渐进聚合多尺度场景上下文特征的伪装物体检测

刘 研'张开华" 樊佳庆" 赵雅倩" 刘青山"

¹⁾(南京信息工程大学自动化学院 南京 210044)

2)(南京信息工程大学计算机与软件学院数字取证教育部工程研究中心 南京 210044)

³⁾(南京航空航天大学计算机科学与技术学院 南京 210016)

4)(苏州浪潮智能科技有限公司 江苏 苏州 215101)

摘 要 伪装物体检测通过模仿人类的视觉检测机理,实现在复杂场景下对伪装物体的定位与识别.然而,多数伪 装物体检测方法在遇到相似外形目标干扰时,仅通过目标的局部表观特征无法准确识别伪装目标.为此,本文提出 一种渐进聚合多尺度场景上下文特征的伪装物体检测网络,通过聚合多阶段语义增强的场景上下文特征来实现准 确的伪装物体判别.具体来说,所提网络主要包含两个创新设计:U型上下文感知模块和跨级特征聚合模块.前者 旨在感知复杂场景中物体的细节轮廓、纹理特征和颜色变化等丰富的局部-全局场景上下文信息.后者则结合坐标 方向的注意力和多层级残差渐进特征聚合机制,逐级渐进聚合相邻层级之间的互补特征,实现对伪装物体全局语 义的强化和局部细节的补充.本文方法在 CHAMELEON、CAMO-Test、COD10K-Test 和 NC4K 等4 个非常具有挑 战性的基准数据集上进行了评测.评测结果表明,本文方法相比于最新方法达到了领先的性能.

关键词 伪装物体检测;场景上下文;深度学习:注意力机制 中图法分类号 TP391 DOI号 10.11897 SP.J.1016.2022.02637

Progressively Aggregating Multi-Scale Scene Context Features for Camouflaged Object Detection

LIU Yan¹⁾ ZHANG Kai-Hua²⁾ FAN Jia-Qing³⁾ ZHAO Ya-Qian⁴⁾ LIU Qing-Shan²⁾

 ¹⁾ (School of Automation, Nanjing University of Information Science and Technology, Nanjing 210044)
 ²⁾ (School of Computer and Software, Engineering Research Center of Digital Forensics, Ministry of Education, Nanjing University of Information Science and Technology, Nanjing 210044)

³⁾ (College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 210016) ⁴⁾ (Inspur Suzhou Intelligent Technology Co., Ltd, Suzhou, Jiangshu 215101)

Abstract Camouflaged object detection (COD) is a computer vision task that imitates human visual mechanisms to recognize and locate camouflaged objects in complex scenes. However, the current COD methods cannot accurately discriminate the camouflage objects only by the local appearance features of the objects when meeting distractors with similar appearances. To this end, this paper presents a COD network based on progressively aggregating multi-scale scene context features, so that the accurate camouflaged object discrimination is realized by aggregating multi-stage semantic enhanced scene context features. Specifically, the network mainly has two novel designs: U-shape Context-Aware Module(UCAM) and Cross-level Feature Aggregation Module(CFAM). The UCAM aims to sense rich local to global context information such as detailed boundaries, texture features, and color changes of camouflaged objects. The CFAM

收稿日期:2022-03-14;在线发布日期:2022-09-29.本课题得到科技创新 2030-"新一代人工智能"重大项目(2018AAA0100400)、国家自 然科学基金项目(62276141,61825601)、江苏省 333 工程人才项目(BRA2020291)资助.刘研,硕士研究生,主要研究方向为伪装物体 检测.E-mail: liuyan0698@163.com. 张开华(通信作者),博士,教授,中国计算机学会(CCF)会员,主要研究领域为视觉目标跟踪、协同 显著性检测.E-mail: zhkhua@gmail.com. 樊佳庆,博士研究生,主要研究方向为视频目标分割.赵雅倩,高级工程师,中国计算机学会 (CCF)高级会员,主要研究领域为人工智能.刘青山,博士,教授,中国计算机学会(CCF)高级会员,主要研究领域为计算机视觉和机器学习.

combines the coordinate direction attention and the multi-level residual progressive feature aggregation mechanism to gradually aggregate complementary features between adjacent levels, strengthen the global semantics of camouflage objects and supplement local details. Extensive evaluations on 4 extremely challenging benchmarks including CHAMELEON, CAMO-Test, COD10K-Test, and NC4K, the experimental results have demonstrated that our model has achieved leading performance compared with state-of-the-art methods.

Keywords camouflaged object detection; scene context; deep learning; attention mechanism

1 引 言

自然界中某些生物为了更好的隐藏自己而不被 天敌捕食,往往会选择将自己融入到周围环境之中, 这种技能被称为"伪装".最近伪装物体检测(Camouflaged Object Detection,COD)。成为了计算机视觉 领域的研究热点之一.在 COD 中,能够快速、准确 地提取伪装物体是一项极具挑战性的任务,其在息 肉分割^[2]、肺部感染分割^[3]、工业产品瑕疵检测^[4]等 领域具有广泛的应用价值.

COD 与传统的显著物体检测(Salient Object Detection,SOD)^[5-12]之间存在共性和特性关联.其共性在于都是对图像中前景物体和背景之间的检测与分割.而特性则是因为"伪装"与"显著"本身是一对相反的概念.SOD 更多关注于显著的物体与区域,而 COD 则关注隐藏在背景下的物体.所以,SOD 的物体和背景之间的全局-局部对比度很高,而 COD 则相反.因此,COD 比 SOD 更具挑战性.长期以来一直有大量致力于 COD 方法的研究.其中,早期方法^[13-15]主要通过手工设计的特征(如纹理、颜色、强度等)来区分伪装/非伪装物体.然而,这类特征区分伪装/非伪装的能力十分有限,泛化性较弱,仅适用于相对简单的场景.

近年来,随着深度学习的兴起,基于深度学习的 COD 方法在性能方面取得了突破性进展,如最新的 方法^[16-19]在四个标准 COD 数据集上都取得了领先 的性能. Zhai 等人^[19]考虑了边缘先验信息,设计了 基于图的互学习 COD 网络 MGL,并以循环的方式 交互捕获边缘和语义信息;Mei 等人^[18]将分心的概 念引入 COD 任务,提出了基于分心挖掘的 COD 网 络 PFNet. PFNet 首先定位潜在物体,然后,逐步聚 焦发现并去除分心区域来细化检测结果;Lv 等 人^[17]提出了可以同时对伪装程度不同的物体进行 定位、分割和排序的网络 LSR;Yang 等人^[16]结合概 率表示模型与 Transformer,设计了基于不确定性

引导 Transformer 推理的 COD 模型 UGTR.

相比于传统方法而言,虽然上述方法极大提高 了检测性能.但是,由于伪装物体与背景之间存在内 在的视觉相似性,仅靠目标的局部表观特征难以准 确区分它们.此外,现实中存在大量极具挑战性的场 景(如图1所示,实际场景中存在颜色高度相似、外 观细长、伪装目标歧义等挑战).这些挑战主要由表 观信息不足、局部细节丢失、目标语义歧义等原因产 生,导致最先进的方法[16-19]也难以准确识别复杂场 景下的伪装物体.为了应对这类挑战场景下的 COD, 一方面,需要利用更多的低级特征来捕获丰富的纹 理、颜色、边缘等信息,然而,产生低级特征的网络层 通常较浅,无法获得较大感受野,导致难以捕获到充 足的场景上下文信息;另一方面,需要充分利用多级 互补特征.通常,网络为了避免参数量过大,需要对 特征映射层不断下采样来获得多级特征,然而,这极 易稀释部分特征,导致模型在实际推理过程中丢失 局部信息,从面会对小目标伪装物体的检测产生严 重影响.

针对上述问题,本文提出了一种渐进聚合多尺 度场景上下文特征的 COD 网络.具体而言,主要设计 了两个关键模块,即 U 型上下文感知模块(U-shape Context-Aware Module,UCAM)和跨级特征聚合 模块(Cross-level Feature Aggregation Module, CFAM).UCAM 从局部到全局、从小尺度到大尺度 全面地挖掘多级特征中潜在的伪装物体场景上下文 信息,而 CFAM 则使用残差渐进聚合的方式充分地 捕获相邻层级之间的互补信息,有效弥补了小物体 特征被稀释、局部细节丢失以及由于缺乏全局语义 引导而导致的目标物体歧义的问题,从而由粗到细 地逐步细化得到精确的预测结果.综上所述,本文的 主要贡献总结如下:

(1)本文提出了一种新颖的 COD 网络,综合考虑了互补的跨级特征和丰富的局部-全局场景上下 文信息.

(2)本文设计了一个U型上下文感知模块

UCAM,通过使用并行级联搜索块(Parallel Cascaded Search Block, PCSB)和多次池化下采样编码后再逐级上采样级联聚合解码来捕获丰富的场景上下文信息.

(3)本文提出了一个跨级特征聚合模块 CFAM,

结合残差渐进特征聚合机制由粗到细地聚合相邻层 级之间的互补特征.

(4)本文模型在四个具有挑战性的测试数据集 上均表现优异,充分证明了其在 COD 任务上的有 效性.



