

分心感知的伪装物体分割

梅海洋¹, 杨鑫^{1*}, 周运铎¹, 季葛鹏², 魏小鹏¹, 范登平³

1. 大连理工大学, 大连 116024, 中国

2. 澳大利亚国立大学, 堪培拉 2601, 澳大利亚

3. 南开大学, 天津 300350, 中国

* 通信作者. E-mail: xinyang@dlut.edu.cn

国家自然科学基金 (批准号: 61972067、U21A20491、U1908214)、中国国家重点研发计划 (2021ZD0112400)、大连市创新科技基金 (批准号: 2020JJ26GX036) 资助

摘要 本文致力于设计一个有效且高效的伪装物体分割 (Camouflaged Object Segmentation, COS) 模型. 为此, 本文开发了一个生物启发的框架, 称为金字塔定位和聚焦网络 (Pyramid Positioning and Focus Network, PFNet+), 其模仿了自然界中的捕食过程. 具体地, 本文的 PFNet+ 包含三个关键模块, 即: 上下文增强模块 (Context Enrichment, CEn)、金字塔定位模块 (Pyramid Positioning Module, PPM) 和聚焦模块 (Focus Module, FM). CEn 通过整合上下文信息来增强骨干特征的表征能力, 从而提供更有辨别性的骨干特征. PPM 被设计用来模仿捕食中的检测过程, 以金字塔的方式从全局的角度定位潜在的目标物体. 然后 FM 被用来执行捕食中的识别过程, 通过在歧义区域的聚焦来逐步细化初始的预测结果. 值得注意的是, 在 FM 中, 本文开发了一个新颖的分心挖掘策略以用于分心区域的发现和去除, 以提高预测的性能. 大量的实验证明本文的 PFNet+ 能够实时运行 (56fps), 在四个标准度量指标下, PFNet+ 在三个具有挑战性的数据集上都显著优于现有的 20 个最新模型, 在其他视觉任务 (如息肉分割) 上的实验进一步证明了 PFNet+ 的泛化能力.

关键词 伪装物体, 分心, 上下文增强, 上下文探索, 金字塔, 分割

1 引言

伪装是动物或物体通过材料、颜色或者光照的任意组合实现的隐藏, 以使目标物体难以被看见 (隐身) 或伪装成其他物体 (模仿) [1]. 受益于发现 “无缝” 嵌入其周围环境的伪装物体的能力, 伪装物体分割 (COS) 在医学诊断 (如息肉分割 [2~4] 和肺部感染分割 [5])、工业 (如在自动生产线上检查不合格产品)、农业 (如蝗虫检测, 以防止入侵)、安全和监视 (如搜索和救援任务, 以及恶

引用格式: 梅海洋, 杨鑫, 周运铎, 等. 分心感知的伪装物体分割. 中国科学: 信息科学, 在审文章

Mei H Y, Yang X, Zhou Y D, et al. Distraction-Aware Camouflaged Object Segmentation (in Chinese). Sci Sin Inform, for review

劣天气中针对自动驾驶的行人或障碍物的检测)、科学研究(如稀有物种发现)和艺术(如逼真的融合和娱乐艺术)等不同领域具有广泛的应用价值。

然而, COS 是一项极具挑战性的任务, 因为伪装策略是通过欺骗观察者的视觉感知系统来工作的 [1], 因此需要大量的视觉感知知识 [6] 来消除由于目标物体和背景之间高度的内在相似性所引发的歧义性。伪装物体分割的研究在生物学和艺术等领域有着悠久而丰富的历史 [1]。早期方法致力于根据纹理 [7]、三维凸度 [8] 和运动 [9] 等手工设计的低级特征来区分前景和背景。然而, 这类特征对伪装和非伪装物体的区分能力有限, 因而在复杂场景中往往失效。尽管最近提出的基于深度学习的方法 [10~12] 在一定程度上提高了分割性能, 但在探索准确而有效的伪装物体分割方法方面仍有很大的空间。

在自然界中, 被捕食动物利用伪装等机制误导捕食者的视觉感官, 以降低被发现的风险 [1]。在自然选择的压力下, 捕食动物为成功的捕食进化出了敏锐的感官和聪明的大脑等多种适应能力。捕食过程可分为三个阶段, 即检测、识别和捕获 [13]。这激发了本文的仿生解决方案, 即通过模仿捕食的前两个阶段来分割伪装物体。

本文提出了金字塔定位和聚焦网络 (PFNet+), 极大提高了伪装物体分割的性能。本文的 PFNet+ 包含三个关键模块, 即上下文增强模块 (CEn)、金字塔定位模块 (PPM) 和聚焦模块 (FM)。其中, CEn 通过整合上下文信息来增强骨干特征的表示能力, 从而提供更有辨别性的骨干特征。PPM 被设计用来模仿捕食中的检测过程, 以金字塔的方式从全局的角度定位潜在的目标物体。然后 FM 被用来执行捕食中的识别过程, 通过在歧义区域的聚焦来逐步细化粗糙的预测结果。具体而言, PPM 利用四个分支中的四个定位模块 (PMs) 来推断目标物体的初始位置, 以在不同尺度下感知区域之间的相关性, 从而减少伪装物体的形状变化和不同位置分布所带来的影响。PM 由一个通道注意块和一个空间注意块组成, 它们都以非局部的方式实现, 以捕获通道和空间位置方面的长程语义依赖, 从而从全局角度推断伪装物体的初始位置。FM 首先基于前景注意(或者背景注意)特征进行多尺度的上下文探索, 发现假阳性(或者假阴性)干扰区域, 然后去除这些干扰, 得到目标物体更纯净的特征表示。这种分心挖掘策略以隐式方式实现, 作用于不同层次的特征上, 以逐步细化分割结果, 使得 PFNet+ 模型具有很强的精确分割伪装物体的能力(如图1所示)。综上所述, 本文的贡献如下:

- 本文将分心的概念引入伪装物体分割任务, 并开发了一种新的分心挖掘策略来进行分心区域的发现和去除, 以帮助伪装物体的精确分割。
- 本文提出了一个新颖的伪装物体分割方法。该方法首先通过探索长程语义依赖关系来定位潜在的目标物体, 然后聚焦于分心区域的发现和去除以逐步细化分割结果。
- 所提出的方法在三个基准数据集上实现了优异的伪装物体分割性能, 实验结果证明了本文方法的有效性。

本文的初步版本在 CVPR 2021 会议上发布 [14], 并在以下四个方面做了扩展:

- 本文引入了一个上下文增强模块来增强骨干特征的表示能力, 以提供更有辨别性的骨干特征, 从而提升伪装物体的定位和聚焦效果。
- 本文将原来的定位模块 (PM) 扩展为金字塔定位模块 (PPM), 以金字塔的方式实现对伪装物体更鲁棒的定位。
- 本文表明新的模型(称为 PFNet+) 比原来的 PFNet [14] 具有更优异的分割性能。并且本文提

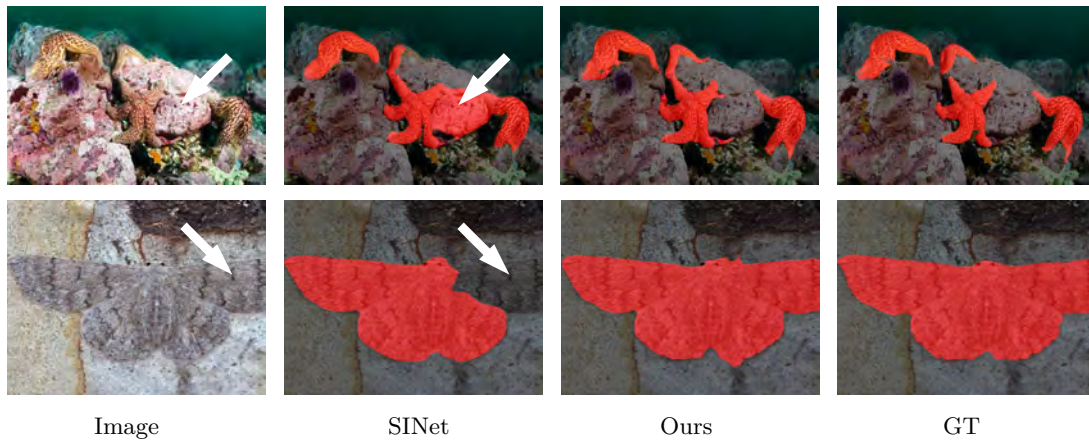


图 1 伪装物体分割的视觉样例. 现有的先进方法 SINet [10] 被与伪装物体具有相似外观的背景区域 (第一行中的箭头所指区域) 或混杂在背景中的伪装物体区域 (第二行中的箭头所指区域) 所迷惑, 本文的方法可以消除这些干扰, 产生准确的分割结果.

Figure 1 Visual examples of camouflaged object segmentation. While the state-of-the-art method SINet [10] confused by the background region which shares similar appearance with the camouflaged objects (pointed to by an arrow in the top row) or the camouflaged region that cluttered in the background (pointed to by an arrow in the bottom row), our method can eliminate these distractions and generate accurate segmentation results.

供了更丰富的消融研究和分析, 以揭示模型设计的合理性与有效性.

- 本文通过将 PFNet+ 模型应用于息肉分割任务, 进一步验证了它的泛化能力.

2 相关工作

在本节中, 首先简要回顾了不同领域中最先进的检测/分割方法, 包括普通目标检测、显著性目标检测、特定区域分割以及伪装物体分割, 然后讨论了一些关于上下文特征学习的代表性工作.

2.1 普通目标检测

普通目标检测 (Generic Object Detection, GOD) 是在自然图像中从多个预定义的一般类别中定位目标实例 [15], 是计算机视觉中最基本、最具挑战性的问题之一, 是解决复杂或高级视觉任务 (如分割 [16]、场景理解 [17] 和目标跟踪 [18]) 的基础. 场景中的普通目标可以是显著的, 也可以是伪装的, 伪装物体可以看作是难样本. 因此, 直接应用 GOD 方法 [19~21] 分割伪装物体可能达不到预期的效果.

2.2 显著性目标检测

显著性目标检测 (Salient Object Detection, SOD) 的目的是识别并分割输入图像中最引人注目的目标. 在过去的几十年中, 已经提出了数百种基于图像的显著性目标检测方法 [22]. 早期的方法主要是基于手工设计的低级特征以及启发式先验信息 (例如颜色 [23] 和对比度 [24]). 近年来, 深度卷积神经网络 (CNNs) 在显著性目标检测领域取得了新的进展. 多层次特征融合被用来进行鲁棒的检测 [25~28]. 循环学习和迭代学习策略也被用来逐步细化检测结果 [29, 30]. 由于对特征增强的有

效性, 注意力机制 [31, 32] 也被应用于显著性检测 [33~35]. 此外, 边界线索也被用来细化显著性检测结果 [36~39]. 然而, 将上述 SOD 方法应用于伪装物体分割可能不合适, 因为“显著”本质上与“伪装”是相反的.

2.3 特定区域分割

特定区域分割 (Specific Region Segmentation, SRS) 是指在场景中分割特定区域, 例如阴影 [40~43]、镜子 [44~46]、玻璃 [47~50] 和水 [51] 等区域. 这些区域是特殊的, 对视觉系统有着至关重要的影响. 对于水、阴影和镜子区域, 前景和背景之间通常存在光强或内容的不连续性. 相反, 伪装物体和背景的光强和内容都很相似, 这给伪装物体分割带来了很大的挑战. 此外, 与玻璃区域相比, 伪装物体通常具有更复杂的结构, 因此增加了准确分割的难度.

