛选。

2.4 注意力模块

通过简单的拼接或相加的方式融合的特征往往是复杂且低效的,大量的噪声及低置信度的信息混杂在一起会对网络的学习造成巨大的困难。为此,设计了一个简单的注意力模块,对CFM融合后的特征进一步结合并筛选,模块结构图见图4。首先将两组相邻层级的特征 f_i, f_{i+1} , $i \in \{2, 3\}$ 进行卷积处理,再分别通过一个 1×1 卷积将通道维度由 C 降为1。即:

$$f_i^{1 \times h \times w} = \text{Conv}_1(\text{Conv}(f_i^{c \times h \times w})), \quad (6)$$

其中: h, w 分别表示特征的高度、宽度两个维度的大小, c 为通道数。将处理后的矩阵拼接并在通道维度上进行Softmax函数计算,可以得到两个不同的权重矩阵 $\omega_1^{1 \times h \times w}, \omega_2^{1 \times h \times w}$ 。将两个权重矩阵分别与对应的特征图相乘,可以得到:

$$f_i^{(x,y)'} = f_i^{(x,y)} \cdot \omega_1^{(x,y)}, \quad (7)$$

$$f_{i+1}^{(x,y)'} = f_{i+1}^{(x,y)} \cdot \omega_2^{(x,y)}, \quad (8)$$

$f^{(x,y)}$ 表示该特征矩阵上对应于 (x, y) 位置的元

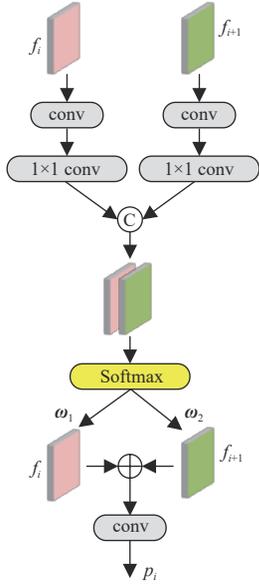


图 4 层间注意力模块

Fig. 4 Cross-layer attention module

素。其中 $\omega_1^{(x,y)} + \omega_2^{(x,y)} = 1$, ω 值越大, 表示该位置的信息越应该被保留, 反之则被舍去。最后, 将两个特征相加, 并经过卷积平滑操作, 得到最终的输出结果:

$$p_i = \text{Conv}(f_i' + f_{i+1}'). \quad (9)$$

采用这种注意力机制可以有效地筛选出各级特征中更有效的信息, 抑制噪声, 增强不同尺度特征的表示能力。需要注意的是, 由于 f_i 是由较高 3 个层的特征聚合得到, 在卷积的过程中已经被过滤掉了大量的细节和噪声信息, 因此将 f_i 不经过注意力机制筛选而直接输出, 即:

$$p_4 = f_4. \quad (10)$$

最后, 将 p_2, p_3, p_4 分别与全局边缘引导 f_e 拼接, 经过 1×1 卷积降维后, 上采样回原图尺寸, 得到最终的预测图 P_2, P_3, P_4 。其中 P_2 作为最终的预测结果用来评估网络性能。

2.5 损失函数

在伪装目标分割过程中, 使用了加权二元交叉熵损失^[18] (weighted binary cross entropy loss, wBCE) 及加权交并比损失^[18] (weighted intersection-over-union loss, wIOU) 两部分作为损失函数。加权二元交叉熵函数的公式如式(11)所示:

$$L_{\text{wBCE}} = - \sum_{i=1}^N \omega_i \cdot y_i \cdot \log z_i + (1 - y_i) \cdot \log(1 - z_i), \quad (11)$$

其中: y_i 是二元标签 0 或 1, z_i 为输出属于 y_i 标签的概率, ω_i 为权重向量。加权交并比损失的公式如式(12)所示:

$$L_{\text{wIOU}} = 1 - \frac{|P \cap G|}{|P \cup G|}, \quad (12)$$

其中: P 为预测值, G 为真实标注值。两种损失函数分别计算全局损失和像素级损失。与标准的损失函数不同, 加权损失更侧重于给较难的像素更大的权重, 而不是给每个像素相同的权重。