图 1 伪装物体检测的视觉样例.现有最先进的方法 UGTR^[16]、LSR^[17]、PFNet^[18]、MGL^[19]无法准确识别出颜色高度相似物体、外观细小的物体(红色方框为细小物体)与伪装目标歧义物体(红色与黄色方框分别表示真实伪装物体与歧义伪装物体), 而本文的方法可以解决这些问题,产生准确的预测结果

2 相关工作

2.1 伪装物体检测

早期方法^[13-15]主要根据图像的纹理、颜色、强度 等手工提取特征来区分伪装的前景物体和背景. Bhajantri等人^[13]设计了一种利用共生矩阵捕获伪 装物体纹理特征的 COD 模型. Zhang 等人^[14]基于 颜色对比的思想提出了一种全局和局部建模相结合 的 COD 模型. 此外,输入强度也可以作为辅助特征 检测伪装物体,如 Pan 等人^[15]设计了一种基于三维 凸性的 COD 方法,通过设计相应的算子来反映图 像中每个像素的灰度,利用灰度值确定合适的阈值 以限制检测区域,从而确定伪装目标的位置并将其 从复杂背景中分离出来.

近年来,基于深度学习的方法^[1,16-19]成为 COD 领域的主流.Fan 等人^[1]构建了一种目前最具挑战 性的大规模伪装物体数据集 COD10K,进一步推动 了深度学习在 COD 任务上的发展,Fan 等人^[1]还开 发了一个简单有效的 COD 网络 SINet,包含用于扩 大感受野的搜索模块和细化预测结果的识别模块. Zhai 等人^[19]提出一种基于图的互学习 COD 网络 MGL.MGL 利用区域诱导图推理模块来挖掘高级 语义引导信息,并配合边缘约束图推理模块提供的 边缘先验信息增强 COD 的底层表征.最后,以循环 的方式交互吸收两个模块之间有价值的信息.Mei 等人^[18]提出了一种基于分心挖掘的 COD 算法 PF-Net:先通过基于自注意力机制设计的定位模块来 定位获取伪装物体的语义信息,再使用聚焦模块滤除冗余干扰,逐步细化当前级特征、上一级特征和多级预测结果.Lyu等人^[17]设计了一种同时对伪装程度不同的物体进行定位、分割和排序的 COD 方法LSR,并提出了一个目前最大的伪装物体测试数据集NC4K.Yang等人^[16]结合概率表示模型与 Transformer,设计了一种基于不确定性引导 Transformer 推理的 COD 模型 UGTR.首先学习一个具有条件概率分布的主下网络输出,用于产生伪装图像的不确定性信息,然后,结合不确定性信息引导 Transformer 推理产生最终的预测结果.

2.2 上下文感知学习

上下文信息在像素级检测任务中起着重要作 用,长期以来有大量工作致力于利用此信息来增强 特征的表达能力,Chen等人^[20]和 Zhao等人^[21]都通 过在主干网络的最后一层建立多尺度像素级表示来 获得丰富的上下文信息.Liu等人^[22]通过并行使用 多个不同大小、不同空洞率的卷积操作来扩大感受 野以获得多尺度上下文信息.Fu等人^[23]使用非局 部模块来捕获上下文的长范围依赖关系.Chen等 人^[24]利用全局上下文模块将不同阶段的特征和全 局语义特征进行聚合.然而,伪装物体与背景之间具 有内在相似性,需要捕获更充分的场景上下文信息 来强化它们之间的表征差异信息.为此,本文设计了 具有更强感知能力的场景上下文感知模块.

2.3 多层级特征聚合

为了充分地利用高层语义信息和低层细节信

息,Lin 等人^[25]提出了特征金字塔网络 FPN,采用 自底向上的方式逐渐聚合不同层次的特征,巧妙地 结合语义与细节信息.此外,考虑到具有较大差异的 特征直接采用简单相加的方式聚合易引入噪声, Wei 等人^[26]设计了一个交叉特征聚合模块来聚合 相邻层级的特征,有效地避免了因为引入过多冗余 信息而破坏原始特征的问题. Ma 等人[27] 通过逐层 收缩聚合相邻特征信息的方式开发了一种金字塔特 征收缩网络,能够更加有效地聚合语义和细节信息. 上述方法探索了不同思路的特征聚合方式,并设计 了相应的特征聚合模块.尽管这些设计在一定程度 上提升了特征聚合的效率和效果. 然而, 它们对于空 间信息的捕捉能力仍然不足,存在聚合后部分特征 被弱化而导致有用信息丢失的问题.本文针对以上 存在的问题进行了探索,致力于设计出更加高效的 特征聚合机制与模块.

3 本文方法

如图 2 所示,本文提出的网络是一种经典的编码-解码结构.给定一张含有伪装物体的 RGB 图像 $T \in \mathbb{R}^{3 \times H \times W}$.首先,将其送入 Res2Net50^[28]骨干网 络中提取多级特征 F_k ,k=1,…,5,其中, F_k 为第 k 层特征,并将最后四层特征通道压缩到 64.然后,在 F_5 层特征上使用空洞空间金字塔池化(Atrous Spatial Pyramid Pooling, ASPP)来获取含有粗略物体 定位信息的全局语义特征.同时,在 F_2 , F_3 , F_4 层特征上使用 UCAM 来捕获富含局部细节的多尺度场 景上下文信息.最后,CFAM 采用残差渐进的方式 聚合多层级相邻层级之间的互补信息,由粗到细地 细化得到精确的预测结果.技术细节将在 3.1 节、3.2 节、3.3 节、3.4 节中分别介绍.



图 2 网络的整体结构图

3.1 空洞空间金字塔池化

输入的伪装物体图像经过五层残差块编码后会 得到具有强语义表征能力的特征 F₅,其主要反映了 潜在伪装物体在图像中的大致位置信息,这对于能 否准确检测出伪装物体至关重要.为了充分利用这 种特性,本文使用 Chen 等人^[20]提出的 ASPP 模块

2641

结合图像级特征来捕获伪装物体全局多尺度语义信息.具体而言,如图 2 所示,将通道降维后的特征 $F_5 \in \mathbb{R}^{64 \times H/32 \times W/32}$ 同时送入 5 个并行分支 l_i ,i=1,...,5 中,其中 l_1 分支使用1×1 的卷积操作, l_5 分支使用全局平均池化操作,剩下三个分支均使用大小为 3×3, 空洞率分别为 d=3,5,7 的卷积操作.随后,将 5 个分支的输出在通道维进行拼接,并利用 1×1 的卷积操作将拼接后特征降维到与输入同样维度大小的特征 $H_5 \in \mathbb{R}^{64 \times H/32 \times W/32}$.

3.2 U型上下文感知模块

通常,伪装物体与背景在纹理、边缘、颜色等方 面具有很强的内在相似性.所以,伪装物体的局部-全局对比度较小,精确检测这些伪装物体需要充足 的低层信息作为支撑.然而,每层特征图上的感受野 大小是固定的,使得网络关注到的场景信息往往会 局限在一定范围之内,这导致网络无法结合场景上 下文信息进行推理.但是,主干网络提取的多级特征 中的低层特征蕴含丰富的细节信息,为了能够捕获 到物体的局部与全局特征之间的细微差异,受Qin等 人^[29]提出的残差 U 型块(Residual U-blocks, RSU) 启发,本文设计了U型上下文感知模块(UCAM)用于 捕获主干网络提取的 F_2 , F_3 , F_4 中丰富的场景上下 文信息. UCAM_n (n=2,3,4)的结构如图 3 所示,其 形状呈 U 型,类似编码-解码的结构.其中,n 表示在 编码-解码过程中下采样(上采样)的次数.对通道压 缩后的特征 $F_k \in \mathbb{R}^{64 \times H/m \times W/m}$ 分别送入 UCAM_n后得 到输出特征 $H_k \in \mathbb{R}^{64 \times H/m \times W/m}$,其中,m = 4, 8, 16, m指第 k 层残差块编码特征下采样的倍率.

以下以 $F_2 \in \mathbb{R}^{64 \times H/4 \times W/4}$ 为输入来介绍 UCAM₃ 的过程,其余两个 UCAM 均参考如下流程.首先, 输入特征分别经过三次 PCSB 和 Maxpooling 交替 进行上下文信息搜索编码和下采样操作得到编码后 的特征 $F_2^3 \in \mathbb{R}^{64 \times H/32 \times W/32}$:

$$\begin{cases} F_{2}^{1} = f_{\text{down}}(f_{\text{PCSB}}^{1}(F_{2})), \\ F_{2}^{2} = f_{\text{down}}(f_{\text{PCSB}}^{2}(F_{2}^{1})), \\ F_{2}^{3} = f_{\text{down}}(f_{\text{PCSB}}^{3}(F_{2}^{2})), \end{cases}$$
(1)

式中, $f_{PCSB}(\cdot)$, $i=1, \dots, 3$ 为第 i 次 PCSB 编码, $f_{down}(\cdot)$ 为 Maxpooling 操作. 再将编码后的特征使 用双线性插值上采样到与输入特征相同大小. 然后, 通过残差连接的方式聚合之前多次下采样的多尺度 特征,并使用一个大小为 3×3 的卷积操作细化聚 合后的特征. 最后,得到解码后的输出的特征 $H_2 \in \mathbb{R}^{64 \times H/4 \times W/4}$:



$$H_{2}^{1} = f_{CBR}(f_{up}(f_{CBR}(F_{2}^{3})) + f_{PCSB}^{3}(F_{2}^{2})),$$

$$H_{2}^{2} = f_{CBR}(f_{up}(H_{2}^{1}) + f_{PCSB}^{2}(F_{2}^{1})),$$

$$H_{2} = f_{CBR}(f_{up}(H_{2}^{2}) + f_{PCSB}^{1}(F_{2})),$$
(2)

式中, $f_{CBR}(\bullet)$ 指卷积、批归一化(BN)和 ReLU 激活函数的组合操作, $f_{up}(\bullet)$ 为上采样操作.