2.4 伪装物体分割

伪装物体分割 (Camouflaged Object Segmentation, COS) 在生物学和艺术等领域有着悠久而丰富的研究历史 [1], 这受益于两项杰出研究 [52, 53] 的巨大影响. 早期与伪装相关的工作致力于根据手工设计的纹理 [7]、三维凸度 [8] 和运动 [9] 等低级特征来区分前景和背景. 这些方法适用于一些简单的情况, 但在复杂的场景中往往失效. 最近, Le 等人 [11] 提出了一种将分类信息集成到像素级分割中的端到端伪装物体分割网络. Yan 等人 [12] 进一步引入了对抗性攻击来提高分割精度. Fan 等人 [10] 开发一个简单而有效的框架, 称为 SINet, 并构建了当前最大的伪装物体分割数据集 COD10K 来促进伪装物体分割在深度学习时代的发展. Lv 等人 [54] 提出了基于排序的方法来同时对伪装物体进行定位、分割、以及检测难度排序. Zhai 等人 [55] 实现了基于相互图学习的伪装物体分割方法. Li 等人 [56] 构建了一种利用矛盾信息来增强显著性物体检测和伪装物体检测的检测能力的框架. Pang 等人 [57] 设计了一个用于伪装物体分割的混合尺度三元组网络. Ji 等人 [58] 在反向重校准网络中引入了伪装物体的边缘先验信息. Zhuge 等人 [59] 通过在编解码器中引入一种特殊的 X 型连接, 以充分融合多层伪装特征. Cheng 等人 [60] 介绍了用于视频中伪装物体分割的隐式运动处理. Ji 等人 [61] 提出一种基于纹理梯度学习的高效框架用于分割伪装物体.

2.5 上下文特征学习

上下文特征学习 (Contextual Feature Learning) 在许多计算机视觉任务中都扮演着重要的角色. 许多工作致力于利用上下文来增强特征表示的能力. 具体地说, [62~64] 开发了多尺度的上下文, [65~67] 提取了多层次的上下文, [47, 68] 捕获了大视场上下文特征, [40] 探索了方向感知的上下文, [69] 开发了空间衰减上下文, [44, 46, 70] 利用了对比的上下文. 然而, 不加区分地探究上下文特征可能对伪装物体分割的贡献不大, 因为上下文往往会被显著目标的特征所支配. 与上述方法不同的是, 本文的方法侧重于从前景/背景注意的特征中探索上下文, 用于上下文的推理和分心区域的发现. 实验章节验证了本文方法的有效性.

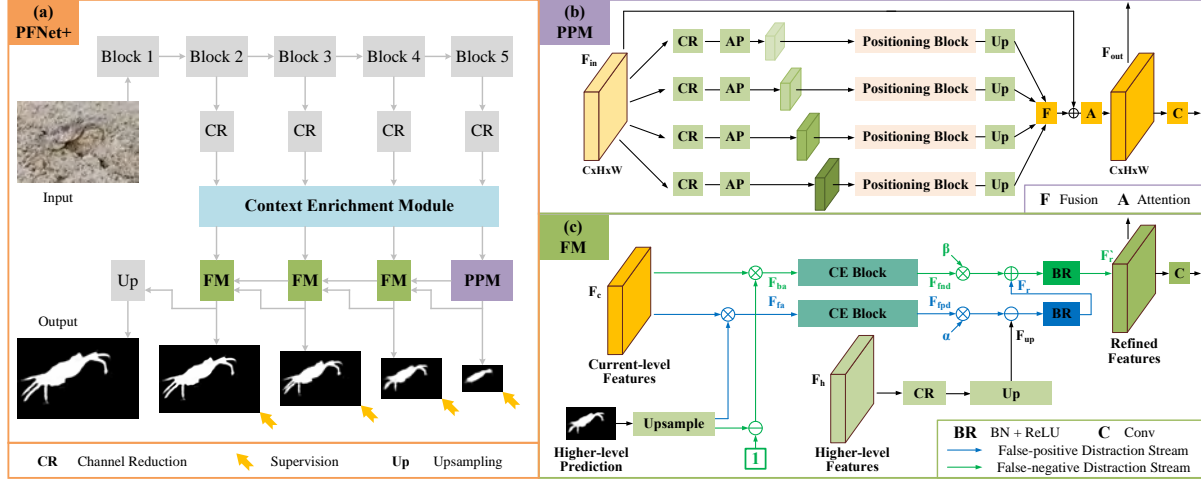


图 2 (a) 本文的金字塔定位与聚焦网络 (PFNet+) 及其两个主要组成模块: (b) 金字塔定位模块 (PPM) 和 (c) 聚焦模块 (FM)。

Figure 2 (a) Overview of our pyramid positioning and focus network (PFNet+) and its two main building blocks: (b) a pyramid positioning module (PPM) and (c) a focus module (FM).

3 方法

本文设计了一个金字塔定位和聚焦网络 (PFNet+), 其包含三个关键模块, 即上下文增强模块 (CEn)、金字塔定位模块 (PPM) 和聚焦模块 (FM)。CEn 通过整合上下文信息来增强骨干特征的代表能力, 从而提供更有辨别性的骨干特征; PPM 以金字塔的方式从全局的角度定位潜在的目标物体; 最后 FM 通过在歧义区域的聚焦来逐步细化粗糙的预测结果。

3.1 概述

本文提出的网络的结构如图2 (a) 所示。给定一幅 RGB 图像, 首先将其送入 ResNet-50 [71] 骨干网络中提取多级特征, 然后将这些特征送入四个卷积层中进行通道缩减。所得到的特征由上下文增强 (CEn) 模块加强。然后, 在最深层特征上应用金字塔定位模块 (PPM) 对潜在物体进行定位。最后, 利用多个聚焦模块 (FMs) 逐步发现和去除假阳性和假阴性干扰, 实现伪装物体的准确分割。

3.2 上下文增强

从骨干网络中提取的特征是包含低级局部细节和高层全局上下文的多层次信息。上下文增强 (CEn) 模块旨在充分利用多层次信息, 为网络的后续部分提供更有辨别性的表示。图3展示了精心设计的 CEn 模块的详细结构。给定输入的最后四级骨干特征 $B^i, i \in \{2, 3, 4, 5\}$, CEn 模块通过扩大感受野并捕捉/整合更多的上下文信息, 以提高骨干特征的代表能力, 从而输出上下文丰富的特征 $F^i, i \in \{2, 3, 4, 5\}$ 。具体来说, 本文从深层到浅层逐步处理骨干特征。首先, 通过对最高级别骨干特征 B^5 应用全局平均池化和两个卷积层来感知全局上下文, 然后通过元素级加法将全局上下文信息整合到 B^5 中。其中每个卷积层包括一个 1×1 卷积、一个批归一化 (BN) 和一个 ReLU 非线性操

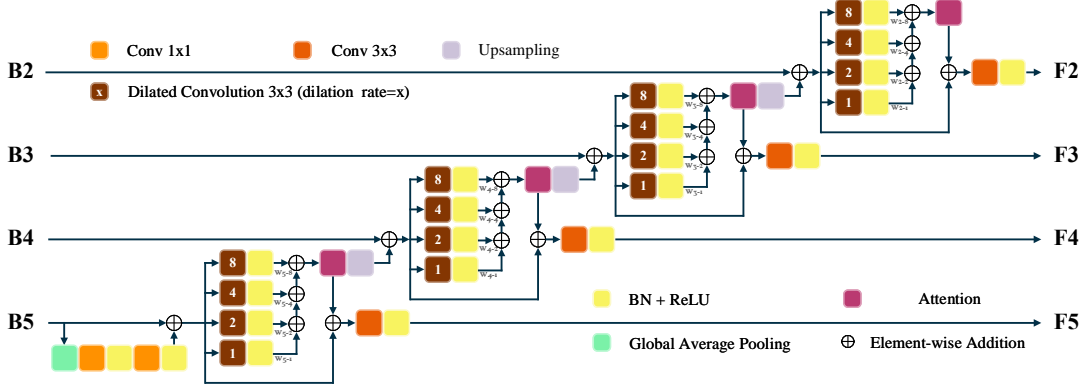


图 3 本文的上下文增强模块 (CEn) 的结构.

Figure 3 The architecture of our context enrichment (CEn) module.

作. 其次, 将特征输入到四个并行的具有不同扩张率的 3×3 卷积中, 以获取不同感受野的上下文. 每个扩张卷积后是一个 BN 层和一个 ReLU 运算. 然后, 提取的上下文通过加权元素级加法和一个注意块进行融合 [72], 并通过元素级加法操作进行整合. 在此基础上, 再通过卷积核为 3×3 的卷积层、BN 层和 ReLU 运算输出上下文丰富的最高级特征 F^5 . 然后, 本文对最高层的融合上下文进行上采样并整合到较浅层的特征 B^4 中, 以增强 B^4 的语义. 然后, 语义增强的特征通过上下文提取、融合和整合, 并生成 F^4 的输出特征. 这样, 骨干特征 $B^i, i \in \{2, 3, 4, 5\}$ 就逐渐被优化为更具辨别性的表示 $F^i, i \in \{2, 3, 4, 5\}$.

3.3 金字塔定位

图2 (b) 展示了精心设计的金字塔定位模块 (PPM) 的详细结构. 给定输入的最深层次特征, PPM 的目的是获取语义增强的深层次特征, 并进一步生成初始分割结果. 它利用四个分支的四个定位模块 (PMs) (图4) 来推断目标物体的初始位置, 以感知不同尺度下区域之间的相关性, 从而减少伪装物体的形状变化和位置的影响.

定位模块. PM 由通道注意块和空间注意块组成. 这两个块都是以非局部的方式实现的, 以获取通道和空间位置方面的长程依赖关系, 从全局角度增强特征的语义表示. 图4展示了定位模块 PM 的详细结构. 具体来说, 给定输入特征 $F \in \mathbb{R}^{C \times H \times W}$, 其中 C 、 H 和 W 分别表示特征的通道数量、高度和宽度, 本文先分别改变 F 的形状以分别得到查询 Q 、键 K 和值 V , 其中 $\{Q, K, V\} \in \mathbb{R}^{C \times N}$, $N = H \times W$ 是像素个数. 然后本文在 Q 和 K 的转置之间执行矩阵乘法, 并应用 SoftMax 层来得到通道注意图 $X \in \mathbb{R}^{C \times C}$:

$$x_{ij} = \frac{\exp(Q_{i:} \cdot K_{j:})}{\sum_{j=1}^C \exp(Q_{i:} \cdot K_{j:})}, \quad (1)$$

其中 $Q_{i:}$ 表示矩阵 Q 的第 i 行, x_{ij} 代表了第 j 个通道对于第 i 个通道的影响. 然后, 本文在 X 和 V 之间执行矩阵乘法, 并将整合的注意特征的形状改变为 $\mathbb{R}^{C \times H \times W}$. 最终, 为了提高容错能力, 本文将结果乘以一个可学习的比例参数 γ 并执行跳跃连接操作以获得最终输出 $F' \in \mathbb{R}^{C \times H \times W}$:

$$F'_{i:} = \gamma \sum_{j=1}^C (x_{ij} V_{j:}) + F_{i:}, \quad (2)$$

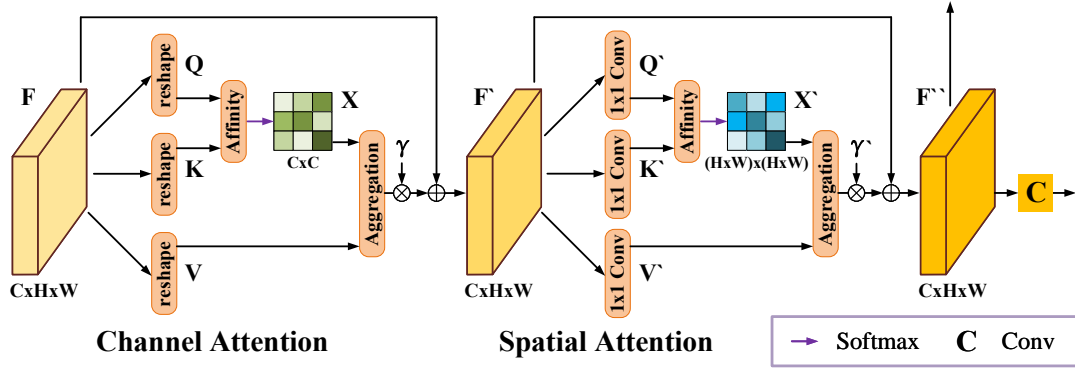


图4 本文的定位模块 (PM) 的结构.