基于上述分析, 使用联合损失函数对 P_2, P_3, P_4 3 个预测图及边缘预测图 f_e 进行深度监督。整体的损失函数可以表示为:

$$L = \lambda_1 L_{\text{edge}} + \lambda_2 L_{\text{pred}}, \quad (13)$$

$$L_{\text{edge}} = L_{\text{wBCE}}(f_e, E_g), \quad (14)$$

$$L_{\text{pred}} = \sum_{i \in \{2, 3, 4\}} L_{\text{wBCE}}(p_i, p_g) + L_{\text{wIOU}}(p_i, p_g), \quad (15)$$

其中: E_g 表示边缘标注值, 可以通过对二元标注图进行 Canny 边缘检测得到。 p_g 表示二元标注图。没有对边缘使用 L_{wIOU} 损失函数监督是因为边缘预测图的前景区域与背景区域差别过大, 可能导致损失函数下降不稳定。 λ_1 和 λ_2 用来平衡两个不同损失函数对整体损失的贡献, 其中 λ_1 设置为 5, λ_2 设置为 1。

3 实验与分析

3.1 数据集

我们在 4 个通用的伪装目标标准数据集上进行实验: CHAMELEON^[11] 是一个小样本的伪装目标数据集, 其中包含 76 张图像, 每张图像至少有一个伪装目标; CAMO^[6] 包含 1 000 张用于训练的图像和 250 张用于测试的图像, 数据集涵盖了大量自然及人工场景下的伪装目标; COD10K^[7] 是迄今为止最大的基准数据集, 它包含 5 个大类和 69 个子类, 共有 3 040 张训练图像和 2 026 张测试图像; NC4K^[12] 是规模最大的伪装目标测试数据集, 包含 4 121 张图像, 可用来评估模型的泛化能力。仿照之前的工作^[7], 本文将 CAMO 的训练集和 COD10K 的训练集结合, 作为完整的训练数据集 (其中包含 4 040 张图片), 并在余下的数据集上测试模型的性能。

3.2 评价指标

本文使用了 4 个广泛使用的评价指标: 结构

性度量(S_a)^[19]、 E 指标(E_o)^[20]、带权重的 F 指标(F_β^w)^[21]及平均绝对误差(MAE)^[22]。

结构性度量(S_a)评估预测结果及标注图像之间的区域级和对象级结构相似性,如式(16)所示:

$$S_a = \alpha \cdot S_o(P, P_g) + (1 - \alpha) \cdot S_r(P, P_g), \quad (16)$$

其中, S_o 和 S_r 分别表示对象级和区域级的结构相似性。根据其他研究中的经验^[19],这里的 α 设置为0.5。

E 指标(E_o)使用一个矩阵(O_{FM})联合计算图像级的统计信息和像素级的匹配信息,可以同时衡量预测的整体完整性和局部精确性,如式(17)所示:

$$E_o = \frac{1}{w \times h} \sum_{x=1}^w \sum_{y=1}^h O_{FM}(x, y), \quad (17)$$

其中: w 表示图像宽度, h 表示图像宽。

带权重的 F 指标(F_β^w)定义一个加权精度(P^w)和加权召回率(R^w)来衡量预测的准确性和完整性:

$$F_\beta^w = (1 + \beta^2) \times \frac{P^w \times R^w}{\beta^2 \times P^w + R^w}, \quad (18)$$

其中, β^2 是一个平衡系数,根据其他研究中的经验^[21], β^2 设置为0.3。

平均绝对误差(MAE)用来衡量预测结果与标注图像之间的像素级差异,其被广泛应用于各类分割任务:

$$MAE = \frac{1}{w \times h} \sum_{x=1}^w \sum_{y=1}^h |P(x, y) - P_g(x, y)|. \quad (19)$$