图 3 中的 PCSB 由四个并行的上下文信息搜索 分支构成,每个分支包含用于通道缩减的 1×1 卷积 操作和用于局部信息搜索的大小分别为 3×3、5× 5、7×7 的卷积操作,每个卷积操作都包含 BN 层和 ReLU.其中,对于分支一,特征被级联送到分支二 中使用更大的卷积核进一步提取特征.以此类推,完 成四个分支之间级联信息交互,使得更高一级的分 支能够获得更大的感受野.最后,将四个并行分支提 取的特征在通道维度上进行拼接得到最终输出. PCSB 编码随着下采样深度越来越深,感受野越来越 大,捕获到的全局信息也越来越丰富.然后,UCAM 在全局信息引导下渐进上采样,并使用残差连接聚 合下采样时的多尺度信息,从而在获得全局感知的 同时避免了部分局部细节信息的丢失.

然而,对于输入的同一张伪装图像,UCAM 在



图 4 不同多级特征使用 U 型上下文感知模块的输入、输出和中间层特征可视化对比

感知它的不同级特征时发挥的作用也各有侧重.针 对特征 F₂与 F₃而言,它是蕴含丰富细节信息的低 层特征,UCAM 侧重于对边缘细节信息的感知与增 强,对比特征可视化图 4 中的特征 F₂、F₃与 F₃、H₃ 可以观察到,随着 UCAM 的逐步下采样再上采样, 全局感受野不断扩大,逐渐强化了边缘细节的表征. 针对特征 F₄而言,它是具有全局语义的深层特征. UCAM 则侧重于引导全局语义信息的表征,对比特 征可视化图 4 中的特征 F₄与 H₄可以观察到,UCAM 下采样过程中进一步扩大了感受野范围,更加有利 于全局信息的搜索与聚焦.此外,对于小目标物体而 言,特征可视化图 3 中还展示了 UCAM 对于小目 标物体的感知过程,可以观察到,聚合下采样和上采 样的多尺度信息后,并没有稀释掉小目标的特征,反 而显著强化了语义表征.

3.3 跨级特征聚合模块

低级特征往往蕴含丰富的细节信息,而高级特 征则具有一致的语义信息和清晰的前景界限.因此, 这两种特征优势互补.为了充分发挥高低级特征的 优势,本文设计了 CFAM(见图 5)来高效聚合不同 空间尺度的特征.给定相对而言的高级特征 X^h和低 级特征 Xⁱ,首先,使用双线性插值将 X^h上采样到和 Xⁱ一样大小.然后,通过逐元素相加将两种原始特 征聚合,并使用一个大小为 3×3 的卷积操作细化得 到初步融合特征 X^{add}:

$$X^{\text{add}} = f_{\text{CBR}}(X^h + X^l) \tag{3}$$

X^{add}集中了两种不同层级特征的优势,并通过相加操作进一步强化了共有部分.

为了充分捕获特征的空间结构信息,受 Hou 等 人^[30]提出的坐标注意力启发,CFAM 将特征分别沿 着水平和垂直坐标方向进行特征编码,即将特征图 沿着水平和垂直两个方向做池化,从而生成一对具 有方向感知的特征映射 z^{*}, z⁹:

$$\begin{cases} z^{x} = f_{X-\text{GAP}}(X^{\text{add}}), \\ z^{y} = f_{Y-\text{GAP}}(X^{\text{add}}) \end{cases}$$

$$\tag{4}$$

式中, $f_{x-GAP}(\cdot)$, $f_{Y-GAP}(\cdot)$ 表示水平和垂直方向的 平均池化操作.

然后,将两个方向的特征映射沿着空间维度拼



图 5 跨级特征聚合模块

接,再使用一个1×1卷积操作获得两个方向的合成 注意力特征 X^{eat}:

$$X^{\text{cat}} = f_{\text{CBh}}(f_{\text{S-Cat}}[z^x; z^y])$$
(5)

式中, f_{s-Cat} [•;•]指空间维度拼接操作, f_{CBh} (•)指 卷积、BN和h-swish^[31]激活函数的组合操作,其中

$$h\text{-swish}(x) = x \frac{\text{Re}LU(x+3)}{6}$$
(6)

然后,将输出结果重新按照水平和垂直方向划 分为两组特征,再分别使用1×1卷积操作获得水平 和垂直方向的坐标加权注意力特征 z₄,z₄:

$$z_{\omega}^{x}, z_{\omega}^{y} = f_{\rm CS}(f_{\rm Split}(X^{\rm Cat})) \tag{7}$$

式中, f_{split}(•)指特征分块操作, f_{cs}(•)指卷积操 作,包括一个 Sigmoid 激活函数.得到的加权特征再 分别和原始的高低特征图相乘,进一步强化两个层 级上伪装物体的语义细节信息.

然后,分别使用3×3的卷积细化相乘后的特

征. 最后,将两种特征相加后得到最终输出特征: $X^{\text{ngg}} = f_{\text{CBR}}(z_{\omega}^{x} \otimes z_{\omega}^{y} \otimes X^{l}) + f_{\text{CBR}}(z_{\omega}^{x} \otimes z_{\omega}^{y} \otimes X^{h})$ (8)

3.4 多层级残差渐进特征聚合机制

多级特征聚合可充分提取不同特征之间的互补 信息。先前的工作^[24-26]主要通过简单的相加或拼 接操作来聚合多尺度信息。这类直接聚合策略将导 致差异大的特征聚合后易产生噪声.为此,本文提出 了结合 CFAM 的多层级残差渐进特征聚合机制,图 6 展示了不同场景下伪装物体的多级特征残差渐进 聚合过程.示意图如图 7 所示,通过跨层渐进聚合从 高层语义到低层细节的多级特征,有效避免了因特 征尺度跨度大而产生的特征冗余.同时,本方法还采 用残差连接来避免聚合过程中的信息损失.从中可 以观察到,在聚合过程中物体的全局语义不断被强 化,局部细节趋于完善.







图 7 多层级残差渐进聚合机制示意图

具体来说, 新进特征聚合分为三个阶. 第一阶段: 首先, 将特征 H₅上采样到与特征 H₄相同尺度 大小,由 CFAM 得到聚合后的特征 X¹⁴⁸⁸. 然后, 将特 征 H₅上采样得到作为下一阶段聚合使用的残差特 征 X^{1es}; 第二阶段: 首先, 将第一阶段聚合后的特征 X¹⁴⁸⁵与残差特征 X^{1es}相加并上采样到与特征 H₃相同 尺度大小,由 CFAM 得到聚合特征 X¹⁴⁸⁶2. 然后,将第 一阶段聚合后的特征 X¹⁴⁸⁶与残差特征 X^{1es}相加并上 采样得到作为下一阶段聚合使用的残差特征 X^{res}; 第三阶段:首先,将第二阶段聚合后的特征 X^{res}与残 差特征 X^{res}相加并上采样到与特征 H₂相同尺度大 小,由 CFAM 得到聚合特征 X^{res}₃. 然后,将第二阶段 聚合后的特征 X^{res}与残差特征 X^{res}₂ 相加并上采样得 到作为最终的残差特征 X^{res}₃. 最后,将特征 X^{res}₃ 与残 差特征 X^{res}相加并上采样 4 倍到与原图相同尺度大 小,得到最终的输出特征 X^{out}:

$$\begin{cases}
X_{1}^{agg} = f_{CFAM}[f_{up}(H_{5}); H_{4}], \\
X_{1}^{res} = f_{up}(H_{5}), \\
X_{2}^{agg} = f_{CFAM}[f_{up}(X_{1}^{agg} + X_{1}^{res}); H_{3}], \\
(10)
\end{cases}$$

$$\begin{array}{l} (X_2^{\text{res}} = f_{\text{up}}(X_1^{\text{res}} + X_1^{\text{res}}) \\ [X_3^{\text{agg}} = f_{\text{CFAM}} [f_{\text{up}}(X_2^{\text{agg}} + X_2^{\text{res}}); H_2], \end{array}$$

$$\int X_3^{\text{res}} = f_{\text{up}}(X_2^{\text{agg}} + X_2^{\text{res}}),$$
 (11)

$$X^{\rm out} = f_{\rm up} (X_3^{\rm agg} + X_3^{\rm res})$$
(12)

式中, f_{CFAM}[•;•]指 CFAM 操作.