Figure 4 The architecture of our positioning module (PM).

其中, γ 从初始值 1 逐渐学习权重. 最后的特征 F' 建模了特征图通道之间的长程语义依赖关系, 因此比输入特征 F 更具辨别性.

之后, 本文将通道注意块的输出特征作为空间注意块的输入. 本文首先在输入特征 F' 上应用三个 1×1 的卷积层并对卷积结果进行形状改变, 以分别生成三个新的特征 Q' 、 K' 和 V' , 其中 $\{Q', K'\} \in \mathbb{R}^{C_1 \times N}$, $C_1 = C/8$, 并且 $V' \in \mathbb{R}^{C \times N}$. 然后, 本文在 Q' 的转置和 K' 之间执行矩阵乘法, 并使用 SoftMax 归一化来生成空间注意图 $X' \in \mathbb{R}^{N \times N}$:

$$x'_{ij} = \frac{\exp(Q'_{i:} \cdot K'_{j:})}{\sum_{j=1}^N \exp(Q'_{i:} \cdot K'_{j:})}, \quad (3)$$

其中 $Q'_{i:}$ 表示矩阵 Q' 的第 i 列, x'_{ij} 代表了第 j 个位置对于第 i 个位置的影响. 此外, 本文在 V' 和 X' 的转置之间进行了矩阵乘法并将结果的形状改变为 $\mathbb{R}^{C \times H \times W}$. 类似于通道注意块, 本文将结果乘以一个可学习的比例参数 γ' 并使用跳跃连接以获得最终输出 $F'' \in \mathbb{R}^{C \times H \times W}$:

$$F''_{:i} = \gamma' \sum_{j=1}^N (V'_{j:} x'_{ji}) + F'_{:i}, \quad (4)$$

其中 γ' 也被初始化为 1. 在 F' 的基础上, F'' 进一步感知了各个位置之间的语义关联, 从而增强了特征的语义表示.

金字塔定位模块. 伪装物体的尺寸通常在很大范围内变化. 受 [63] 的启发, 本文将定位模块扩展为金字塔结构, 以减少形状变化带来的影响, 为不同尺寸的伪装物体提供更准确的初始位置. 如图2 (b) 所示, 金字塔定位模块 (PPM) 由四个分支组成, 每个分支依次进行通道缩减、自适应最大池化、定位和双线性上采样. 在每个分支中, 特征首先被处理成特定大小, 然后用于全局依赖性建模. 在特定的尺寸下, 特征的每个像素都是输入图像中特定尺寸区域的表示 (即具有特定尺寸的感受野). 如果一个像素的感受野不足以覆盖整个伪装物体, 那么 PM 可能会发生相关性感知不充分的情况. 另一方面, 如果这个感受野比小的伪装物体大得多, 覆盖伪装物体的像素的语义就会被背景所支配, 导致 PM 的相关性感知不理想. 因此, 本文改变自适应平均最大池化的目标大小, 实现金字塔的相关性感知. 对于 PPM 中的四个分支, 本文将送入 PM 的特征的边长分别设置为 5、7、9 和 11. 然后, 四个分支的输出特征通过以下通道维度上的串联操作进行融合和整合: 卷积层、跳

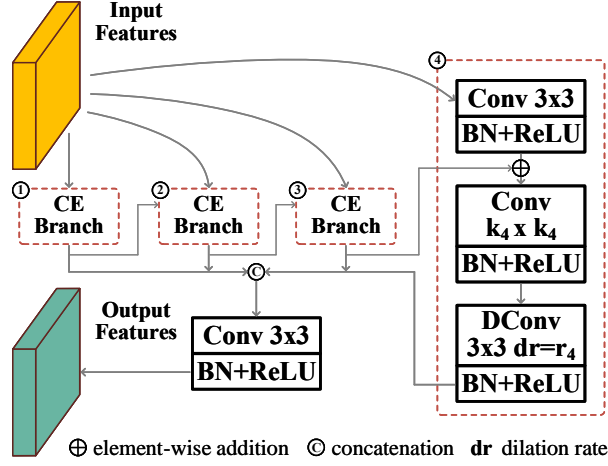


图 5 本文的上下文探索模块 (CEx) 的结构.

Figure 5 The architecture of our context exploration (CEx) block.

跃连接和注意块 [72], 并生成 PPM 的输出特征 F_{out} . 最后, 本文可以通过在 F_{out} 上应用填充为 3 的 7×7 卷积来得到目标的初始位置图. F_{out} 和初始位置图将由下面的聚焦模块 (FM) 逐步完善.

3.4 聚焦

伪装物体通常与背景具有相似的外观, 因此在初始分割结果中自然会出现假阳性和假阴性的预测. 聚焦模块 (FM) 的设计目的是发现并消除这些错误预测. FM 将当前级特征、上级特征和预测结果作为输入, 输出细化后的特征和更准确的预测结果.

分心发现. 本文注意到, 人类在经过仔细的分析之后可以很好地分辨分心区域. 本文的观察是人类会进行上下文推理, 即比较歧义区域和高置信度区域的模式, 例如纹理和语义, 来做出最终决策. 这启发本文对所有预测的前景 (或背景) 区域进行上下文探索, 以发现与高置信度前景 (或背景) 预测区域异质的假阳性分心区域 (或假阴性分心区域). 如图 2 (c) 所示, 本文首先对更高级别的预测进行上采样, 然后使用 Sigmoid 层对其进行归一化. 然后, 本文将此归一化图及其取反版本与当前级别的特征 F_c 相乘, 分别生成前景注意的特征 F_{fa} 和背景注意的特征 F_{ba} . 最后, 将这两种类型的特征送入两个并行的上下文探索 (CEx) 模块中来执行上下文推理, 以分别发现假阳性分心 F_{fpd} 和假阴性分心 F_{fnd} .

如图 5 所示, CEx 模块由四个上下文探索分支组成, 每个分支包括用于通道缩减的 3×3 卷积、用于局部特征提取的 $k_i \times k_i$ 卷积、以及用于上下文感知的卷积核为 3×3 且扩张率为 r_i 的扩张卷积. 本文分别将 $k_i, i \in \{1, 2, 3, 4\}$ 设置为 1、3、5、7, 并将 $r_i, i \in \{1, 2, 3, 4\}$ 设置为 1、2、4、8. 每个卷积后都跟有一个批归一化 (BN) 层和一个 ReLU 非线性运算. 第 $i, i \in \{1, 2, 3\}$ 个分支的输出将被送入到第 $(i+1)$ 个分支, 在更大的感受野中被进一步处理. 然后, 本文将所有四个分支的输出在通道维度上叠加, 并通过 3×3 的卷积进行融合. 通过这种设计, CEx 模块获得了在大范围内感知丰富上下文的能力, 因此可以用于上下文推理和分心发现.

分心去除. 在分心发现之后, 本文可以按以下方式进行分心去除:

$$\begin{aligned} F_{up} &= U(CBR(F_h)), \\ F_r &= BR(F_{up} - \alpha F_{fpd}), \\ F'_r &= BR(F_r + \beta F_{fnd}), \end{aligned} \quad (5)$$

其中, F_h 和 F_r 分别表示输入的上级特征和输出的精细特征, CBR 代表卷积、批归一化 (BN) 和 ReLU 的组合, U 是双线性上采样, α 和 β 是可学习的比例参数且初始值均为 1. 在这里, 本文使用逐元素减法运算来消除歧义的背景 (即假阳性分心) 和逐元素的加法操作来补充缺失的前景 (即假阴性干扰).

最后, 在细化后的特征上应用卷积层, 得到更准确的预测结果 F'_r . 本文使用真值图来监督生成的预测图, 来强迫 F'_r 成为一个更纯净的表达, 即分心去除的特征. 这会引导 CEx 模块发现特定形式的分心, 使整个聚焦模块以一种隐式的方式进行分心的发现和去除. 值得注意的是, 本文没有采用特定的分心图来显式地监督 F_{fpd} 和 F_{fnd} , 这基于以下两点考虑: (i) 标注假阳性和假阴性分心既昂贵又主观, 因此很难获得足够和有代表性的分心图; (ii) 对所有聚焦模块使用固定的分心图进行监控是次优的, 因为每个聚焦模块输入的上级特征是不同的, 本文希望发现和去除的分心应该随着逐渐细化的输入上级特征而动态变化.

讨论. 分心线索已在许多视觉任务中被探索, 例如显著性目标检测 [34,73]、语义分割 [74] 和视觉跟踪 [75]. 现有的工作利用假阳性分心 [73~75] 或假阴性分心 [34] 来获得更准确的结果. 与上述方法不同的是, 本文同时探索了这两种类型的分心, 并提出了一个精心设计的聚焦模块来发现并去除这些分心. 虽然 [43] 中的阴影分心感知模块同时考虑了两种类型的分心, 本文提出的聚焦模块在以下三个方面与该模块有着本质的区别. 首先, 阴影分心感知模块根据相同的输入特征来提取特征并预测两种类型的分心, 而本文的聚焦模块是从前景注意特征中发现假阳性分心, 从背景注意特征中发现假阴性分心. 其次, 阴影分心感知模块中的特征提取器仅包含两个 3×3 的卷积, 而本文的上下文探索模块由四个分支组成, 能够感知多尺度的上下文以更好地发现分心. 最后, 阴影分心感知模块的监督是根据现有阴影检测模型 [40~42] 的预测结果与真值之间的差异得到的, 这种显式的监督策略会受到具体方法的限制, 因而通用性有限. 相比之下, 本文设计了一个隐式的分心挖掘策略, 通过对分心去除的特征施加真值监督, 迫使每个上下文探索模块探索特定形式的分心. 据本文所知, 本文是第一个利用分心挖掘来解决伪装物体分割的工作, 本文所提出的分心挖掘策略可为其他视觉任务提供启发.

3.5 损失函数

PFNet+ 中有四个输出结果, 其中一个来自金字塔定位模块 (PPM), 三个来自聚焦模块 (FM). 对于 PPM, 本文对其输出使用二值交叉熵 (BCE) 损失 ℓ_{bce} 和交并比损失 ℓ_{iou} [36], 即 $\mathcal{L}_{ppm} = \ell_{bce} + \ell_{iou}$, 来引导定位模块探索目标物体的初始位置. 对于聚焦模块, 本文希望它能更多地关注分心区域, 这类区域通常位于物体的边界、细长区域或孔处, 因此本文结合加权 BCE 损失 ℓ_{wbce} [76] 和加权 IoU 损失 ℓ_{wioU} [76], 即 $\mathcal{L}_{fm} = \ell_{wbce} + \ell_{wioU}$, 来迫使聚焦模块将注意力放在可能的分心区

域. 最后, 总体损失函数为:

$$\mathcal{L}_{overall} = \mathcal{L}_{ppm} + \sum_{i=2}^4 2^{(4-i)} \mathcal{L}_{fm}^i, \quad (6)$$

其中, \mathcal{L}_{fm}^i 表示对于 PFNet+ 中第 i 级聚焦模块的输出的损失.