为了进行公平的对比,我们使用相同的代码,对不同数据集的4种评价指标进行计算。

3.3 实验细节

本文模型基于Pytorch框架构建,在NVIDIA GeForce RTX 2080TiGPU上进行所有实验。使用在ImageNet上预训练的权重文件初始化ResNet-50骨干网络的参数,其他参数由网络默认生成。在训练之前,所有训练图像及标注图像均被调整为 352×352 大小,并且不使用任何数据增强策略。批量大小设置为8并在训练过程中使用了Adam优化器,初始的学习率设置为 $1e-4$,并且每30次迭代后,学习率除以10,网络共训练60轮,大约需要5.5h。在测试过程中,测试图像同样被调整为 352×352 大小,随后输入网络。预测图通过双线性上采样操作缩放到原始大小以评估结果。

3.4 实验结果及对比

将本文方法与现有的11种COD方法进行比较,包括BASNet^[23]、EGNet^[24]、CPD^[16]、F3Net^[18]、PraNet^[25]、SINet^[7]、PFNet^[8]、C2FNet^[26]、SINetV2^[27]、LSR^[12]、UGTR^[28]。为了公平比较,我们直接使用作者在网络上开源的预测图,用相同的公式进行评估。如果缺少预测图,则使用作者提供的预训练完成的模型生成预测图。本文总结了在4个数据集上不同基线模型的定量结果。从表1可以看出,本文方法在不同的数据集上都优于其他模型。

为了进行更广泛的对比,本文使用在目标检测领域的YOLOv5模型进行分割对比实验。在训练开始前,所有图片参照官方代码库(<https://github>).

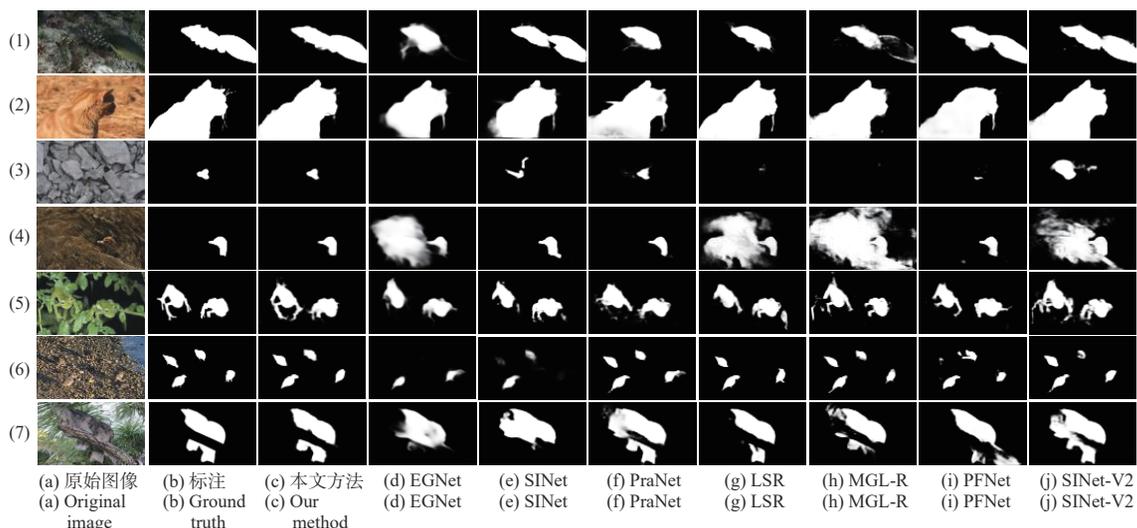


图 5 本文方法与其他方法的视觉对比

Fig. 5 Vision comparison of our method with other methods

表 1 不同模型在 4 个数据集 (CHAMELEON, CAMO-test, COD10k-test, NC4K) 上对 4 种评价指标的定量结果

Tab. 1 Quantitative results of different models for four evaluation metrics on four dataset (CHAMELEON, CAMO-test, COD10k-test, NC4K)