3.5 损失函数

本文模型共有四个输出结果,其中,前三个输出 分别是 CFAM 每聚合一次后加上残差特征的输出, 最后的输出是 CFAM 聚合三次后加上残差特征再上 采样到的与原图一样大小的输出.本文所有输出结果 均使用统一的损失函数来监督训练,对于前三个输 出,采用将真实值下采样到与输出预测值相同大小的 方式来监督训练,因此,构造如下损失函数^[26]:

 $\mathcal{L}(\mathcal{P},\mathcal{G}) = \mathcal{L}^{u}_{low}(\mathcal{P},\mathcal{G}) + \omega \mathcal{L}_{ME}(\mathcal{P},\mathcal{G})$ (13) 式中, \mathcal{P} 和 \mathcal{G} 分别代表输出预测值和真实值, \mathcal{L}^{u}_{low} 和 \mathcal{L}_{BCE} 分别代表加权交并比(Intersection of Union, IoU) 损失和二值交叉熵(Binary Cross-Entropy, BCE)损 失. 加权损失能够从像素级上对全局和局部结构进 行有效约束,并更关注于难挖掘像素样本.

最后,本文方法的总体损失函数定义如下:

$$\mathcal{L}_{\text{total}} = \sum_{i=1}^{4} \lambda_i \ \mathcal{L}_i(\mathcal{P}, \mathcal{G})$$
(14)

式中,λ_i为不同损失函数的权重系数.

4 实验设置与结果分析

4.1 数据集

本文在 4 个基准数据集上对算法模型进行了评估:CHAMELEON^[32]包含 76 张通过谷歌搜索"伪装"关键词得到的经过人工逐像素标注好的高分辨率图像. CAMO^[33]包含 1250 张伪装物体图像. 其中,1000 张用于训练,250 张用于测试. COD10K^[1] 是目前最具挑战性的大规模 COD 数据集,包含 5 个大类以及 69 个子类,共 5066 张图片,其中,3040 张用于训练,2026 张用于测试. NC4K^[17]是目前最大的 COD 测试数据集,包含 4121 张从互联网下载并经过人工标注的伪装图像.本文模型使用的训练集共计 4040 张图片,分别来自 CAMO(1000 张)和 COD10K(3040 张)的组合.

4.2 评估方法

本文使用平均绝对误差(MAE)^[34]、平均 E-mea-

sure(E_{δ})^[35]、结构度量(S_{a})^[36]以及平均 F-measure (F_{β})^[37]和加权 F-measure(F_{β}^{ω})^[38]作为评价指标.其中,MAE 计算预测值和真实值之间的逐像素平均绝对误差:

$$MAE = \frac{1}{H \times W} \sum_{i=1}^{W} \sum_{j=1}^{H} |\mathcal{P}(i,j) - \mathcal{G}(i,j)|$$
(15)

式中, *P*指预测值, *G*指真实值, *W*和*H*分别代表宽和高, MAE的值被正则化在[0,1]区间内.

*E*_{\$}用于获取图像水平统计信息及其局部像素 匹配信息:

$$E_{\phi} = \sum_{i=1}^{W} \sum_{j=1}^{H} \phi_{FM}(i,j)$$
(16)

式中, *φ*_{FM}表示增强对齐矩阵, 其用来评估伪装物体 预测结果的整体和局部的精确度.

*S*_a着重评估预测图的结构信息,旨在评估区域 感知 *S*_r和目标感知 *S*_a之间的结构相似性:

$$S_{\alpha} = \alpha \times S_o + (1 - \alpha)S_r \tag{17}$$

式中,α是平衡参数,其默认值为 0.5.

F^{*}_β通过对准确率与查全率设置不同的权重来 评估计算:

$$F_{\beta}^{\omega} = \frac{(1+\beta^2) \times Recall^{\omega} \times Precison^{\omega}}{\beta^2 \times Precison^{\omega} + Recall^{\omega}} \quad (18)$$

式中, β设置为 0.3.

4.3 实现细节

本文模型使用 Pytorch^[39]框架来实现,训练和 测试都在一张 GeForce RTX 2080Ti GPU上进行. 模型训练阶段,输入图像的大小被调整为416×416, 并使用随机载剪、水平翻转和颜色增强进行数据扩充,批大小设定为16,初始学习率为1e-4,训练周期 为100,每经历50个周期学习率将除以10衰减一次,优 化器采用 Adam^[40].在模型测试阶段,本文不使用任 何后处理手段,只是将输入图片大小调整为416× 416,然后送入模型进行推理,最后将输出使用双线性 插值调整到原图大小.式(14)中的λ_i分别设置为0.4、 0.6、0.8、1.0.

4.4 与其他算法的性能比较

为了证明本文方法的有效性,本文将所提方法 与13个最新的基于深度学习的方法进行了对比,包 括FPN^[25]、PSPNet^[21]、PiCANet^[6]、UNet++^[41]、 CPD^[8]、BASNet^[7]、EGNet^[9]、PraNet^[2]、SINet^[1]、 MGL^[19]、PFNet^[18]、LSR^[17]、UGTR^[16].为了公平比 较,上述方法的所有预测图均来自文献[1]或相应的 论文和开源代码推理生成.此外,所有预测图都是使 用同一评测代码进行评估.

定量比较:表1和表2列出了不同方法在四个

测试数据集的定量指标比较结果.从表中结果可以 看出,本文方法的各项指标均表现优异,在 CAMO-Test、COD10K-Test、NC4K 上全面超越最先进的模型.在数量最少的数据集 CHAMELEON 上 S_a和F_a 指标表现略显不足,但差距甚微. F_{β}^{*} 在四个测试集 上分别提升了 0.7%、5.4%、3.8%、2.1%; E_{*} 分别 提升了 0.2%、1.8%、1.2%、0.9%;而 MAE 则分别 下降了 1%、10%、4%、4%.