4 伪装物体分割实验

4.1 实验设置

数据集. 本文在三个基准数据集上评估了本文的方法: CHAMELEON [77]、CAMO [11] 以及 COD10K [10]. CHAMELEON [77] 包含 76 张通过谷歌搜索引擎、使用“伪装动物”作为关键词从互联网上收集的图片, 以及相应的人工标注的目标级真值. CAMO [11] 包含 1250 张不同类别的伪装图像, 分为 1000 张训练图像和 250 张测试图像. COD10K [10] 是目前最大的基准数据集, 它包括从多个摄影网站下载的 5066 张伪装图片 (其中 3040 张用于训练, 2026 张用于测试), 涵盖 5 个大类和 69 个子类. 本文仿照之前工作 [10] 的做法, 使用 CAMO [11] 和 COD10K [10] 的训练集作为训练集 (4040 张图片), 其余图片作为测试集.

评估指标. 本文使用四个广泛使用的标准度量来评估本文的方法: 结构度量 (S_α) [78]、自适应 E 度量 (E_ϕ^{ad}) [79]、加权 F 度量 (F_β^ω) [80] 以及平均绝对误差 (M). S_α 着重评估预测图的结构信息, 其定义为: $S_\alpha = \alpha S_o + (1-\alpha) S_r$, 其中 S_o 和 S_r 分别表示物体感知和区域感知的结构相似性, α 同 [78] 一样被设置为 0.5. E_ϕ 同时评估像素级匹配和图像级统计信息, 其被证明与人类的视觉感知相关 [79]. 因此, 本文使用这个指标来评估伪装物体分割结果的整体和局部的精度. F_β 是一个综合的关于预测图精确度和召回率的评估方法. 最近的研究 [78, 79] 证明加权的 F_β (即 F_β^ω [80]) 能提供更可靠的评估结果. 因此, 本文在比较中也考虑了这个指标. 平均绝对误差 (M) 广泛应用于前景背景分割任务中, 它计算预测图和真值之间的像素差异.

实现细节. 本文采用 PyTorch [81] 实现提出的模型. 训练和测试均使用一台 8 核电脑, 配备 Intel core i7-9700K 3.6 GHz CPU (64GB RAM) 和 NVIDIA GeForce RTX 2080Ti GPU (11GB 显存). 模型训练阶段, 输入图像的大小被调整为 480×480 , 并通过随机水平翻转和颜色抖动进行数据扩充. 编码器网络的参数由预先在 ImageNet 上训练的 ResNet-50 模型 [71] 初始化, PFNet+ 中的其余层则随机初始化. 本文使用动量为 0.9 且权重衰减率为 5×10^{-4} 的随机梯度下降 (SGD) 优化器进行网络优化. 本文将批大小设置为 12, 并通过 poly 策略 [82] 调整学习率, 其中基础学习率为 0.001, 幂次为 0.9. 网络训练 60 轮后收敛, 耗时仅需约 3 小时. 测试阶段, 本文首先将图像的大小调整为 480×480 以进行网络推断, 然后将输出图的大小调整回输入图像的原始大小. 两次调整大小的过程都使用双线性插值. 本文不使用任何后处理 (如全连接的条件随机场 (CRF) [83]) 来进一步增强最终输出. 480×480 大小图像的推断仅需 0.018 秒 (约 56 帧/秒).

4.2 和最新方法的比较

为了证明本文的 PFNet+ 的有效性, 本文将其与 20 个最新方法进行了比较: 目标检测方法 FPN [19], 语义分割方法 PSPNet [63], 实例分割方法 Mask RCNN [20]、HTC [84] 以及 MSRCNN [21],

阴影检测方法 DSC [40] 以及 BDRAR [42], 医学图像分割方法 UNet++ [85] 以及 PraNet [2], 显著性目标检测方法 PiCANet [33]、BASNet [36]、CPD [86]、PFANet [87]、EGNet [37]、F3Net [76]、GCPANet [88] 以及 MINet-R [89] 和伪装物体分割方法 SINet [10]、ERRNet [58] 以及 PFNet [14]. 其中, FPN [19] 利用了深度卷积网络固有的多尺度、金字塔层次结构来构建特征金字塔; PSPNet [63] 采用金字塔池化模块来挖掘全局上下文信息; Mask RCNN [20] 在目标检测框架的基础上增加 mask 预测分支, 实现了实例分割; HTC [84] 通过将检测和分割特征交织在一起以进行联合多阶段处理, 将级联集成到实例分割中; MSRCNN [21] 能够学习预测实例掩码的质量; DSC [40] 以方向感知的方式分析图像上下文; BDRAR [42] 利用跨深层和浅层的空间上下文, 以及迭代地整合上下文; UNet++ [85] 是一种基于编码器-解码器结构的神经网络框架, 其中编码器和解码器子网络通过一系列嵌套的、密集的跳过路径连接以弥合特征间的语义差异; PraNet [2] 是一种平行反注意网络, 能够整合高层语义信息和挖掘边界信息; PiCANet [33] 是一种新颖的像素级上下文注意力网络, 选择性地关注每个像素的信息性上下文位置; BASNet [36] 和 EGNet [37] 探索了边界信息以提升方法性能; CPD [86]、PFANet [87] 以及 GCPANet [88] 利用全局上下文信息引导分割过程; MINet-R [89] 是针对分割任务中存在的多尺度问题而设计的一种具有转换-交互-融合策略的方法. 为了公平比较, 上述方法的所有预测图由公共网站提供, 或者通过运行利用开源代码重新训练的模型生成. 此外, 所有的预测图都使用相同的代码进行评估.

定量评价. 表1报告了 PFNet+ 与其他 20 种最新方法在三个基准数据集上的定量结果. 可以看到, 本文的方法在所有四个标准评估指标下都优于所有其他方法. 例如, 与最先进的伪装物体分割方法 SINet [10] 相比, 本文的方法将 F_{β}^w 在 CHAMELEON [77]、CAMO [11]、以及 COD10K [10] 数据集上分别提升了 9.8%、10.7%、以及 12.6%. 值得注意的是, 本文的方法也比 SINet 更快 (即 56 帧/秒对比 51 帧/秒).

定性评价. 图6展示了本文的方法与其他方法的定性比较结果. 可以看出, 本文的方法能够准确地分割出小的伪装物体 (前两行)、大的伪装物体 (第三行和第四行)、以及多个伪装物体 (第五行和第六行). 这主要是因为金字塔定位模块能够通过挖掘长程语义依赖关系, 为后续的分心挖掘提供不同尺度伪装物体的初始位置. 现有的方法通常会被和伪装物体具有相似外观的背景 (第七行) 或混杂于背景中的前景区域 (第八行) 所困扰, 相比之下, 本文的方法可以成功地推断出真实的伪装物体区域. 这主要得益于本文提出的分心挖掘策略, 该策略有助于抑制假阳性分心区域和补充假阴性分心区域. 此外, 受益于上下文增强和分心发现过程中的多尺度上下文探索, 本文的方法能够获得精细的分心信息, 从而能够对具有复杂结构的伪装物体进行精细的分割 (最后一行).

4.3 消融分析

本文进行消融实验, 以验证为准确伪装物体分割量身定做的两个关键部件 (即定位模块 (PM) 和聚焦模块 (FM)) 以及两个扩展模块 (即上下文增强模块 (CEn) 和金字塔定位模块 (PPM)) 的有效性, 并将结果报告在表2、表3、表4、表5和图7、图8、图9中.

定位模块的有效性. 从表2中可以看出, 在基础模型 (a) 上引入通道注意块 (b) 或空间注意块 (c) 可以在一定程度上提高分割性能, 二者的结合 (d) 可以获得更好的分割效果. 这证实了定位模块有利于伪装目标的准确分割. 此外, 为了验证在 PM 中使用非局部结构的优势, 本文将 PM 中的

表 1 本文提出的方法与相关的 20 种最新方法在三个基准数据集上四个评估指标 (即结构度量 S_α (越大越好)、自适应 E 度量 E_ϕ^{ad} (越大越好)、加权 F 度量 F_β^ω (越大越好)、以及平均绝对误差 M (越小越好)) 下的比较结果. 所有预测结果都使用相同的代码进行评估. 最好的结果已加粗显示. \circ : 目标检测方法, \bullet : 语义分割方法, $*$: 实例分割方法, Δ : 阴影检测方法, \S : 医学图像分割方法, \dagger : 显著性目标检测方法, $*$: 伪装物体分割方法. 在所有三个基准数据集上, 本文的方法在所有四个标准评估指标下都比其他方法有很大的优势.

Table 1 Comparison of our proposed method and other 20 state-of-the-art methods in the relevant fields on three benchmark datasets in terms of the structure-measure S_α (larger is better), the adaptive E-measure E_ϕ^{ad} (larger is better), the weighted F-measure F_β^ω (larger is better), and the mean absolute error M (smaller is better). All the prediction maps are evaluated with the same code. The best results are marked in bold. \circ : object detection method. \bullet : semantic segmentation method. $*$: instance segmentation methods. Δ : shadow detection methods. \S : medical image segmentation methods. \dagger : sod methods. $*$: cos methods. Our method outperforms other counterparts with a large margin under all four standard evaluation metrics on all three benchmarks.

Methods	Pub.'Year	CHAMELEON (76 images)				CAMO-Test (250 images)				COD10K-Test (2,026 images)			
		$S_\alpha \uparrow$	$E_\phi^{ad} \uparrow$	$F_\beta^\omega \uparrow$	$M \downarrow$	$S_\alpha \uparrow$	$E_\phi^{ad} \uparrow$	$F_\beta^\omega \uparrow$	$M \downarrow$	$S_\alpha \uparrow$	$E_\phi^{ad} \uparrow$	$F_\beta^\omega \uparrow$	$M \downarrow$
FPN $^\circ$ [19]	CVPR'17	0.794	0.835	0.590	0.075	0.684	0.791	0.483	0.131	0.697	0.711	0.411	0.075
PSPNet $^\bullet$ [63]	CVPR'17	0.773	0.814	0.555	0.085	0.663	0.778	0.455	0.139	0.678	0.688	0.377	0.080
Mask RCNN $*$ [20]	ICCV'17	0.643	0.780	0.518	0.099	0.574	0.716	0.430	0.151	0.613	0.750	0.402	0.080
UNet++ § [85]	DLMIA'17	0.695	0.808	0.501	0.094	0.599	0.740	0.392	0.149	0.623	0.718	0.350	0.086
DSC $^\Delta$ [40]	CVPR'18	0.850	0.888	0.714	0.050	0.736	0.830	0.592	0.105	0.758	0.788	0.542	0.052
PiCANet † [33]	CVPR'18	0.769	0.836	0.536	0.085	0.609	0.753	0.356	0.156	0.649	0.678	0.322	0.090
BDRAR $^\Delta$ [42]	ECCV'18	0.779	0.881	0.663	0.064	0.759	0.825	0.664	0.093	0.753	0.836	0.591	0.051
HTC $*$ [84]	CVPR'19	0.517	0.490	0.204	0.129	0.476	0.442	0.174	0.172	0.548	0.521	0.221	0.088
MSRCNN $*$ [21]	CVPR'19	0.637	0.688	0.443	0.091	0.617	0.670	0.454	0.133	0.641	0.708	0.419	0.073
BASNet † [36]	CVPR'19	0.687	0.742	0.474	0.118	0.618	0.719	0.413	0.159	0.634	0.676	0.365	0.105
CPD † [86]	CVPR'19	0.853	0.878	0.706	0.052	0.726	0.802	0.550	0.115	0.747	0.763	0.508	0.059
PFANet † [87]	CVPR'19	0.679	0.732	0.378	0.144	0.659	0.735	0.391	0.172	0.636	0.619	0.286	0.128
EGNet † [37]	ICCV'19	0.848	0.879	0.702	0.050	0.732	0.827	0.583	0.104	0.737	0.777	0.509	0.056
F3Net † [76]	AAAI'20	0.854	0.899	0.749	0.045	0.779	0.840	0.666	0.091	0.786	0.832	0.617	0.046
GCPANet † [88]	AAAI'20	0.876	0.891	0.748	0.041	0.778	0.842	0.646	0.092	0.791	0.799	0.592	0.045
PraNet § [2]	MICCAI'20	0.860	0.898	0.763	0.044	0.769	0.833	0.663	0.094	0.789	0.839	0.629	0.045
MINet-R † [89]	CVPR'20	0.844	0.919	0.746	0.040	0.749	0.835	0.635	0.090	0.759	0.832	0.580	0.045
SINet $*$ [10]	CVPR'20	0.869	0.899	0.740	0.044	0.751	0.834	0.606	0.100	0.771	0.797	0.551	0.051
ERRNet $*$ [58]	PR'22	0.868	0.917	0.787	0.039	0.779	0.851	0.679	0.085	0.786	0.845	0.630	0.043
PFNet $*$ [14]	CVPR'21	0.882	0.942	0.810	0.033	0.782	0.852	0.695	0.085	0.800	0.868	0.660	0.040
PFNet+$*$	Ours	0.893	0.946	0.838	0.028	0.791	0.861	0.713	0.080	0.806	0.879	0.677	0.037