基线模型	CHAMELEON				CAMO-test			
	$S_a \uparrow$	$E_o \uparrow$	$F_{\beta}^w \uparrow$	MAE \downarrow	$S_a \uparrow$	$E_o \uparrow$	$F_{\beta}^w \uparrow$	MAE \downarrow
BASNet	0.687	0.721	0.474	0.118	0.618	0.661	0.413	0.159
EGNet	0.848	0.870	0.702	0.050	0.732	0.768	0.583	0.104
CPD	0.853	0.866	0.706	0.052	0.726	0.729	0.550	0.115
F3Net	0.848	0.917	0.744	0.047	0.711	0.780	0.564	0.109
PraNet	0.860	0.907	0.760	0.044	0.769	0.824	0.663	0.094
SINet	0.869	0.891	0.740	0.044	0.751	0.771	0.606	0.100
PFNet	0.882	0.942	0.810	0.033	0.782	0.852	0.695	0.085
C2FNet	0.888	0.946	0.828	0.032	0.820	0.864	0.752	0.066
SINetV2	0.888	0.942	0.816	0.030	0.820	0.882	0.743	0.070
LSR	0.890	0.948	0.820	0.030	0.787	0.852	0.696	0.080
UGTR	0.888	0.918	0.794	0.031	0.785	0.859	0.686	0.086
YOLOv5m-seg	0.834	0.884	0.767	0.045	0.699	0.711	0.556	0.098
YOLOv5x-seg	0.851	0.902	0.796	0.042	0.740	0.789	0.592	0.087
本文方法	0.892	0.952	0.828	0.027	0.828	0.889	0.817	0.061
基线模型	COD10K-test				NC4K			
	$S_a \uparrow$	$E_o \uparrow$	$F_{\beta}^w \uparrow$	MAE \downarrow	$S_a \uparrow$	$E_o \uparrow$	$F_{\beta}^w \uparrow$	MAE \downarrow
BASNet	0.634	0.678	0.365	0.105	0.695	0.785	0.546	0.095
EGNet	0.737	0.779	0.509	0.056	0.777	0.864	0.639	0.075
CPD	0.747	0.770	0.508	0.059	0.787	0.852	0.696	0.072
F3Net	0.739	0.819	0.544	0.051	0.780	0.848	0.656	0.070
PraNet	0.789	0.861	0.629	0.045	0.822	0.877	0.724	0.059
SINet	0.771	0.806	0.551	0.051	0.808	0.876	0.723	0.058
PFNet	0.800	0.868	0.660	0.040	0.839	0.892	0.745	0.053
C2FNet	0.813	0.890	0.636	0.036	0.838	0.898	0.762	0.049
SINetV2	0.815	0.887	0.680	0.037	0.840	0.907	0.770	0.048
LSR	0.804	0.889	0.673	0.037	0.840	0.907	0.770	0.048
UGTR	0.818	0.850	0.667	0.035	0.839	0.892	0.746	0.052
YOLOv5m-seg	0.715	0.732	0.527	0.048	0.753	0.780	0.634	0.066
YOLOv5x-seg	0.794	0.832	0.590	0.044	0.802	0.868	0.715	0.053
本文方法	0.822	0.890	0.673	0.034	0.850	0.908	0.775	0.044

注:加粗表示最好结果

com/ultralytics/yolov5)的设置被重新调整为 640×640 大小,其他参数使用默认设置。实验选取与本文方法参数量相近的YOLOv5m-seg模型(22.67M)与性能最优的YOLOv5x-seg模型,结果见表1。本文方法在模型大小相近的情况下性能远远领先YOLOv5m-seg模型,与YOLOv5x-seg模型相比有着巨大的优势。

图5展示了本文方法与其他模型的视觉对比结果。可以看出,在不同的具有挑战性的场景下(第1~2行大尺寸伪装目标,第3~4行小伪装目

标,第5~6行模糊边缘),本文方法都能产生优于其他模型的预测图。在目标被部分遮挡的情况下(第7行),该方法也能准确定位目标区域并产生精确的边缘细节。因此,本文方法相比于其他方法在伪装目标分割任务中具有更优秀的性能。另外,本文提供了本文方法与其他11种模型的P-R曲线和F曲线,如图6所示。