表 1 不同方法在 CHAMELEON 和 CAMO-Test 测试数据集的指标结果比较

年八		CHAME	ELEON(76	张图像)			CAMO	-Test(250	张图像)	
平田 -	S_{α} \blacklozenge	E_{Φ} \uparrow	F^{ω}_{β} \blacklozenge	F_{β} \uparrow	$M \not \bullet$	S_{α} \uparrow	E_{Φ} \uparrow	F^{ω}_{β} \blacklozenge	F_{β} \uparrow	$M \not \downarrow$
CVPR2017	0.794	0.783	0.590	0.648	0.075	0.684	0.677	0.483	0.676	0.131
CVPR2017	0.773	0.758	0.555	0.630	0.085	0.663	0.659	0.455	0.520	0.139
CVPR2018	0.769	0.749	0.536	0.618	0.085	0.609	0.584	0.356	0.573	0.156
DLMIA2018	0.695	0.762	0.501	0.557	0.094	0.599	0.653	0.392	0.460	0.149
CVPR2019	0.687	0.721	0.474	0.795	0.118	0.618	0.661	0.413	0.503	0.159
CVPR2019	0.853	0.866	0.706	0.771	0.052	0.726	0.802	0.550	0.618	0.115
ICCV2019	0.848	0.870	0.702	0.702	0.050	0.732	0.768	0.583	0.670	0.104
MICCAI2020	0.860	0.898	0.763	0.789	0.044	0.769	0.833	0.663	0.710	0.094
CVPR2020	0.869	0.891	0.740	0.827	0.044	0.751	0.771	0.606	0.702	0.100
CVPR2021	0.893	0.923	0.813	0.824	0.030	0.775	0.847	0.673	0.725	0.088
CVPR2021	0.882	0.930	0.810	0.826	0.033	0.782	0.840	0.695	0.746	0.085
CVPR2021	0.890	0.935	0.822	0.841	0.030	0.787	0.838	0.696	0.744	0.080
ICCV2021	0.888	0.918	0.796	0.804	0.031	0.785	0.859	0.686	0.747	0.086
ırs	0.882	0. 937	0.829	0.838	0.029	0.808	0.877	0.750	0. 789	0.070
	年份 - CVPR2017 CVPR2018 DLMIA2018 CVPR2019 CVPR2019 ICCV2019 MICCAI2020 CVPR2021 CVPR2021 CVPR2021 CVPR2021 ICCV2021 ICCV2021	年份 S _a ↑ CVPR2017 0.794 CVPR2017 0.773 CVPR2018 0.769 DLMIA2018 0.695 CVPR2019 0.687 CVPR2019 0.853 ICCV2019 0.848 MICCAI2020 0.860 CVPR2021 0.889 CVPR2021 0.882 CVPR2021 0.880 ICCV2021 0.880 ICCV2021 0.885	年份 年份	年份 年份 S_a ↑ E_{ϕ} ↑ F_{β}^{w} ↑ CVPR2017 0.794 0.783 0.590 CVPR2017 0.773 0.758 0.555 CVPR2018 0.769 0.749 0.536 DLMIA2018 0.695 0.762 0.501 CVPR2019 0.687 0.721 0.474 CVPR2019 0.853 0.866 0.706 ICCV2019 0.848 0.870 0.702 MICCAI2020 0.860 0.898 0.763 CVPR2021 0.883 0.923 0.813 CVPR2021 0.882 0.930 0.810 CVPR2021 0.882 0.935 0.822 ICCV2021 0.882 0.918 0.796 mrs 0.882 0.937 0.829	年份 CHAMELEON(76 张图像) S _a ↑ E _Φ ↑ F ^a _β ↑ F _β ↑ CVPR2017 0.794 0.783 0.590 0.648 CVPR2017 0.773 0.758 0.555 0.630 CVPR2018 0.769 0.749 0.536 0.618 DLMIA2018 0.695 0.762 0.501 0.557 CVPR2019 0.687 0.721 0.474 0.795 CVPR2019 0.853 0.866 0.706 0.771 ICCV2019 0.848 0.870 0.702 0.702 MICCAI2020 0.860 0.898 0.763 0.789 CVPR2021 0.882 0.930 0.810 0.827 CVPR2021 0.888 0.935 0.822 0.841 ICCV2021 0.888 0.935 0.822 0.841 ICCV2021 0.888 0.937 0.829 0.838	年份 CHAMELEON(76 张图像) S _a ↑ E _φ ↑ F _β ↑ F _β ↑ M↓ CVPR2017 0.794 0.783 0.590 0.648 0.075 CVPR2017 0.773 0.758 0.555 0.630 0.085 CVPR2018 0.769 0.749 0.536 0.618 0.085 DLMIA2018 0.695 0.762 0.501 0.557 0.094 CVPR2019 0.687 0.721 0.474 0.795 0.118 CVPR2019 0.853 0.866 0.706 0.771 0.052 ICCV2019 0.848 0.870 0.702 0.702 0.50 MICCAI2020 0.860 0.898 0.763 0.789 0.044 CVPR2021 0.882 0.930 0.813 0.824 0.030 CVPR2021 0.882 0.930 0.810 0.826 0.033 CVPR2021 0.888 0.935 0.822 0.841 0.030 CVPR2021 0.888 <td< td=""><td>CHAMELEON(76 张图像)年份S_{α} ↑E_{ϕ} ↑F_{β}^{α} ↑R_{β} ↑$M \downarrow$$S_{\alpha}$ ↑CVPR20170.7940.7830.5900.6480.0750.684CVPR20170.7730.7580.5550.6300.0850.663CVPR20180.7690.7490.5360.6180.0850.609DLMIA20180.6950.7620.5010.5570.0940.599CVPR20190.6870.7210.4740.7950.1180.618CVPR20190.8530.8660.7060.7710.0520.726ICCV20190.8480.8700.7020.7020.0500.732MICCAI20200.8600.8980.7630.7890.0440.769CVPR20210.8820.9300.8130.8240.0300.775CVPR20210.8820.9300.8100.8260.0330.782Irs0.8820.9370.8290.8380.0290.808</td><td>年份 CHAMELEON(76 张图像) CAMO S_a ↑ E_{\phi} ↑ F^o_B ↑ F_b ↑ M ↓ S_a ↑ E_{\phi} ↑ CVPR2017 0.794 0.783 0.590 0.648 0.075 0.684 0.677 CVPR2017 0.773 0.758 0.555 0.630 0.085 0.663 0.659 CVPR2018 0.769 0.749 0.536 0.618 0.085 0.609 0.584 DLMIA2018 0.695 0.762 0.501 0.557 0.094 0.599 0.653 CVPR2019 0.687 0.721 0.474 0.795 0.118 0.618 0.661 CVPR2019 0.848 0.870 0.702 0.702 0.726 0.802 ICCV2019 0.848 0.870 0.702 0.702 0.732 0.768 MICCAI2020 0.860 0.898 0.763 0.789 0.044 0.769 0.833 CVPR2021 0.882 0.930 0.810 0.827 0.044 0.751 0.771 CVPR2021 0.889 0.930<td>年份 CHAMELEON(76 张图像) CAMO-Test(250 - 50 - 50 - 50 - 50 - 50 - 50 - 50 -</td><td>年份CHAMELEON(76 张图像)CAMO-Test(250 张图像)Sa ↑E ϕ ↑Fg ↑Fβ ↑M ↓Sa ↑E ϕ ↑Fg ↑Fβ ↑CVPR20170.7940.7830.5900.6480.0750.6840.6770.4830.676CVPR20170.7730.7580.5550.6300.0850.6630.6590.4550.520CVPR20180.7690.7490.5360.6180.0850.6090.5840.3560.573DLMIA20180.6950.7620.5010.5570.0940.5990.6530.3920.460CVPR20190.6870.7210.4740.7950.1180.6180.6610.4130.503CVPR20190.8840.8700.7020.7020.0500.7320.7680.5830.663ICCV20190.8480.8700.7020.7020.0500.7320.7680.5830.663MICCAI20200.8600.8980.7630.7890.0440.7690.8330.6630.710CVPR20210.8820.9300.8100.8260.0330.7870.8470.6730.725CVPR20210.8820.9300.8100.8260.0330.7870.8380.6960.744ICCV20210.8820.9350.8220.8410.0300.7870.8590.6860.747Irrs0.8820.9370.8290.8380.0290.8080.8770.7500.787</td></td></td<>	CHAMELEON(76 张图像)年份 S_{α} ↑ E_{ϕ} ↑ F_{β}^{α} ↑ R_{β} ↑ $M \downarrow$ S_{α} ↑CVPR20170.7940.7830.5900.6480.0750.684CVPR20170.7730.7580.5550.6300.0850.663CVPR20180.7690.7490.5360.6180.0850.609DLMIA20180.6950.7620.5010.5570.0940.599CVPR20190.6870.7210.4740.7950.1180.618CVPR20190.8530.8660.7060.7710.0520.726ICCV20190.8480.8700.7020.7020.0500.732MICCAI20200.8600.8980.7630.7890.0440.769CVPR20210.8820.9300.8130.8240.0300.775CVPR20210.8820.9300.8100.8260.0330.782Irs0.8820.9370.8290.8380.0290.808	年份 CHAMELEON(76 张图像) CAMO S _a ↑ E _{\phi} ↑ F ^o _B ↑ F _b ↑ M ↓ S _a ↑ E _{\phi} ↑ CVPR2017 0.794 0.783 0.590 0.648 0.075 0.684 0.677 CVPR2017 0.773 0.758 0.555 0.630 0.085 0.663 0.659 CVPR2018 0.769 0.749 0.536 0.618 0.085 0.609 0.584 DLMIA2018 0.695 0.762 0.501 0.557 0.094 0.599 0.653 CVPR2019 0.687 0.721 0.474 0.795 0.118 0.618 0.661 CVPR2019 0.848 0.870 0.702 0.702 0.726 0.802 ICCV2019 0.848 0.870 0.702 0.702 0.732 0.768 MICCAI2020 0.860 0.898 0.763 0.789 0.044 0.769 0.833 CVPR2021 0.882 0.930 0.810 0.827 0.044 0.751 0.771 CVPR2021 0.889 0.930 <td>年份 CHAMELEON(76 张图像) CAMO-Test(250 - 50 - 50 - 50 - 50 - 50 - 50 - 50 -</td> <td>年份CHAMELEON(76 张图像)CAMO-Test(250 张图像)Sa ↑E ϕ ↑Fg ↑Fβ ↑M ↓Sa ↑E ϕ ↑Fg ↑Fβ ↑CVPR20170.7940.7830.5900.6480.0750.6840.6770.4830.676CVPR20170.7730.7580.5550.6300.0850.6630.6590.4550.520CVPR20180.7690.7490.5360.6180.0850.6090.5840.3560.573DLMIA20180.6950.7620.5010.5570.0940.5990.6530.3920.460CVPR20190.6870.7210.4740.7950.1180.6180.6610.4130.503CVPR20190.8840.8700.7020.7020.0500.7320.7680.5830.663ICCV20190.8480.8700.7020.7020.0500.7320.7680.5830.663MICCAI20200.8600.8980.7630.7890.0440.7690.8330.6630.710CVPR20210.8820.9300.8100.8260.0330.7870.8470.6730.725CVPR20210.8820.9300.8100.8260.0330.7870.8380.6960.744ICCV20210.8820.9350.8220.8410.0300.7870.8590.6860.747Irrs0.8820.9370.8290.8380.0290.8080.8770.7500.787</td>	年份 CHAMELEON(76 张图像) CAMO-Test(250 - 50 - 50 - 50 - 50 - 50 - 50 - 50 -	年份CHAMELEON(76 张图像)CAMO-Test(250 张图像)Sa ↑E ϕ ↑Fg ↑Fβ ↑M ↓Sa ↑E ϕ ↑Fg ↑Fβ ↑CVPR20170.7940.7830.5900.6480.0750.6840.6770.4830.676CVPR20170.7730.7580.5550.6300.0850.6630.6590.4550.520CVPR20180.7690.7490.5360.6180.0850.6090.5840.3560.573DLMIA20180.6950.7620.5010.5570.0940.5990.6530.3920.460CVPR20190.6870.7210.4740.7950.1180.6180.6610.4130.503CVPR20190.8840.8700.7020.7020.0500.7320.7680.5830.663ICCV20190.8480.8700.7020.7020.0500.7320.7680.5830.663MICCAI20200.8600.8980.7630.7890.0440.7690.8330.6630.710CVPR20210.8820.9300.8100.8260.0330.7870.8470.6730.725CVPR20210.8820.9300.8100.8260.0330.7870.8380.6960.744ICCV20210.8820.9350.8220.8410.0300.7870.8590.6860.747Irrs0.8820.9370.8290.8380.0290.8080.8770.7500.787