非局部注意替换为传统注意 (使用卷积和全连接层来生成空间和通道注意力) .. 从表3中显示的结果可以看出, 非局部注意的效果更好, 这证实了长程语义依赖关系有利于伪装目标的准确分割.

聚焦模块的有效性. 在 (a) 的基础上, 引入本文提出的假阳性分心挖掘 (e) 或假阴性分心挖掘 (f), 将大大提高分割效果. 同时考虑到两种类型的分心, 即 (h), 本文获得了更好的结果. 例如, 引入聚焦模块分别将 E_ϕ^{ad} 和 F_β^ω 提升了 5.7% 和 5.8%. 实验结果表明, 聚焦模块使本文的方法具有很强的对伪装物体进行准确分割的能力. 当去除来自上级预测的引导时, 即 (g), 分割性能会有一定程度的下降. 这是因为不加区分地从输入特征中挖掘分心会增加分心发现的难度, 从而阻碍了分心的有效去除. 这验证了本文的从注意的输入特征中学习分心的设计的合理性. 根据实验结果 (i-l), 可以看到当增加部分或者全部的聚焦模块时, 上述结论依然成立. 此外, 本文将最后一个聚焦模块中

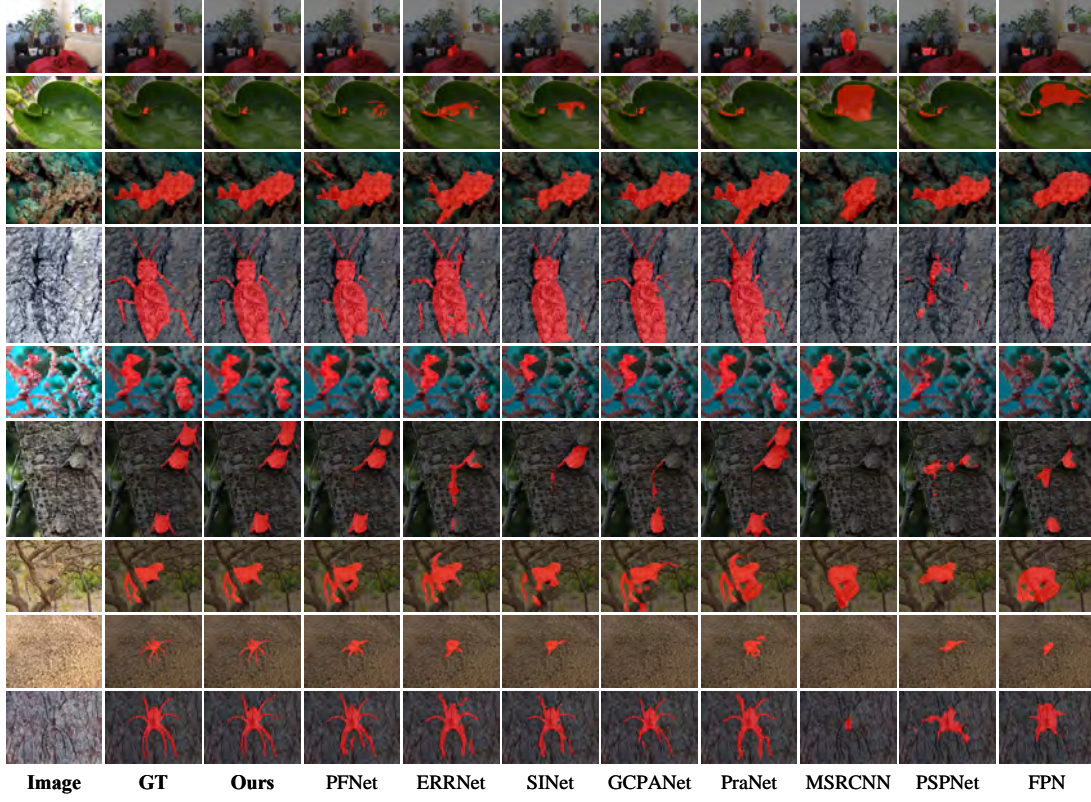


图 6 本文模型与最新方法的视觉比较. 显然, 本文的方法能够更准确地分割不同环境下的各种伪装物体.

Figure 6 Visual comparison of the proposed model with state-of-the-art methods. Obviously, our approach is capable of segmenting various camouflaged objects concealed in different environments more accurately.

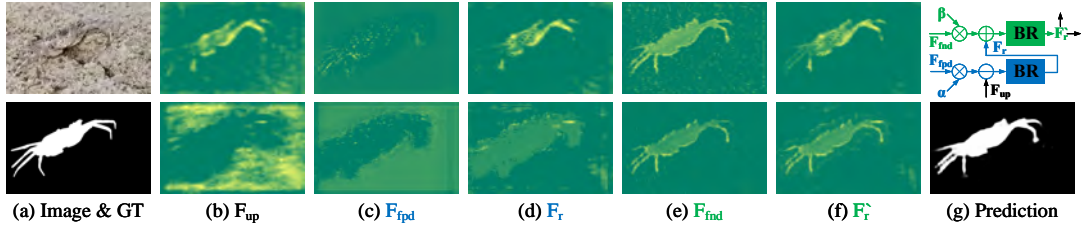


图 7 最后一个聚焦模块中的特征图可视化结果, 最好以彩色和放大方式观看.

Figure 7 Visualizing feature maps in the last FM. Best viewed in color and zoomed-in.

的特征图进行了可视化 (参见图7). 通过挖掘假阳性分心 (c), 可以极大地抑制 (b) 中的假阳性预测 (d). 通过挖掘假阴性分心 (e), 本文可以得到伪装物体的更为纯净的表示 (f). 这清晰地证明了所提出的旨在发现和去除分心的分心挖掘策略的有效性.

为了进行隐式分心监督与显式分心监督的比较, 本文首先根据最先进的 COS 方法 SINET [10] 的预测结果和真值之间的差异获得 F_{fpd} 和 F_{fnd} 的监督图, 然后利用这些“真值”分心图来训练本文的模型, 并将结果显示在表4中. 由表4可以看出, 使用显式策略的表现不如使用隐式策略的表现. 这是因为本文设计的隐式策略使网络摆脱了只能消除由特定方法或人类定义的固定分心的限制, 从而有助于网络更好地应对不同的挑战性场景.

表 2 消融分析. “B” 代表在本文的网络中从定位模块 (“PM”) 中移除通道注意块 (“CA”) 和空间注意块 (“SA”) 并将聚焦模块 (“FM”) 中的假阳性分心分支 (“FPD”) 和假阴性分心分支 (“FND”) 替换为简单的跳跃连接. “w/o A” 表示在聚焦模块中没有将上级预测结果作为注意图来引导当前级特征. 可以观察到, 每一个提出的部件都能对分割性能做出贡献.

Table 2 Ablation analyses. “B” denotes our network with the channel attention block (“CA”) and spatial attention block (“SA”) removed from positioning module (“PM”) and the false-positive distraction stream (“FPD”) and false-negative distraction stream (“FND”) in the focus module (“FM”) replaced by a simple skip-connection. “w/o A” denotes that the higher-level prediction is not used as the attention map to guide the current-level features in the focus module. As can be observed, each proposed component plays an important role and contributes to the performance.

Networks	B	PM		FM			COD10K-Test			
		CA	SA	FPD	FND	w/o A	$S_\alpha \uparrow$	$E_\phi^{ad} \uparrow$	$F_\beta^\omega \uparrow$	$M \downarrow$
(a)	✓						0.779	0.803	0.591	0.051
(b)	✓	✓					0.788	0.819	0.618	0.046
(c)	✓		✓				0.791	0.826	0.624	0.046
(d)	✓	✓	✓				0.792	0.835	0.631	0.045
(e)	✓			✓			0.790	0.844	0.632	0.043
(f)	✓				✓		0.790	0.837	0.628	0.043
(g)	✓			✓	✓	✓	0.796	0.843	0.639	0.042
(h)	✓			✓	✓		0.797	0.860	0.649	0.041
(i)	✓	✓	✓	✓			0.796	0.854	0.645	0.042
(j)	✓	✓	✓		✓		0.796	0.847	0.644	0.043
(k)	✓	✓	✓	✓	✓	✓	0.796	0.851	0.647	0.042
(l)	✓	✓	✓	✓	✓		0.800	0.868	0.660	0.040

表 3 在 PM 中使用不同注意力机制之间的比较

Table 3 Comparison between attention mechanisms used in the PM

PFNet	COD10K-Test			
	$S_\alpha \uparrow$	$E_\phi^{ad} \uparrow$	$F_\beta^\omega \uparrow$	$M \downarrow$
(i) w/ traditional attention	0.796	0.849	0.642	0.042
(ii) w/ non-local attention	0.800	0.868	0.660	0.040

表 4 在 FM 中使用不同分心监督之间的比较

Table 4 Comparison between distraction supervision strategies used in the FM

PFNet	COD10K-Test			
	$S_\alpha \uparrow$	$E_\phi^{ad} \uparrow$	$F_\beta^\omega \uparrow$	$M \downarrow$
(A) w/ explicit supervision	0.794	0.853	0.641	0.042
(B) w/ implicit supervision	0.800	0.868	0.660	0.040

上下文增强模块的有效性. 表5的 (1) 和 (2) 以及图8的第二列和第三列分别定量和定性的对比了添加 CEn 和不添加 CEn 时伪装物体分割的结果, 突出了丰富的上下文对提高 COS 性能的作用. CEn 模块探索了丰富的上下文信息, 便于后续对伪装物体的定位和边界划定, 从而实现更精确的伪装物体分割.