本文方法与其他方法在模型复杂度、参数量和实时性上也进行了对比。所有算法在相同的硬件环境下(RTX2080Ti显卡)进行实验。其中

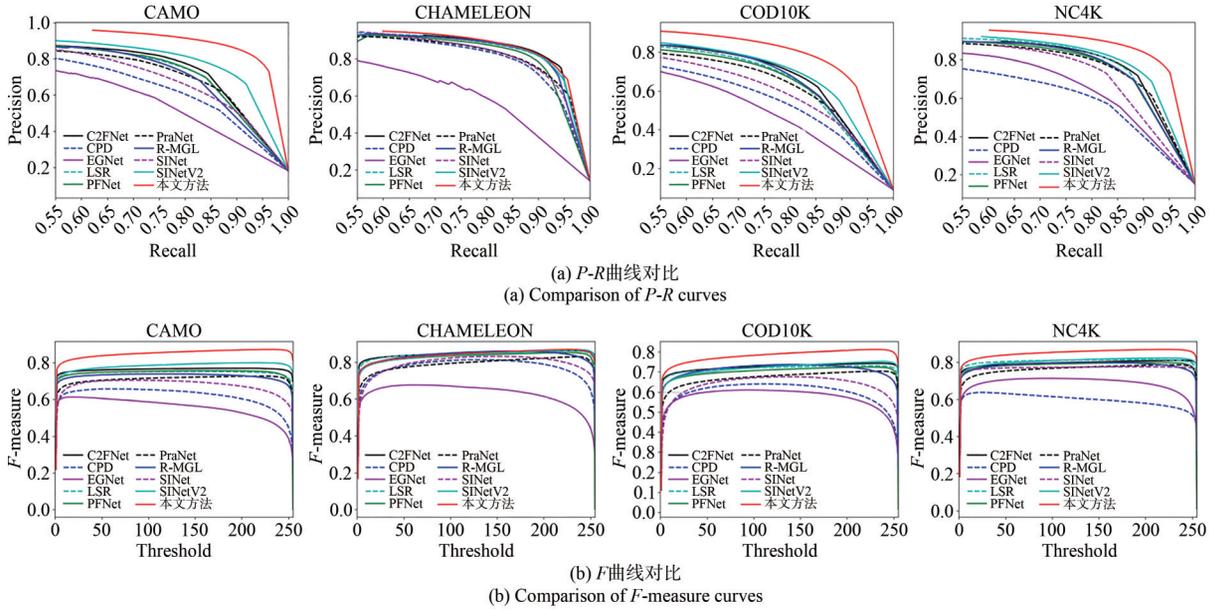


图 6 10 种不同方法在 4 个基准数据集上的 P - R 曲线和 F 曲线,本文方法为红色实线。 P - R 曲线越接近右上角, F 曲线越接近坐标系上部,表示模型的性能越好。

Fig. 6 P - R curves and F -measure curves of 10 different methods on four benchmark datasets. Our method is shown with a solid red line. The closer the P - R curve is to the upper right corner and the higher the F -measure curve is, the better the performance of the model is.

浮点运算次数 (Floating Point Operations, FLOPs) 可用来衡量算法复杂度,为 21.26G;模型参数量 (Parameters, Params) 为 29.47M;FPS (Frame Per

Second) 为 44.2。为了公平比较,所有模型均使用 352×352 的图片计算。如表 2 所示,本文方法在提升准确性的同时也保证了实时性能。

表 2 不同模型的速度和模型复杂度分析

Tab. 2 Speed and model complexity analysis on multiple models

Method	本文方法	SINet	PFNet	UGTR	MGL-R	LSR	SINetV2
FLOPs/G	21.26	38.04	19.00	1 024	549.62	66.64	31.26
Params./M	29.47	48.95	49.50	48.87	63.58	50.95	24.93
FPS	44.2	56	62	18	12	56	25

3.5 消融实验

为了验证每个模块的有效性,本文设计了一系列消融实验,对边缘提取模块、多级特征增强模块 (MFEM)、跨层级特征聚合模块 (CFM)、注意力模块等逐步解耦,以验证其有效性,实验结果见表 3。为了验证损失函数及对应的超参数对网络性能的影响,本文对一系列不同的超参数设置进行了定量评价。