表 2 不同方法在 COD10K 和 COD10K-Test 测试数据集的指标结果比较

之 外	左 //\		COD10K	-Test(2026	张图像)			NC4	K(4121 张	图像)	
万伝	平田 -	S _α ↑	E_{Φ} \uparrow	F^{ω}_{β} A	F_{β} \uparrow	$M \not \bullet$	S_{α} \uparrow	E_{Φ} \uparrow	F^{ω}_{β} \blacklozenge	F_{β} \uparrow	$M \not \downarrow$
FPN	CVPR2017	0.697	0.691	0.411	0.481	0.075	_	_	_	_	_
PSPNet	CVPR2017	0.678	0.680	0.377	0.457	0.080	_				
PiCANet	CVPR2018	0.649	0.643	0.322	0.489	0.090					
UNet++	DLMIA2018	0.623	0.672	0.350	0.408	0.086	_				
BASNet	CVPR2019	0.634	0.678	0.365	0.486	0.105	0.695	0.761	0.546	0.610	0.095
CPD	CVPR2019	0.747	0.770	0.508	0.595	0.059	0.716	0.724	0.551	0.597	0.092
EGNet	ICCV2019	0.737	0.779	0.509	0.582	0.056	0.766	0.792	0.626	0.689	0.067
PraNet	MICCAI2020	0.789	0.839	0.629	0.671	0.045	0.822	0.876	0.724	0.762	0.059
SINet	CVPR2020	0.771	0.806	0.551	0.679	0.051	0.808	0.873	0.722	0.769	0.057
MGL	CVPR2021	0.814	0.865	0.666	0.710	0.035	Τ_	<u></u>			
PFNet	CVPR2021	0.800	0.868	0.660	0.701	0.040	0.829	0.887	0.745	0.784	0.053
LSR	CVPR2021	0.804	0.880	0.672	0.714	0.036	0.840	0.894	0.765	0.804	0.048
UGTR	ICCV2021	0.818	0.850	0.667	0.711	0.035					
Οι	ırs	0.820	0.892	0.710	0.740	0.031	0.845	0.903	0.786	0.816	0.044

本文方法在场景复杂的大规模挑战性测试数据 集 COD10K-Test 上各项指标全面大幅度领先最先 进的方法.表 3 和表 4 列出了不同算法在 COD10K-Test 的四个常见父类(即:两栖动物类、水生动物 类、飞行动物类、陆生动物类)上的定量指标比较结 果.可以看出,本文方法在四个父类上的指标结果均 全面大幅度领先最先进的方法, F_{β}^{ω} 分别提升了 1.9%、3.1%、4.1%、4.5%.进一步充分验证了本文 方法在 COD10K-Test 上的有效性和先进性.

图 8 和图 9 是本文方法与不同方法的 PR 曲 线、F_β曲线示意图.其中,红色连续曲线是本文方法 的结果.从上下两幅图中的八个子图可以进一步直 观地看出,本文方法在指标性能上的表现最优. 定性比较:图 10 是本文方法和其它方法的定性 比较结果.可以看出,本文方法能够准确识别出目前 最先进的几种方法无法识别出的颜色高度相似物 体,如第一行海水中的乌龟和第八行的白色花朵上 的白蜘蛛.与此同时,本文方法可以分辨出场景中存 在语义歧义的伪装物体,如第四行近处的狮子和远 处的斑马.此外,本文方法还能够解决预测结果细节 丢失的问题,如第三行、第六行和第九行中的物体, 目前最先进的几种方法都丢失了部分细节,尤其是 最为细长的部分,而本文方法还能识别出场景中的小 目标伪装物体,如第七行图中树枝上的鸟,在所有方 法中只有本文方法能够准确识别出来.

表 3 不同方法在 COD10K-Test 测试集的 4 个父类上的指标结果比较

大 社	年八		两栖动物类(124 张图像)			水生动物类(474 张图像)	
刀伝	平顶 —	$S_{\alpha} \uparrow$	E_{Φ} \uparrow	F_{β} \uparrow	$M \downarrow$	S_{α} \bigstar	E_{Φ} \uparrow	F_{β} \uparrow	$M \not \bullet$
FPN	CVPR2017	0.745	0.776	0.497	0.065	0.684	0.732	0.432	0.103
PSPNet	CVPR2017	0.736	0.774	0.463	0.072	0.659	0.712	0.396	0.111
PiCANet	CVPR2018	0.686	0.702	0.405	0.079	0.616	0.631	0.335	0.115
UNet++	DLMIA2018	0.677	0.745	0.434	0.079	0.599	0.673	0.347	0.121
BASNet	CVPR2019	0.707	0.740	0.476	0.087	0.619	0.666	0.373	0.134
CPD	CVPR2019	0.794	0.839	0.587	0.051	0.739	0.792	0.529	0.082
EGNet	ICCV2019	0.785	0.854	0.606	0.047	0.725	0.793	0.528	0.080
PraNet	MICCAI2020	0.842	0.905	0.717	0.035	0.781	0.883	0.696	0.065
SINet	CVPR2020	0.827	0.866	0.654	0.042	0.758	0.803	0.570	0.073
MGL	CVPR2021	0.854	0.885	0.733	0.028	0.807	0.853	0.686	0.050
PFNet	CVPR2021	0.847	0.911	0.740	0.031	0.792	0.868	0.675	0.055
LSR	CVPR2021	0.845	0.905	0.751	0.030	0.802	0.875	0.693	0.052
UGTR	ICCV2021	0.856	0.896	0.737	0.029	0.809	0.859	0.685	0.050
Ou	urs	0.860	0.922	0.770	0.025	0.815	0.894	0.724	0.045

表 4 不同方法在 COD10K-Test 测试集的 4 个父类上的指标结果比较

		F	飞行动物类((714 张图像)			陆生动物类(699 张图像)	
万伝	+107 -	S_{α}	E_{Φ} \blacklozenge	F_{eta} \bigstar	$M \not \bullet$	$S_{\alpha} \uparrow$	E_{Φ} \uparrow	F_{β} \uparrow	$M \not \bullet$
FPN	CVPR2017	0.726	0.766	0.440	0.061	0.601	0.656	0.353	0.109
PSPNet	CVPR2017	0.644	6 0.767	0.449	0.063	0.669	0.718	0.352	0.071
PiCANet	CVPR2018	0.663	0.676	0.347	0.069	0.658	0.708	0.273	0.074
UNet++	DLMIA2018	0.659	0.727	0.397	0.068	0.608	0.749	0.288	0.070
BASNet	CVPR2019	0.664	0.710	0.403	0.085	0.601	0.645	0.301	0.108
CPD	CVPR2019	0.777	0.827	0.544	0.046	0.714	0.771	0.445	0.058
EGNet	ICCV2019	0.766	0.826	0.543	0.044	0.700	0.775	0445	0.053
PraNet	MICCAI2020	0.819	0.888	0.669	0.033	0.756	0.835	0.565	0.046
SINet	CVPR2020	0.798	0.828	0.580	0.040	0.743	0.778	0.491	0.050
MGL	CVPR2021	0.839	0.872	0.699	0.025	0.785	0.822	0.604	0.035
PFNet	CVPR2021	0.823	0.902	0.691	0.029	0.772	0.854	0.605	0.040
LSR	CVPR2021	0.830	0.906	0.706	0.026	0.772	0.854	0.610	0.038
UGTR	ICCV2021	0.842	0.872	0.698	0.026	0.789	0.822	0.605	0.036
O.	urs	0.846	0.913	0.747	0.022	0. 791	0.867	0.655	0.033



图 8 不同方法在四个测试数据集上的 PR 曲线示意图





以上定量与定性实验分析,充分展现了本文方 法的鲁棒性和有效性.这主要得益于 UCAM 能够 从局部到全局、从小尺度到大尺度逐步搜索捕获场 景上下文信息,弥补主干网络丢失的局部细节信息; 此外,CFAM 能够高效渐进地聚合 UCAM 捕获到 的多层相邻级特征之间的互补信息,使得预测图的 局部细节更完善、边界更精细. 模型参数比较:表5是本文方法与最新几种方 法的模型参数比较.可以看出,本文方法的模型参数 量是其中最少的,FLOPs 比最低的 SINet 略微高出 一点.因为,CFAM在聚合特征过程时只使用了简 单的池化、卷积、元素加与乘的操作,并不涉及复杂 运算;而且,UCAM 进行多次下采样后特征图不断 缩小,即使使用多次PSCB编码操作也并不会显著

方法	输入图像大小	参数量/MB	FLOPs/GB
$SINet_{20}$	352 imes 352	48.95	19.4
MGL_{21}	473×473	73.73	378.2
$PFNet_{21}$	416×416	46.17	26.5
LSR_{21}	480×480	55.45	55.1
$UGTR_{21}$	473×473	37.54	123.3
Ours	416×416	26.82	21.4

表 5 模型参数比较

增加参数量.因此,堆叠多个 UCAM 与 CFAM 来搭 建的模型产生的最终计算量也很小.

4.5 消融实验

为了充分证明本文方法的有效性,本文进一步在 三个最具挑战的测试数据集 CHAMELEON、CAMO-Test、COD10K-Test 上进行了广泛的消融实验: ASPP 的有效性:ASPP 使用在主干网络的最后一层,不仅通过全局池化增强了全局语义,还由不同大小空洞率的卷积操作捕获了多尺度信息.从表 6中的消融实验(1)和(2)的对比结果可以看出,使用ASPP 后模型可以更好地捕获到伪装物体全局位置信息,从而使得各项指标均有显著提升.