金字塔定位模块的有效性. 本文用 PPM 代替 PFNet 中的 PM 进行了一次实验, 以验证 PPM 相比于 PM 的优势. 表5展示了实验的定量结果, 同时图9中展示的定性比较结果也支撑着该结论. 当 PM 可能会被杂乱的背景区域 (如图9的第一行) 或伪装的前景区域 (如图9的第二行) 所迷惑时, PPM 仍可以借助金字塔式的定位方式帮助获得更准确的定位结果.

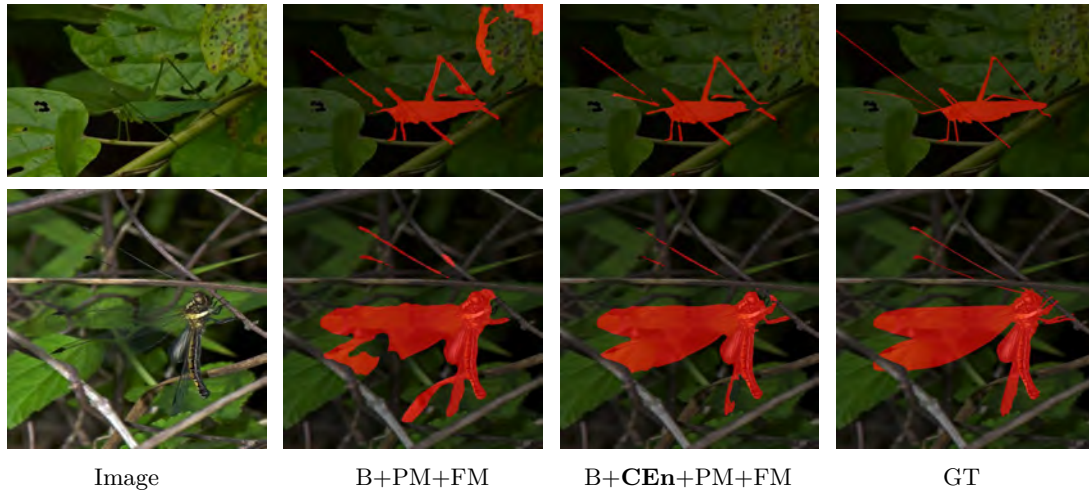


图 8 上下文增强模块有效性的可视化结果.

Figure 8 Visual demonstration of the effectiveness of CEn.

表 5 消融分析. “CEn” 和 “PPM” 表示上下文增强模块和金字塔定位模块.

Table 5 Ablation analyses. “CEn” and “PPM” denote the context enrichment and pyramid positioning module.

Networks	COD10K-Test			
	$S_{\alpha} \uparrow$	$E_{\phi}^{ad} \uparrow$	$F_{\beta}^{\omega} \uparrow$	$M \downarrow$
(1) B + PM + FM (即 PFNet)	0.800	0.868	0.660	0.040
(2) B + CEn + PM + FM	0.806	0.873	0.674	0.039
(3) B + PPM + FM	0.804	0.874	0.670	0.038
(4) B + CEn + PPM + FM (即 PFNet+)	0.806	0.879	0.677	0.037

5 息肉分割实验

本文提出的 PFNet+ 借助精心设计的上下文增强模块、金字塔定位模块和聚焦模块, 可以分割隐藏在不同环境中的各种伪装物体, 因此有潜力处理其他具有挑战性的视觉任务/应用. 在本节中, 本文考虑息肉分割任务. 息肉是指人体组织表面长出的赘生物, 可在结肠、直肠、胃甚至喉咙中发现. 由于息肉具有癌变的可能性, 因此监测息肉生长的大小、位置以及是否成为恶性肿瘤是非常重要的. 结肠息肉由于其固有的隐蔽性, 可以被看作是一种伪装的物体. 因此, 本文对息肉分割任务进行实验, 以验证本文方法的泛化能力.

5.1 实验设置

本文遵循 [2, 90] 使用从 Kvasir [91] 和 CVC-ClinicDB [92] 中选择的 1450 对图像作为训练集. 本文选择广泛使用的 CVC-300 数据集用于测试, 它是 EndoScene [93] 中的一个测试数据集. 本文将为息肉分割重新训练的 PFNet+ 命名为 PFNet+^p. 模型训练阶段, 输入图像的大小被调整为 480×480 , 并通过随机水平翻转和颜色抖动进行数据扩充, 批大小设置为 10, 网络训练 100 轮后收敛, 耗时需约 90 分钟. 测试阶段, 本文首先将图像的大小调整为 480×480 以进行网络推断, 然后将输出

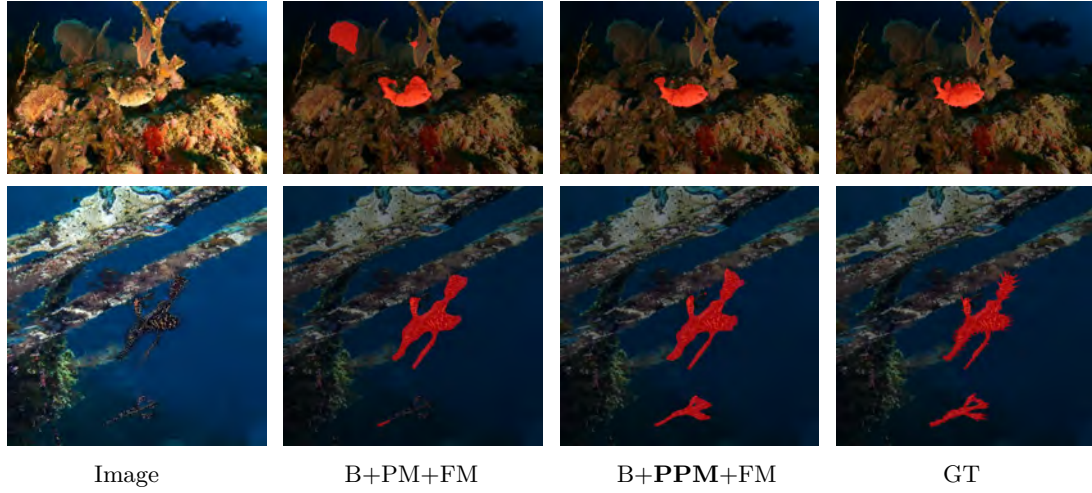


图 9 金字塔定位模块有效性的可视化结果.

Figure 9 Visual demonstration of the effectiveness of PPM.

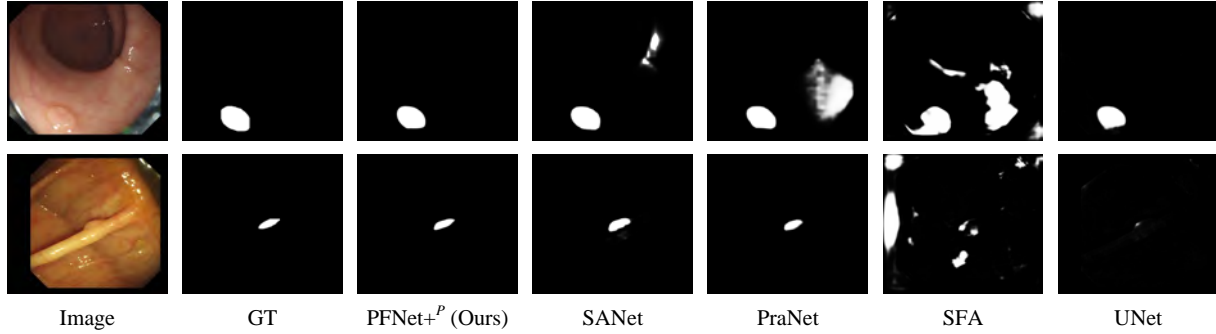


图 10 本文模型与最新息肉分割方法的视觉比较.

Figure 10 Visual comparison of the proposed model with state-of-the-art polyp segmentation methods.

图的大小调整回输入图像的原始大小. 两次调整大小的过程都使用双线性插值. 本文使用在息肉分割任务中广泛使用的标准度量来评估本文的方法: 结构度量 (S_α) [78]、自适应 E 度量 (E_ϕ^{ad}) [79]、加权 F 度量 (F_β^ω) [80] 以及平均绝对误差 (M).

5.2 和最新方法的比较

本文将 $\text{PFNet}+^P$ 与 9 个最新方法进行了比较: 包括 UNet [94]、UNet++ [85]、SFA [95]、MSEG [96]、DCRNet [97]、PraNet [2]、EU-Net [98]、SANet [99]、PolypPVT [90]. 为了公平比较, 上述方法的所有息肉分割预测图由作者提供, 或者通过运行利用开源代码重新训练的模型生成. 此外, 所有的预测图都使用相同的代码进行评估. 表6和图10分别展示了定量和定性比较结果, 可以看到 $\text{PFNet}+^P$ 在息肉分割上有很好的效果, 清晰地证明了 $\text{PFNet}+^P$ 的泛化能力.

表 6 本文提出的方法与相关的 9 种最新息肉分割方法在 CVC-300 基准数据集上四个评估指标 (即结构度量 S_α (越大越好)、自适应 E 度量 E_ϕ^{ad} (越大越好)、加权 F 度量 F_β^ω (越大越好)、以及平均绝对误差 M (越小越好)) 下的比较结果. 所有预测结果都使用相同的代码进行评估. 最好的结果已加粗显示. 本文的方法有很好的效果.

Table 6 Comparison of our method and other 9 state-of-the-art polyp segmentation methods on the CVC-300 benchmark dataset in terms of the structure-measure S_α (larger is better), the adaptive E-measure E_ϕ^{ad} (larger is better), the weighted F-measure F_β^ω (larger is better), and the mean absolute error M (smaller is better). all the prediction maps are evaluated with the same code. the best results are marked in bold. our method achieves superior segmentation performance.

Methods	Pub.'Year	CVC-300			
		$S_\alpha \uparrow$	$E_\phi^{ad} \uparrow$	$F_\beta^\omega \uparrow$	$M \downarrow$
UNet [94]	MICCAI'15	0.843	0.867	0.684	0.022
UNet++ [85]	DLMIA'18	0.839	0.884	0.687	0.018
SFA [95]	MICCAI'19	0.640	0.604	0.341	0.065
MSEG [96]	arXiv'21	0.924	0.951	0.852	0.009
DCRNet [97]	arXiv'21	0.921	0.934	0.830	0.010
PraNet [2]	MICCAI'20	0.925	0.938	0.843	0.010
EU-Net [98]	CRV'21	0.904	0.915	0.805	0.015
SANet [99]	MICCAI'21	0.928	0.948	0.859	0.008
PolypPVT [90]	arXiv'21	0.935	0.973	0.884	0.007
PFNet+^p	Ours	0.945	0.969	0.890	0.006

6 结论

本文致力于实现准确的伪装物体分割. 本文设计了一种新颖的金字塔定位和聚焦网络(PFNet+), 其通过分心挖掘策略来发现和去除分心, 在三个基准数据集上实现了顶尖的伪装物体分割性能. 未来工作包括探索该方法在其他领域 (例如通用物体分割或透明物体分割) 的应用潜力, 并尝试改进模型使之能够准确分割视频中的伪装物体.