基线模型选择一个类似 U-net^[29] 结构的分割网络,编码器部分为 ResNet-50 网络,解码器逐级上采样并与较浅层特征结合,逐渐恢复到原尺寸。从表 3 可以看出,不同的模块对模型的性能

提升都有贡献:在 U-net 架构基础上加入多尺度特征增强模块和跨层级特征聚合模块后,模型的 4 个评价指标 S_a 、 E_o 、 F_β^w 、MAE 分别提升了 4.8%、3.5%、15.7%、11.1%,证明模型的层内和层间多尺度特征表达能力有了一定加强;在此基础上加入边缘提取模块,4 个指标进一步提升了 1.6%、1.8%、4.2%、10%,说明边缘先验信息在该分割任务中做出了重要的贡献;在加入注意力模块后, S_a 、 E_o 、 F_β^w 分别提升了 0.006、0.018、0.007, MAE 指标则下降了 0.002。

图 7 为逐步解耦各个子模块后的可视化效果对比。从图 7(d) 可以看出,在去除了边缘提取模

表 3 不同模块的有效性分析

Tab. 3 Effectiveness analysis of different modules

边缘提取模块	多尺度特征增强模块	跨层级特征聚合模块	注意力模块	COD10K-test			
				$S_a \uparrow$	$E_o \uparrow$	$F_{\beta}^w \uparrow$	MAE \downarrow
				0.766	0.828	0.552	0.045
	✓	✓		0.803	0.857	0.639	0.040
✓		✓		0.812	0.864	0.660	0.038
	✓	✓	✓	0.812	0.868	0.652	0.038
✓		✓	✓	0.819	0.875	0.664	0.036
✓	✓	✓		0.816	0.872	0.666	0.036
✓	✓	✓	✓	0.822	0.890	0.673	0.034

注:加粗表示最好结果

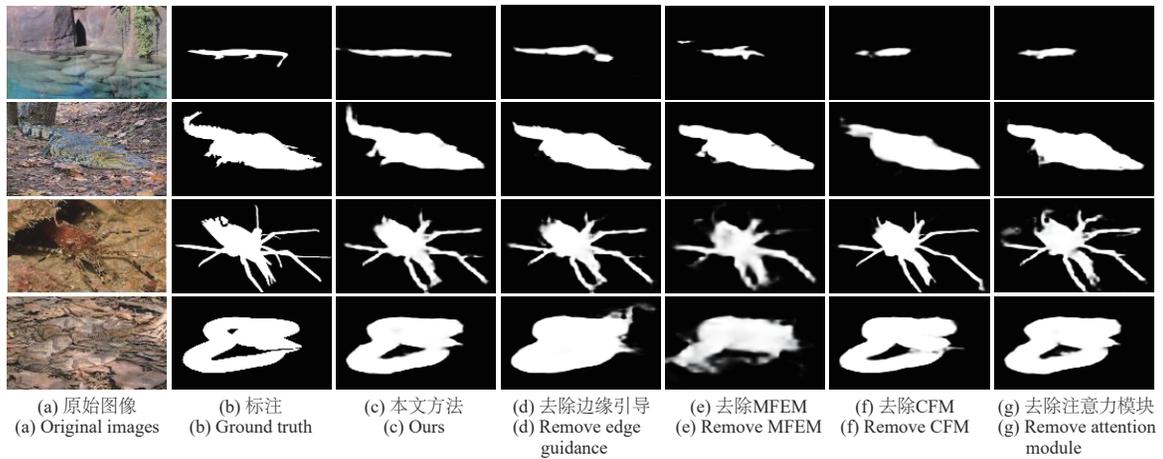


图 7 去除不同模块的视觉比较

Fig. 7 Vision comparison of removed different modules

块后,预测结果的边界存在大量的冗余,一些较复杂的边缘结构难以被清晰地分割,说明边缘提取模块对目标边界像素的提纯至关重要。多尺度的特征更有利于定位复杂场景下的伪装物体,在分别去掉多尺度特征增强模块(图7(e))和跨层级特征聚合模块(图7(f))后,模型不能准确地找到目标所在的位置,出现了目标区域模糊不清、目标结构被错判和伪装区域连通性的问题。从图7(c)和图7(g)可以看出,注意力模块对融合后的特征进一步去噪,使网络更关注于预测目标区域,对背景区域进行抑制,目标细节更明显,置信度较低的噪声干扰被去除。