此外,为了验证 ASPP 中的卷积设置不同空洞 率对实验结果产生的影响,本文通过设置不同空洞 率组合,在场景复杂、最具挑战性的测试数据集 COD10K-Test上进行了广泛实验,表 7 中的实验结 果表明,并不是设置越大的空洞率的结果就越好,经 过多次实验探索和验证,最终发现实验结果最优的 空洞率组合应分别设置为 3,5,7.

表 6 本文方法在 CHAMELEON、CAMO-Test、COD10k-Test 三个测试数据集上的消融实验

	ACDD	UCAM	CEAM	6	СНАМ	ELEON			CAM	D-Test			COD10	K-Test	
	ASEE	UCAM	ULAM	S_{α}	E_{Φ} \uparrow	F^{ω}_{β} \blacklozenge	$M \not \bullet$	S_{α} \blacklozenge	E_{Φ} \blacklozenge	F^{ω}_{β} \blacklozenge	$M \not \downarrow$	S_{α} \bigstar	E_{Φ} \blacklozenge	F^{ω}_{β} \blacklozenge	$M \not \bullet$
(1)				0.842	0.916	0.768	0.038	0.765	0.832	0.668	0.088	0.766	0.859	0.618	0.044
(2)	\checkmark			0.850	0.921	0.771	0.034	0.783	0.857	0.701	0.081	0.779	0.873	0.634	0.039
(3)	\checkmark	\checkmark		0.880	0.934	0.822	0.026	0.805	0.863	0.732	0.072	0.812	0.885	0.692	0.033
(4)		\checkmark	\checkmark	0.881	0.932	0.827	0.030	0.800	0.870	0.741	0.074	0.818	0.888	0.706	0.032
(5)	\checkmark		\checkmark	0.862	0.929	0.796	0.032	0.802	0.865	0.737	0.074	0.810	0.882	0.691	0.034
(6)	\checkmark	\checkmark	\checkmark	0.882	0.937	0. 829	0.029	0.808	0.877	0.750	0.070	0.820	0.892	0.710	0.031

表 7 ASPP 中设置不同空洞率的对比实验

空洞率设置	S_{α} \uparrow	E_{Φ} \blacklozenge	F_{β}	$M \not \flat$
6,12,18	0.805	0.879	0.681	0.036
5,8,11	0.818	0.890	0.706	0.032
2,4,6	0.817	0.888	0.705	0.032
3,5,7	0.820	0.892	0.710	0.031

UCAM的有效性:UCAM用于感知主干网络的 多级特征,旨在捕获充足的场景上下文信息.从表 6 中的消融实验(2)和(3)可以观察到,在ASPP基础 上使用UCAM后指标大幅度提升.此外,从消融可 视化图11中可以看出,网络模型不使用UCAM时, 预测图的全局和局部信息均有一定程度的缺失.同 时,第二行和第三行的预测图还存在一定的语义歧 义,将与伪装物体相似的背景区域也识别为伪装物 体.然而,当模型使用了UCAM后,UCAM能够结合 PSCB从局部到全局、从小尺度到大尺度逐步搜索并 捕获场景上下文信息,为模型补充了丢失的局部细 节信息,使得模型能够结合全局和局部上下文信息 综合进行场景分析和推理,最终预测图的局部结构 会更加完整、精细,一定程度上避免了上述情况.

为了验证 PCSB 在 UCAM 中对特征编码的有效性,本文也在测试数据集 COD10K-Test 上进行 了消融实验:使用一个 3×3 的卷积操作来替换 PCSB. 从表 8 可以看出,使用 PCSB 的结果更优.



图 11 模型使用 UCAM 和 CFAM 的消融对比图

PCSB通过并行级联方式,逐级递进的扩大感受野 区域,为 UCAM 能够进一步捕获到局部到全局的 信息做出了一定的贡献.

表 8 UCAM 中使用不同特征编码方式的对比实验

编码方式	S_{α} \blacklozenge	E_{Φ} \blacklozenge	F_{β} \bigstar	$M \not \downarrow$
3×3Conv	0.818	0.888	0.702	0.033
PCSB	0.820	0.892	0.710	0.031

UCAM 是一种 U 型结构,经过 PCSB 编码后, 通过渐进式上采样和残差连接来解码聚合多尺度信息.为了验证使用相加与拼接操作聚合特征的有效 性,本文也在测试数据 COD10k-Test 上进行了消融 实验. 从表 9 可以看出,使用相加的方式结果更优.

CFAM 与残差渐进特征聚合机制的有效性: CFAM 高效聚合相邻层级的互补特征.其中,采用渐

	表 9	UCAM 👎	•使用不同	同特征聚合	方式的对	比实验
_						

聚合方式	S_{α} \bigstar	E_{Φ} \blacklozenge	F_{β}	$M \not \downarrow$
拼接	0.819	0.890	0.706	0.032
相加	0.820	0.892	0.710	0.031

进式特征聚合可以充分捕获相邻两级特征的优势部分,而使用残差连接可以避免特征聚合过程的信息 损失.从表6中的消融实验(3)和(4)以及(2)和(5) 可以看出使用 CFAM 后各项指标再次有了显著提 升.这得益于 CFAM 从水平和垂直两个空间维度上 对特征进行增强,使得模型能够从两个方向关注到 物体的全局和局部信息.从消融可视化图 11 中可以 看出,模型不使用 CFAM 时,预测结果图的局部细 节非常不完整:第一行中右边黄色的毛毛虫预测结 果缺失较多,而第二行预测图蚂蚱的腿部细节也不 够精细.

CFAM 从水平和垂直两个方向对特征进行强 化.为了验证两个方向上的强化对最终结果的影响, 本文在测试数据集 COD10K-Test 上进行了消融实 验,分别是以直接简单相加的方式聚合跨级特征、只 使用水平方向、只使用垂直方向以及两个方向同时 使用来聚合跨级特征.从表 10 可以看出,同时在两 个方向上都对特征进行强化能够获得最优的结果. 从聚合方式可视化对比图 12 中可以观察到,直接聚 合相邻层的特征会给特征带来一定的噪声污染,而 只使用水平方向或垂直方向来增强聚合特征的表现 不如同时使用两个方向得到的特征响应,这是因为, 从两个方向同时进行特征强化可以充分考虑两个方 向上的特征关系,从而获得更好的特征聚合效果.

方向	S_{α} \blacklozenge	E_{Φ} \blacklozenge	Fÿ ↑	$M \not \bullet$
直接聚合	0.812	0.885	0.692	0.033
水平方向	0.816	0.887	0.705	0.032
垂直方向	0.817	0.888	0.705	0.032
水平十垂直	0.820	0.892	0.710	0.031

表	10	CFAM	中使	用不	同强	化方	向	的对	比	实驯
---	----	------	----	----	----	----	---	----	---	----



直接聚合 水平方向聚合 垂直方向聚合 水平+垂直方向聚合 图 12 使用不同特征聚合方式的可视化图

残差渐进特征聚合机制先使用 CFAM 将相邻 两层互补特征聚合,然后通过残差连接上较高级别的特征,旨在减少高级语义特征在聚合过程的丢失.为了验证残差连接的必要性,本文在测试数据集

COD10K-Test上进行了消融实验,分别对比使用和 不使用残差连接对最终结果的影响.从表 11 可以看 出,使用残差连接的结果更优.

表 11 是否使用残差连接的对比实验

有无残差连接	S_{α} \uparrow	E_{Φ} \blacklozenge	F β	$M \not \bullet$
无残差连接	0.818	0.890	0.706	0.032
有残差连接	0.820	0.892	0.710	0.031

特征可视化图 13 展示了渐进特征聚合使用与 不使用残差连接对特征聚合的影响.从中可以直观 观察到使用残差连接后的特征响应更强,从而有效 地恢复了因特征聚合损失的部分信息.这充分说明 了残差连接在渐进特征聚合中发挥了重要作用.



结论

5

本文提出了一种渐进聚合场景上下文特征的 COD网络,主要包括两个创新设计:U型上下文感 知模块(UCAM)和跨级特征聚合模块(CFAM).其 中,UCAM 能够捕获丰富的多尺度场景上下文信 息,从而挖掘到更多的细节信息;CFAM 则将坐标 方向的注意力与多层级残差渐进特征聚合机制相结 合,逐级渐进聚合相邻层之间的互补特征,生成精确 的预测结果图.本文在四个极具挑战性的测试数据 集上进行了广泛评测,并对比了最新的 13 种算法. 实验结果充分证明了本文方法的有效性和优越性.