参考文献

- 1 M. Stevens and S. Merilaita, "Animal camouflage: current issues and new perspectives," *Philosophical Transactions of the Royal Society B*, vol. 364, no. 1516, pp. 423–427, 2009.
- 2 D.-P. Fan, G.-P. Ji, T. Zhou, G. Chen, H. Fu, J. Shen, and L. Shao, "Pranet: Parallel reverse attention network for polyp segmentation," in *MICCAI*, 2020, pp. 263–273.
- 3 G.-P. Ji, G. Xiao, Y.-C. Chou, D.-P. Fan, K. Zhao, G. Chen, and L. Van Gool, "Video polyp segmentation: A deep learning perspective," *Machine Intelligence Research*, vol. 19, no. 6, p. 531–549, 2022.
- 4 G.-P. Ji, Y.-C. Chou, D.-P. Fan, G. Chen, D. Jha, H. Fu, and L. Shao, "Progressively normalized self-attention network for video polyp segmentation," in *MICCAI*, 2021, pp. 142–152.
- 5 D.-P. Fan, T. Zhou, G.-P. Ji, Y. Zhou, G. Chen, H. Fu, J. Shen, and L. Shao, "Inf-net: Automatic covid-19 lung infection segmentation from ct images," *IEEE TMI*, vol. 39, no. 8, pp. 2626–2637, 2020.
- 6 T. Troscianko, C. P. Benton, P. G. Lovell, D. J. Tolhurst, and Z. Pizlo, "Camouflage and visual perception," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 364, no. 1516, pp. 449–461, 2009.
- 7 P. Sengottuvelan, A. Wahi, and A. Shanmugam, "Performance of decamouflaging through exploratory image analysis," in *ETET*, 2008, pp. 6–10.
- 8 Y. Pan, Y. Chen, Q. Fu, P. Zhang, and X. Xu, "Study on the camouflaged target detection method based on 3d convexity," *Modern Applied Science*, vol. 5, no. 4, p. 152, 2011.
- 9 J. Yin, Y. Han, W. Hou, and J. Li, "Detection of the mobile object with camouflage color under dynamic background based on optical flow," *Procedia Engineering*, vol. 15, pp. 2201–2205, 2011.
- 10 D.-P. Fan, G.-P. Ji, G. Sun, M.-M. Cheng, J. Shen, and L. Shao, "Camouflaged object detection," in *CVPR*, 2020, pp. 2777–2787.
- 11 T.-N. Le, T. V. Nguyen, Z. Nie, M.-T. Tran, and A. Sugimoto, "Anabranch network for camouflaged object segmentation," *CVIU*, vol. 184, pp. 45–56, 2019.
- 12 J. Yan, T.-N. Le, K.-D. Nguyen, M.-T. Tran, T.-T. Do, and T. V. Nguyen, "Mirrornet: Bio-inspired camouflaged object segmentation," *IEEE Access*, vol. 9, pp. 43 290–43 300, 2021.

- 13 J. R. Hall, I. C. Cuthill, R. J. Baddeley, A. J. Shohet, and N. E. Scott-Samuel, "Camouflage, detection and identification of moving targets," *RSBS*, vol. 280, no. 1758, p. 20130064, 2013.
- 14 H. Mei, G. Ji, Z. Wei, X. Yang, X. Wei, and D. Fan, "Camouflaged object segmentation with distraction mining," in *CVPR*, 2021, pp. 8772–8781.
- 15 L. Liu, W. Ouyang, X. Wang, P. Fieguth, J. Chen, X. Liu, and M. Pietikinen, "Deep learning for generic object detection: A survey," *IJCV*, vol. 128, no. 2, pp. 261–318, 2020.
- 16 A. Kirillov, K. He, R. Girshick, C. Rother, and P. Dollar, "Panoptic segmentation," in *CVPR*, 2019, pp. 9404–9413.
- 17 L.-J. Li, R. Socher, and L. Fei-Fei, "Towards total scene understanding: Classification, annotation and segmentation in an automatic framework," in *CVPR*, 2009, pp. 2036–2043.
- 18 A. Yilmaz, O. Javed, and M. Shah, "Object tracking: A survey," *ACM Computing Surveys*, vol. 38, no. 4, pp. 13–es, 2006.
- 19 T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *CVPR*, 2017, pp. 2117–2125.
- 20 K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *ICCV*, 2017, pp. 2961–2969.
- 21 Z. Huang, L. Huang, Y. Gong, C. Huang, and X. Wang, "Mask scoring r-cnn," in *CVPR*, 2019, pp. 6409–6418.
- 22 D.-P. Fan, M.-M. Cheng, J.-J. Liu, S.-H. Gao, Q. Hou, and A. Borji, "Salient objects in clutter: Bringing salient object detection to the foreground," in *ECCV*, 2018, pp. 186–202.
- 23 R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," in *CVPR*, 2009, pp. 1597–1604.
- 24 M.-M. Cheng, N. J. Mitra, X. Huang, P. H. Torr, and S.-M. Hu, "Global contrast based salient region detection," *IEEE TPAMI*, vol. 37, no. 3, pp. 569–582, 2014.
- 25 G. Lee, Y.-W. Tai, and J. Kim, "Deep saliency with encoded low level distance map and high level features," in *CVPR*, 2016, pp. 660–668.
- 26 Q. Hou, M.-M. Cheng, X. Hu, A. Borji, Z. Tu, and P. Torr, "Deeply supervised salient object detection with short connections," *IEEE TPAMI*, vol. 41, no. 4, pp. 815–828, 2019.
- 27 P. Zhang, D. Wang, H. Lu, H. Wang, and X. Ruan, "Amulet: Aggregating multi-level convolutional features for salient object detection," in *ICCV*, 2017, pp. 202–211.
- 28 T. Zhao and X. Wu, "Pyramid feature attention network for saliency detection," in *CVPR*, 2019, pp. 3085–3094.
- 29 X. Zhang, T. Wang, J. Qi, H. Lu, and G. Wang, "Progressive attention guided recurrent network for salient object detection," in *CVPR*, 2018, pp. 714–722.
- 30 W. Wang, J. Shen, M.-M. Cheng, and L. Shao, "An iterative and cooperative top-down and bottom-up inference network for salient object detection," in *CVPR*, 2019, pp. 5968–5977.
- 31 A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *NeurIPS*, vol. 30, 2017, pp. 1–11.
- 32 S. Woo, J. Park, J.-Y. Lee, and I. So Kweon, "Cbam: Convolutional block attention module," in *ECCV*, 2018, pp. 3–19.
- 33 N. Liu, J. Han, and M.-H. Yang, "Picanet: Learning pixel-wise contextual attention for saliency detection," in *CVPR*, 2018, pp. 3089–3098.
- 34 S. Chen, X. Tan, B. Wang, and X. Hu, "Reverse attention for salient object detection," in *ECCV*, 2018, pp. 234–250.
- 35 L. Zhu, J. Chen, X. Hu, C.-W. Fu, X. Xu, J. Qin, and P.-A. Heng, "Aggregating attentional dilated features for salient object detection," *IEEE TCSVT*, vol. 30, no. 10, pp. 3358–3371, 2019.
- 36 X. Qin, Z. Zhang, C. Huang, C. Gao, M. Dehghan, and M. Jagersand, "Basnet: Boundary-aware salient object detection," in *CVPR*, 2019, pp. 7479–7489.
- 37 J.-X. Zhao, J.-J. Liu, D.-P. Fan, Y. Cao, J. Yang, and M.-M. Cheng, "Egnet: Edge guidance network for salient object detection," in *ICCV*, 2019, pp. 8779–8788.
- 38 J. Su, J. Li, Y. Zhang, C. Xia, and Y. Tian, "Selectivity or invariance: Boundary-aware salient object detection," in *ICCV*, 2019, pp. 3799–3808.
- 39 Z. Tu, Y. Ma, C. Li, J. Tang, and B. Luo, "Edge-guided non-local fully convolutional network for salient object detection," *IEEE TCSVT*, vol. 31, no. 2, pp. 582–593, 2020.
- 40 X. Hu, L. Zhu, C.-W. Fu, J. Qin, and P.-A. Heng, "Direction-aware spatial context features for shadow detection," in *CVPR*, 2018, pp. 7454–7462.
- 41 H. Le, T. F. Y. Vicente, V. Nguyen, M. Hoai, and D. Samaras, "A+d net: Training a shadow detector with adversarial shadow attenuation," in *ECCV*, 2018, pp. 662–678.
- 42 L. Zhu, Z. Deng, X. Hu, C.-W. Fu, X. Xu, J. Qin, and P.-A. Heng, "Bidirectional feature pyramid network with recurrent

- attention residual modules for shadow detection,” in *ECCV*, 2018, pp. 121–136.
- 43 Q. Zheng, X. Qiao, Y. Cao, and R. W. Lau, “Distraction-aware shadow detection,” in *CVPR*, 2019, pp. 5167–5176.
 - 44 X. Yang, H. Mei, K. Xu, X. Wei, B. Yin, and R. W. Lau, “Where is my mirror?” in *ICCV*, 2019, pp. 8809–8818.
 - 45 H. Mei, B. Dong, W. Dong, P. Peers, X. Yang, Q. Zhang, and X. Wei, “Depth-aware mirror segmentation,” in *CVPR*, 2021, pp. 3044–3053.
 - 46 H. Mei, L. Yu, K. Xu, Y. Wang, X. Yang, X. Wei, and R. W. Lau, “Mirror segmentation via semantic-aware contextual contrasted feature learning,” *ACM TOMM*, 2022.
 - 47 H. Mei, X. Yang, Y. Wang, Y. Liu, S. He, Q. Zhang, X. Wei, and R. W. Lau, “Don’t hit me! glass detection in real-world scenes,” in *CVPR*, 2020, pp. 3687–3696.
 - 48 E. Xie, W. Wang, W. Wang, M. Ding, C. Shen, and P. Luo, “Segmenting transparent objects in the wild,” in *ECCV*, 2020, pp. 696–711.
 - 49 H. Mei, B. Dong, W. Dong, J. Yang, S.-H. Baek, F. Heide, P. Peers, X. Wei, and X. Yang, “Glass segmentation using intensity and spectral polarization cues,” in *CVPR*, 2022.
 - 50 L. Yu, H. Mei, W. Dong, Z. Wei, L. Zhu, Y. Wang, and X. Yang, “Progressive glass segmentation,” *IEEE TIP*, 2022.
 - 51 X. Han, C. Nguyen, S. You, and J. Lu, “Single image water hazard detection using fcnn with reflection attention units,” in *ECCV*, 2018, pp. 105–120.
 - 52 G. H. Thayer and A. H. Thayer, “Concealing-coloration in the animal kingdom : an exposition of the laws of disguise through color and pattern being a summary of abbott h. thayer’s discoveries,” *New York the Macmillan Co*, 1909.
 - 53 H. B. Cott, “Adaptive coloration in animals,” *Methuen & Co. Ltd*, 1940.
 - 54 Y. Lv, J. Zhang, Y. Dai, A. Li, B. Liu, N. Barnes, and D.-P. Fan, “Simultaneously localize, segment and rank the camouflaged objects,” in *CVPR*, 2021, pp. 11 591–11 601.
 - 55 Q. Zhai, X. Li, F. Yang, C. Chen, H. Cheng, and D.-P. Fan, “Mutual graph learning for camouflaged object detection,” in *CVPR*, 2021, pp. 12 997–13 007.
 - 56 A. Li, J. Zhang, Y. Lv, B. Liu, T. Zhang, and Y. Dai, “Uncertainty-aware joint salient object and camouflaged object detection,” in *CVPR*, 2021, pp. 10 071–10 081.
 - 57 Y. Pang, X. Zhao, T.-Z. Xiang, L. Zhang, and H. Lu, “Zoom in and out: A mixed-scale triplet network for camouflaged object detection,” in *CVPR*, 2022.
 - 58 G.-P. Ji, L. Zhu, M. Zhuge, and K. Fu, “Fast camouflaged object detection via edge-based reversible re-calibration network,” *Pattern Recognition*, vol. 123, p. 108414, 2022.
 - 59 M. Zhuge, X. Lu, Y. Guo, Z. Cai, and S. Chen, “Cubenet: X-shape connection for camouflaged object detection,” *PR*, 2022.
 - 60 X. Cheng, H. Xiong, D.-p. Fan, Y. Zhong, M. Harandi, T. Drummond, and Z. Ge, “Implicit motion handling for video camouflaged object detection,” in *CVPR*, 2022.
 - 61 G.-P. Ji, D.-P. Fan, Y.-C. Chou, D. Dai, A. Liniger, and L. Van Gool, “Deep gradient learning for efficient camouflaged object detection,” *Machine Intelligence Research*, 2023.
 - 62 L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, “Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs,” *IEEE TPAMI*, vol. 40, no. 4, pp. 834–848, 2017.
 - 63 H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, “Pyramid scene parsing network,” in *CVPR*, 2017, pp. 2881–2890.
 - 64 H. Mei, Y. Liu, Z. Wei, L. Zhu, Y. Wang, D. Zhou, Q. Zhang, and X. Yang, “Exploring dense context for salient object detection,” *IEEE TCSVT*, 2021.
 - 65 X. Yang, H. Mei, J. Zhang, K. Xu, B. Yin, Q. Zhang, and X. Wei, “Drfn: Deep recurrent fusion network for single-image super-resolution with large factors,” *IEEE TMM*, vol. 21, no. 2, pp. 328–337, 2019.
 - 66 J. Zhang, C. Long, Y. Wang, X. Yang, H. Mei, and B. Yin, “Multi-context and enhanced reconstruction network for single image super resolution,” in *ICME*, 2020, pp. 1–6.
 - 67 J. Zhang, C. Long, Y. Wang, H. Piao, H. Mei, X. Yang, and B. Yin, “A two-stage attentive network for single image super-resolution,” *IEEE TCSVT*, vol. 32, no. 3, pp. 1020–1033, 2021.
 - 68 C. Peng, X. Zhang, G. Yu, G. Luo, and J. Sun, “Large kernel matters—improve semantic segmentation by global convolutional network,” in *CVPR*, 2017, pp. 4353–4361.
 - 69 X. Hu, C.-W. Fu, L. Zhu, T. Wang, and P.-A. Heng, “Sac-net: Spatial attenuation context for salient object detection,” *IEEE TCSVT*, vol. 31, no. 3, pp. 1079–1090, 2020.
 - 70 H. Ding, X. Jiang, B. Shuai, A. Qun Liu, and G. Wang, “Context contrasted feature and gated multi-scale aggregation for scene segmentation,” in *CVPR*, 2018, pp. 2393–2402.
 - 71 K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *CVPR*, 2016, pp. 770–778.