为了分析联合损失函数各参数对网络的影响,按照不同的比例设置两部分损失函数的超参数(其中 λ_1 为边缘损失 L_{edge} 的比重, λ_2 为预测损失 L_{pred} 的比重),进行对比实验。根据表4,模型在给边缘损失较大权重时表现更好,在 $\lambda_1=5$ 、 $\lambda_2=1$

表 4 两种损失函数的比重对网络性能的影响

Tab. 4 Effect of the proportion of the two loss functions on network performance

λ_1	λ_2	COD10K-test			
		$S_a \uparrow$	$E_o \uparrow$	$F_{\beta}^w \uparrow$	MAE \downarrow
1	1	0.813	0.887	0.660	0.036
2	1	0.820	0.890	0.669	0.034
5	1	0.822	0.890	0.673	0.034
10	1	0.825	0.884	0.670	0.035
1	2	0.807	0.878	0.662	0.035
1	5	0.800	0.871	0.654	0.037

注:加粗表示最好结果

时效果最好。这表明在网络不变的情况下,准确的边缘先验能够极大地提升网络的有效性。

4 结 论

本文提出了一种基于边缘增强和特征融合的

伪装目标分割网络。首先设计了一种边缘提取模块,有效利用低级特征,产生精确的边缘先验。其次,采用多尺度特征增强模块和跨层级特征融合模块,分别提取层内和层间的有效多尺度信息。之后,设计了一种简单有效的层间注意力模块,对充分融合的特征进行再次筛选,去除冗余的背景噪声干扰。最后,将各层预测与边缘先验结合,生

成最后的预测图,并采用联合损失函数对不同尺度的预测图进行联合监督。本文方法在4个伪装目标基准数据集上进行实验,在4种不同的评价指标上都优于其他方法。在视觉对比中,本文方法分割出的预测图能够更好地识别复杂场景下的伪装物体,更好地保留了目标轮廓,细节信息更清晰。因此,本文方法对伪装目标分割有更好的效果。

参 考 文 献:

- [1] PAN Y X, CHEN Y W, FU Q, *et al.* Study on the camouflaged target detection method based on 3D convexity [J]. *Modern Applied Science*, 2011, 5(4): 152-157.
- [2] HARALICK R M, SHANMUGAM K, DINSTEN I H. Textural features for image classification [J]. *IEEE Transactions on Systems, Man, and Cybernetics*, 1973, SMC-3(6): 610-621.
- [3] CANNY J. A computational approach to edge detection [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1986, PAMI-8(6): 679-698.
- [4] YIN J Q, HAN Y B, HOU W D, *et al.* Detection of the mobile object with camouflage color under dynamic background based on optical flow [J]. *Procedia Engineering*, 2011, 15: 2201-2205.
- [5] ZUO C, QIAN J M, FENG S J, *et al.* Deep learning in optical metrology: a review [J]. *Light: Science & Applications*, 2022, 11(1): 39.
- [6] LE T N, NGUYEN T V, NIE Z L, *et al.* Anabranch network for camouflaged object segmentation [J]. *Computer Vision and Image Understanding*, 2019, 184: 45-56.
- [7] FAN D P, JI G P, SUN G L, *et al.* Camouflaged object detection [C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle: IEEE, 2020: 2774-2784.
- [8] MEI H Y, JI G P, WEI Z Q, *et al.* Camouflaged object segmentation with distraction mining [C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Nashville: IEEE, 2021: 8768-8777.
- [9] ZHAI Q, LI X, YANG F, *et al.* Mutual graph learning for camouflaged object detection [C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Nashville: IEEE, 2021: 12992-13002.
- [10] LI A X, ZHANG J, LV Y Q, *et al.* Uncertainty-aware joint salient object and camouflaged object detection [C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Nashville: IEEE, 2021: 10066-10076.
- [11] SKUROWSKI P, ABDULAMEER H, BLASZCZYK J, *et al.* Animal camouflage analysis: CHAMELEON database [R]. *Unpublished Manuscript*, 2018.
- [12] LV Y Q, ZHANG J, DAI Y C, *et al.* Simultaneously localize, segment and rank the camouflaged objects [C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Nashville: IEEE, 2021: 11586-11596.
- [13] HE K M, ZHANG X Y, REN S Q, *et al.* Deep residual learning for image recognition [C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas: IEEE, 2016: 770-778.
- [14] 孙劲光,王雪. 基于目标轮廓的实例分割方法[J]. *液晶与显示*, 2022, 37(4): 519-529.
SUN J G, WANG X. Instance segmentation method based on target contour [J]. *Chinese Journal of Liquid Crystals and Displays*, 2022, 37(4): 519-529. (in Chinese)
- [15] FENG M Y, LU H C, DING E R. Attentive feedback network for boundary-aware salient object detection [C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach: IEEE, 2019: 1623-1632.
- [16] WU Z, SU L, HUANG Q M. Cascaded partial decoder for fast and accurate salient object detection [C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach: IEEE, 2019: 3902-3911.