参考文献

- [1] Fan D P, Ji G P, Sun G, et al. Camouflaged object detection//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA, 2020: 2777-2787
- [2] Fan D P, Ji G P, Zhou T, et al. Pranet: Parallel reverse attention network for polyp segmentation//Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham, Swiss: Springer, 2020; 263-273

- [3] Fan D P, Zhou T, Ji G P, et al. Inf-net: Automatic covid-19 lung infection segmentation from ct images. IEEE Transactions on Medical Imaging, 2020, 39(8): 2626-2637
- [4] Qiu L, Wu X, Yu Z. A high-efficiency fully convolutional networks for pixel-wise surface defect detection. IEEE Access, 2019, 7: 15884-15893
- [5] Li T, Song H, Zhang K, et al. Recurrent reverse attention guided residual learning for saliency object detection. Neurocomputing, 2020, 389: 170-178
- [6] Liu N, Han J, Yang M H. Picanet: Learning pixel-wise contextual attention for saliency detection//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. SaltLakeCity, USA, 2018: 3089-3098
- [7] Qin X, Zhang Z, Huang C, et al. Basnet: Boundary-aware salient object detection//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Angeles, USA, 2019: 7479-7489
- [8] Wu Z, Su L, Huang Q. Cascaded partial decoder for fast and accurate salient object detection//Proceedings of the IEEE/ CVF Conference on Computer Vision and Pattern Recognition. Los Angeles, USA, 2019: 3907-3916
- [9] Zhao J X, Liu J J, Fan D P, et al. EGNet: Edge guidance network for salient object detection//Proceedings of the IEEE/CVF International Conference on Computer Vision. LosAngeles, USA, 2019: 8779-8788
- [10] Wang Zheng-Wen, Song Hui-Hui, Fan Jia-Qing, Jin Qing-Shan. Semantic guided feature aggregation network for salient object detection. Acta Automatica Sinica, to appear(in Chinese)

(王正文,宋慧慧,樊佳庆等,基于语义引导特征聚合的显著 性目标检测网络,自动化学报,已录用)

- [11] Chen Bing-Cai, Tao Xin, Chen Hui, et al. Saliency detection via fusion of boundary connectivity and local. Chinese Journal of Computers, 2020, 43(1): 16-28(in Chinese)
 (陈炳才,陶鑫,陈慧等.融合边界连通性与局部对比性的图 像显著性检测.计算机学报,2020,43(1): 16-28)
- [12] Zhang Dong-Ming, Jin Guo-Qing, Dai Feng, et al. Salient object detection based on deep fusion of hand-crafted fearures. Chinese Journal of Computers, 2019, 42(9): 2076-2086(in Chinese)

(张冬明, 靳国庆, 代锋等. 基于深度融合的显著性目标检测 算法. 计算机学报, 2019, 42(9): 2076-2086)

- [13] Bhajantri N U, Nagabhushan P. Camouflage defect identification: A novel approach//Proceedings of the 9th International Conference on Information Technology (ICIT'06). Bhubaneswar, India, 2006: 145-148
- [14] 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, 2016, 27(9): 2001-2013
- [15] Pan Y, Chen Y, Fu Q, et al. Study on the camouflaged target detection method based on 3D convexity. Modern Applied Science, 2011, 5(4): 152
- [16] Yang F, Zhai Q, Li X, et al. Uncertainty-guided transformer reasoning for camouflaged object detection//Proceedings of the IEEE/CVF International Conference on Computer Vision. Virtual, 2021: 4146-4155

- [17] Lv Y, Zhang J, Dai Y, et al. Simultaneously localize, segment and rank the camouflaged objects//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Virtual, 2021: 11591-11601
- [18] Mei H, Ji G P, Wei Z, et al. Camouflaged object segmentation with distraction mining//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Virtual, 2021: 8772-8781
- [19] Zhai Q, Li X, Yang F, et al. Mutual graph learning for camouflaged object detection//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Virtual, 2021: 12997-13007
- [20] Chen L C, Papandreou G, Schroff F, et al. Rethinking atrous convolution for semantic image segmentation. arXiv preprint arXiv:1706.05587, 2017
- [21] Zhao H, Shi J, Qi X, et al. Pyramid scene parsing network//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Venice, Italy, 2017: 2881-2890
- [22] Liu S, Huang D. Receptive field block net for accurate and fast object detection//Proceedings of the European Conference on Computer Vision (ECCV). Munich, Germany, 2018: 385-400
- [23] Fu J, Liu J, Tian H, et al. Dual attention network for scene segmentation//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Angeles, USA, 2019: 3146-3154
- Chen Z, Xu Q, Cong R, et al. Global context-aware progressive aggregation network for salient object detection//
 Proceedings of the AAAI Conference on Artificial Intelligence. New York, USA, 2020, 34(07): 10599-10606
- [25] Lin T Y. Dollár P. Girshick R. et al. Feature pyramid networks for object detection//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Hawaii, USA, 2017: 2117-2125
- [26] Wei J, Wang S, Huang Q. F³ Net: Fusion, feedback and focus for salient object detection//Proceedings of the AAAI Conference on Artificial Intelligence. NewYork, USA, 2020, 34(7): 12321-12328
- [27] Ma M, Xia C, Li J. Pyramidal feature shrinking for salient object detection//Proceedings of the AAAI Conference on Artificial Intelligence. Virtual, 2021, 35(3): 2311-2318
- [28] Gao S H, Cheng M M, Zhao K, et al. Res2net: A new multi-scale backbone architecture. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 43 (2): 652-662
- [29] Qin X, Zhang Z, Huang C, et al. U2-Net: Going deeper with nested U-structure for salient object detection. Pattern Recognition, 2020, 106: 107404
- [30] Hou Q, Zhou D, Feng J. Coordinate attention for efficient mobile network design//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Virtual, 2021: 13713-13722
- [31] Ramachandran P, Zoph B, Le Q V. Searching for activation functions. arXiv preprint arXiv:1710.05941, 2017

- [32] Skurowski P, Abdulameer H, Błaszczyk J, et al. Animal camouflage analysis: Chameleon database. Unpublished Manuscript, 2018, 2(6): 7
- [33] Le T N, Nguyen T V, Nie Z, et al. Anabranch network for camouflaged object segmentation. Computer Vision and Image Understanding, 2019, 184: 45-56
- [34] 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. Rhode, Island, 2012: 733-740
- [35] Fan D P, Gong C, Cao Y, et al. Enhanced-alignment measure for binary foreground map evaluation. arXiv preprint arXiv: 1805.10421, 2018
- [36] Fan D P, Cheng M M, Liu Y, et al. Structure-measure: A new way to evaluate foreground maps//Proceedings of the IEEE International Conference on Computer Vision. Venice, Italy, 2017: 4548-4557
- [37] Achanta R, Hemami S, Estrada F, et al. Frequency-tuned



LIU Yan, M. S. candidate. His research interests include camouflaged object detection.

ZHANG Kai-Hua, Ph. D., professor. His research

Background

Camouflaged object detection(COD) is a meaningful and challenging task in computer vision. It has characteristics and commonalities with the traditional salient object detection task. The two tasks focus on the detection and recognition of hidden objects and salient objects in images respectively. Research on camouflaged objects can be traced back to the 1990s. In the past few decades, a lot of work has been generated around COD algorithm research. Early research mainly used hand-designed features to distinguish camouflaged objects. However, the ability of such features to distinguish camouflage from non-camouflage is very limited, and it is only suitable for relatively simple scenes. In recent years, with the rise of deep learning, deep learning-based COD methods have achieved breakthroughs in performance.

Fan et al. proposes a large-scale COD dataset called COD10K, and designed a simple and effective COD network SINet, which further promotes the development of deep learning on COD tasks. Subsequently, a series of algorithms with advanced performance emerged, such as PFNet, MGL, LSR and UGTR, which designed COD networks through distraction mining, graph-based mutual learning, simultaneously localize, segment and rank the camouflaged Objects and combining probability representation models and Transformer. Those models have achieved a state of the art performance salient region detection//Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami, USA, 2009: 1597-1604

- [38] Margolin R, Zelnik-Manor L, Tal A. How to evaluate foreground maps?//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Columbus, USA, 2014: 248-255
- [39] Paszke A, Gross S, Massa F, et al. Pytorch: An imperative style, high-performance deep learning library. Advances in Neural Information Processing Systems. Vancouver Canada, 2019, 32
- [40] Kingma D P, Ba J. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980, 2014
- [41] Zhou Z, Rahman Siddiquee M M, Tajbakhsh N, et al. Unet ++: A nested u-net architecture for medical image segmentation//Proceedings of the Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Cham, Swiss: Springer, 2018; 3-11

interests include Visual tracking and co-saliency detection.

FAN Jia-Qing, Ph. D. candidate. His research interests include video object segmentation.

ZHAO Ya-Qian, senior engineer. Her research interests include artificial intelligence.

LIU Qing-Shan, Ph. D., professor. His research interests include computer vision and machine learning.

and refined detection results.

In this paper, we propose a COD network based on scene context-aware and progressive feature aggregation. Specifically, the network mainly has two innovative designs: U-shape Context-Aware Module (UCAM) and Cross-level Feature Aggregation Module (CFAM). UCAM is used to fully mine the scene context information of camouflaged objects from local to global and from large scale to small scale, while CFAM incorporates a residual progressive mechanism to aggregate complementary information between high-level semantics and low-level details at adjacent levels. Thereby, making up for the ambiguity caused by the loss of local details and the lack of global semantics, and obtaining fine prediction results.

We have evaluated four extremely challenging benchmark datasets including CHAMELEON, CAMO-Test, COD10K-Test, and NC4K. The evaluation results show that the weight F-measure has increased by 0.7%, 4.2%, 2.6%, and 1.2% on the four test datasets, respectively, and the MAE has decreased by 1%, 10%, 4%, and 4%, respectively. The evaluation results show that the algorithm in this paper has demonstrated that our model has achieved leading performance compared with state-of-the-art methods.