- 72 L. Yang, R.-Y. Zhang, L. Li, and X. Xie, “Simam: A simple, parameter-free attention module for convolutional neural networks,” in *ICML*, 2021, pp. 11 863–11 874.
- 73 H. Xiao, J. Feng, Y. Wei, M. Zhang, and S. Yan, “Deep salient object detection with dense connections and distraction diagnosis,” *IEEE TMM*, vol. 20, no. 12, pp. 3239–3251, 2018.
- 74 Q. Huang, C. Xia, C.-H. Wu, S. Li, Y. Wang, Y. Song, and C.-C. J. Kuo, “Semantic segmentation with reverse attention,” in *BMVC*, 2017, pp. 18.1–18.13.
- 75 Z. Zhu, Q. Wang, B. Li, W. Wu, J. Yan, and W. Hu, “Distractor-aware siamese networks for visual object tracking,” in *ECCV*, 2018, pp. 101–117.
- 76 J. Wei, S. Wang, and Q. Huang, “F³net: fusion, feedback and focus for salient object detection,” in *AAAI*, vol. 34, 2020, pp. 12 321–12 328.
- 77 P. Skurowski, H. Abdulameer, J. Błaszczyk, T. Depta, A. Kornacki, and P. Koziel, “Animal camouflage analysis: Chameleon database,” *Unpublished Manuscript*, vol. 2, p. 7, 2018.
- 78 D.-P. Fan, M.-M. Cheng, Y. Liu, T. Li, and A. Borji, “Structure-measure: A new way to evaluate foreground maps,” in *ICCV*, 2017, pp. 4548–4557.
- 79 D.-P. Fan, G.-P. Ji, X. Qin, and M.-M. Cheng, “Cognitive vision inspired object segmentation metric and loss function,” *SSI*, vol. 6, p. 6, 2021.
- 80 R. Margolin, L. Zelnik-Manor, and A. Tal, “How to evaluate foreground maps?” in *CVPR*, 2014, pp. 248–255.
- 81 A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga *et al.*, “Pytorch: An imperative style, high-performance deep learning library,” in *NeurIPS*, vol. 32, 2019, pp. 1–12.
- 82 W. Liu, A. Rabinovich, and A. C. Berg, “Parsenet: Looking wider to see better,” *arXiv:1506.04579*, 2015.
- 83 P. Krähenbühl and V. Koltun, “Efficient inference in fully connected crfs with gaussian edge potentials,” in *NeurIPS*, vol. 24, 2011, pp. 1–9.
- 84 K. Chen, J. Pang, J. Wang, Y. Xiong, X. Li, S. Sun, W. Feng, Z. Liu, J. Shi, W. Ouyang *et al.*, “Hybrid task cascade for instance segmentation,” in *CVPR*, 2019, pp. 4974–4983.
- 85 Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, “Unet++: A nested u-net architecture for medical image segmentation,” *DLMIA*, pp. 3–11, 2018.
- 86 Z. Wu, L. Su, and Q. Huang, “Cascaded partial decoder for fast and accurate salient object detection,” in *CVPR*, 2019, pp. 3907–3916.
- 87 T. Zhao and X. Wu, “Pyramid feature attention network for saliency detection,” in *CVPR*, 2019, pp. 3085–3094.
- 88 Z. Chen, Q. Xu, R. Cong, and Q. Huang, “Global context-aware progressive aggregation network for salient object detection,” in *AAAI*, vol. 34, 2020, pp. 10 599–10 606.
- 89 Y. Pang, X. Zhao, L. Zhang, and H. Lu, “Multi-scale interactive network for salient object detection,” in *CVPR*, 2020, pp. 9413–9422.
- 90 D. Bo, W. Wenhai, F. Deng-Ping, L. Jinpeng, F. Huazhu, and S. Ling, “Polyp-pvt: Polyp segmentation with pyramid vision transformers,” *arXiv:2108.06932*, 2021.
- 91 D. Jha, P. H. Smedsrud, M. A. Riegler, P. Halvorsen, T. d. Lange, D. Johansen, and H. D. Johansen, “Kvasir-seg: A segmented polyp dataset,” in *MMM*, 2020, pp. 451–462.
- 92 J. Bernal, F. J. Sánchez, G. Fernández-Esparrach, D. Gil, C. Rodríguez, and F. Vilarino, “Wm-dova maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians,” *CMIG*, vol. 43, pp. 99–111, 2015.
- 93 D. Vázquez, J. Bernal, F. J. Sánchez, G. Fernández-Esparrach, A. M. López, A. Romero, M. Drozdal, and A. Courville, “A benchmark for endoluminal scene segmentation of colonoscopy images,” *Journal of Healthcare Engineering*, vol. 2017, 2017.
- 94 O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *MICCAI*, 2015, pp. 234–241.
- 95 Y. Fang, C. Chen, Y. Yuan, and K.-y. Tong, “Selective feature aggregation network with area-boundary constraints for polyp segmentation,” in *MICCAI*, 2019, pp. 302–310.
- 96 C.-H. Huang, H.-Y. Wu, and Y.-L. Lin, “Hardnet-mseg: a simple encoder-decoder polyp segmentation neural network that achieves over 0.9 mean dice and 86 fps,” *arXiv:2101.07172*, 2021.
- 97 Z. Yin, K. Liang, Z. Ma, and J. Guo, “Duplex contextual relation network for polyp segmentation,” *arXiv:2103.06725*, 2021.
- 98 K. Patel, A. M. Bur, and G. Wang, “Enhanced u-net: A feature enhancement network for polyp segmentation,” in *CRV*, 2021, pp. 181–188.
- 99 J. Wei, Y. Hu, R. Zhang, Z. Li, S. K. Zhou, and S. Cui, “Shallow attention network for polyp segmentation,” in *MICCAI*, 2021, pp. 699–708.

Distraction-Aware Camouflaged Object Segmentation

Haiyang MEI¹, Xin YANG^{1*}, Yunduo ZHOU¹, Ge-Peng JI², Xiaopeng WEI¹ & Deng-Ping FAN³

1. *Dalian University of Technology, Dalian 116024, China;*

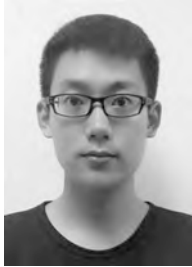
2. *Australian National University, Canberra 2601, Australia;*

3. *Nankai University, Tianjin 300350, China*

* Corresponding author. E-mail: xinyang@dlut.edu.cn

Abstract In this work, our goal is to design an effective and efficient camouflaged object segmentation (COS) model. To this end, we develop a bio-inspired framework, termed pyramid positioning and focus network (PFNet+), which mimics the process of predation in nature. Specifically, our PFNet+ contains three key modules, i.e., a context enrichment (CEn) module, a pyramid positioning module (PPM), and a focus module (FM). The CEn aims at enhancing the representation ability of backbone features via integrating contextual information for providing more discriminative backbone features. The PPM is designed to mimic the detection process in predation for positioning the potential target objects from a global perspective in a pyramid manner and the FM is then used to perform the identification process in predation for progressively refining the initial prediction via focusing on the ambiguous regions. Notably, in the FM, we develop a novel distraction mining strategy for the distraction discovery and removal, to benefit the performance of estimation. Extensive experiments demonstrate that our PFNet+ runs in real-time (56 FPS) and outperforms 20 cutting-edge models on three challenging datasets under four standard metrics. The generalization capability of our PFNet+ is further demonstrated by the experiments on the other vision task (i.e., polyp segmentation).

Keywords Camouflaged object, distraction, context enrichment, context exploration, pyramid, segmentation.



Haiyang Mei received the B.Eng. degree in the Automation from Dalian University of Technology, Dalian China, in 2017. He is currently working toward the Ph.D. degree with the School of Computer Science and Technology, Dalian University of Technology, Dalian, China. His research interests include image processing and computer vision.



Xin Yang is a Professor in the Department of Computer Science at Dalian University of Technology, China. Yang received his B.S. degree in Computer Science from Jilin University in 2007. From 2007 to June 2012, he was a joint Ph.D. student at Zhejiang University and UC Davis for Graphics and received his Ph.D. degree in July 2012. His research interests include computer graphics and robotic vision.



Xiaopeng Wei is a professor at Dalian University of Technology. His research areas include computer animation, computer vision, robots, and intelligent CAD. So far, he has (co-)authored approximately 200 published papers.



Deng-Ping FAN was born in 1988. received his PhD degree from the Nankai University in 2019. He is a Postdoctoral Researcher, working with Prof. Luc Van Gool in Computer Vision Lab @ ETH Zurich. From 2019-2021, he was a research scientist (PI) and team lead of IIAI-CV&Med in IIAI. He received his Ph.D. from Nankai University in 2019 under the supervision of Prof. Ming-Ming Cheng. His research interest is in Computer Vision and Medical Image Analysis. Specifically, he focuses on Dichotomous Image Segmentation (General Object Segmentation, Camouflaged Object Segmentation, Saliency Detection), Multi-Modal AI, etc. He is a member of the TRACE-Zurich project on automated driving. He won the Best Paper Finalist Award at IEEE CVPR 2019, the Best Paper Award Nominee at IEEE CVPR 2020.