- [17] SZEGEDY C, VANHOUCKE V, IOFFE S, *et al.* Rethinking the inception architecture for computer vision [C]. IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 2818-2826.
- [18] WEI J, WANG S H, HUANG Q M. F³Net: fusion, feedback and focus for salient object detection [C]//*Proceedings of the 34th AAAI Conference on Artificial Intelligence*. New York: AAAI, 2020: 12321-12328.
- [19] 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: IEEE, 2017: 4558-4567.
- [20] FAN D P, GONG C, CAO Y, *et al.* Enhanced-alignment measure for binary foreground map evaluation [C]//*Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*. Stockholm: ijcai. org, 2018: 698-704.
- [21] MARGOLIN R, ZELNIK-MANOR L, TAL A. How to evaluate foreground maps [C]. IEEE Conference on Computer Vision and Pattern Recognition. Columbus: IEEE, 2014: 248-255.
- [22] 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: IEEE, 2012: 733-740.
- [23] QIN X B, ZHANG Z C, HUANG C Y, *et al.* BASNet: boundary-aware salient object detection [C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach: IEEE, 2019: 7471-7481.
- [24] ZHAO J X, LIU J J, FAN D P, *et al.* EGNet: edge guidance network for salient object detection [C]//*Proceedings of the IEEE/CVF International Conference on Computer Vision*. Seoul: IEEE, 2019: 8778-8787.
- [25] FAN D P, JI G P, ZHOU T, *et al.* PraNet: parallel reverse attention network for polyp segmentation [C]. 23rd International Conference on Medical Image Computing and Computer-Assisted Intervention. Lima: Springer, 2020: 263-273.
- [26] SUN Y J, CHEN G, ZHOU T, *et al.* Context-aware cross-level fusion network for camouflaged object detection [C]//*Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence*. Montreal: ijcai. org, 2021: 1025-1031.
- [27] 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.
- [28] YANG F, ZHAI Q, LI X, *et al.* Uncertainty-guided transformer reasoning for camouflaged object detection [C]//*Proceedings of the IEEE/CVF International Conference on Computer Vision*. Montreal: IEEE, 2021: 4126-4135.
- [29] ZHOU Z W, SIDDIQUEE M M R, TAJBAKSH N, *et al.* UNet++: redesigning skip connections to exploit multiscale features in image segmentation [J]. *IEEE Transactions on Medical Imaging*, 2020, 39(6): 1856-1867.

作者简介:



李明岩,男,硕士研究生,2018年于长春大学获得学士学位,主要从事计算机视觉、深度学习方面的研究。
E-mail:1019952730@qq.com



吴川,男,博士,研究员,2005年于中国科学院长春光学精密机械与物理研究所获得博士学位,主要从事图像融合、目标跟踪、嵌入式系统开发等方面的研究。E-mail:wuchuan0458@sina.com