基于分心挖掘的伪装物体分割

梅海洋[1] 季葛鹏[2,4] 魏子麒[3,⋆] 杨鑫[1,⋆] 魏小鹏[1] 范登平[4]
[1] 大连理工大学 [2] 武汉大学 [3] 清华大学 [4] 起源人工智能研究院

https://mhaiyang.github.io/CVPR2021_PFNet/index

Abstract

伪装物体分割（*Camouflaged Object Segmentation*, *COS*），旨在识别“完美”嵌入其周围环境的物体，具有广泛的实际应用。*COS* 任务的核心挑战在于目标物体和背景之间存在着高度的相似。在这篇论文中，本文致力于克服实现有效且高效的 *COS* 所面临的挑战。为了达到这个目的，本文开发了一个生物启发的框架，称为定位和聚焦网络（*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* 包含两个关键模块，即定位模块（*Positioning Module*, *PM*）和聚焦模块（*Focus Module*, *FM*）。*PM* 被设计用来模仿捕食中的检测过程，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。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（*72fps*），在四个标准度量下，*PFNet* 在三个具有挑战性的数据集上都显著优于现有的 *18* 个最新模型。

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）[47]。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 [13] 和肺部感染分割 [14]）、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且有价值的应用。

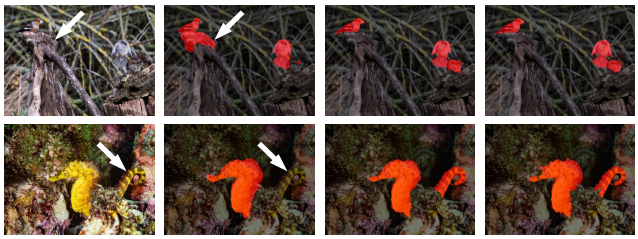


图 1: 伪装物体分割的视觉样例。现有的先进方法 SINet [12] 被与伪装物体具有相似外观的背景区域（第一行中的箭头所指区域）或混杂在背景中的伪装物体区域（第二行中的箭头所指区域）所迷惑，本文的方法可以消除这些干扰，产生准确的分割结果。

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 [47]，因此需要大量的视觉感知知识 [50] 来消除由于目标物体和背景之间的高度内在相似性而引起的歧义。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 [47]。早期的方法致力于根据纹理 [45]、三维凸度 [39] 和运动 [28] 等手工制作的低级特征来区分前景和背景。然而，这些特征对伪装和非伪装物体的区分能力有限，因此基于它们的方法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 [12, 26, 58] 在一定程度上提高了分割性能，但在探索准确伪装物体分割的有效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机制，以降低被发现的风险 [47]。

* 杨鑫（xinyang@dlut.edu.cn）和魏子麒为本文的共同通讯作者。
本文为 CVPR2021 论文 [36] 的中文翻译版。

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 [15]。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。

本文提出了一个定位和聚焦网络（PFNet），极大提高了伪装物体分割的性能。本文的 PFNet 包含两个关键模块，即定位模块（PM）和聚焦模块（FM）。其中，PM 用于模拟捕食中的检测过程，从全局角度对潜在目标进行定位，然后 FM 被用来执行捕食中的识别过程，通过聚焦歧义区域来细化初始的分割结果。具体地说，PM 由一个通道注意力模块和一个空间注意力模块组成，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范围的语义依赖，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，然后去除这些干扰，得到目标物体更纯净的表示。这种分心挖掘策略以隐式方式实现，应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目标的能力（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，称为定位和聚焦网络（PFNet）。该方法首先通过探索长范围的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。

2. 相关工作

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 [31]，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 [23]、场景理解 [29] 和目标跟踪 [61]）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例。因此，直接应用 GOD 方法 [17,22,30] 分割伪装物体可能达不到预期的效果。

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 [9]。早期的方法主要是基于手工制作的低级特征以及启发式先验（例如颜色 [1] 和对比度 [6]）。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 [19,27,63,68]。循环学习和迭代学习策略的也被用来逐步细化检测结果 [52,64]。由于对特征增强的有效性，注意力机制 [51,54] 也被应用于显著性检测 [4,32]。此外，边界线索也被用来细化显著性检测结果 [43,48,67]。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。

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 [20,25,70,72]，镜子 [35,59]，玻璃 [38,57] 和水 [16] 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。

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 [47]，这受益于两项杰出研究 [7, 49] 的巨大影响。早期与伪装相关的工作致力于根据手工制作的纹理 [45]、三维凸度 [39] 和运动 [28] 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 [26] 提出了一种将分类信息集成到像素级分割中的端到端的伪装物体分割网络。Yan 等人 [58] 进一步引入了对抗性攻击来提高分割精度。Fan 等人 [12] 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。

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，[3,37,65] 开发了多尺度的上下文，[60,62] 提取了多层次的上下文，[38,42] 捕获了大视场上下文特征，[20] 探索了方向感知的上下文，[8,59] 利用了对比的上下文。然而，不加区分

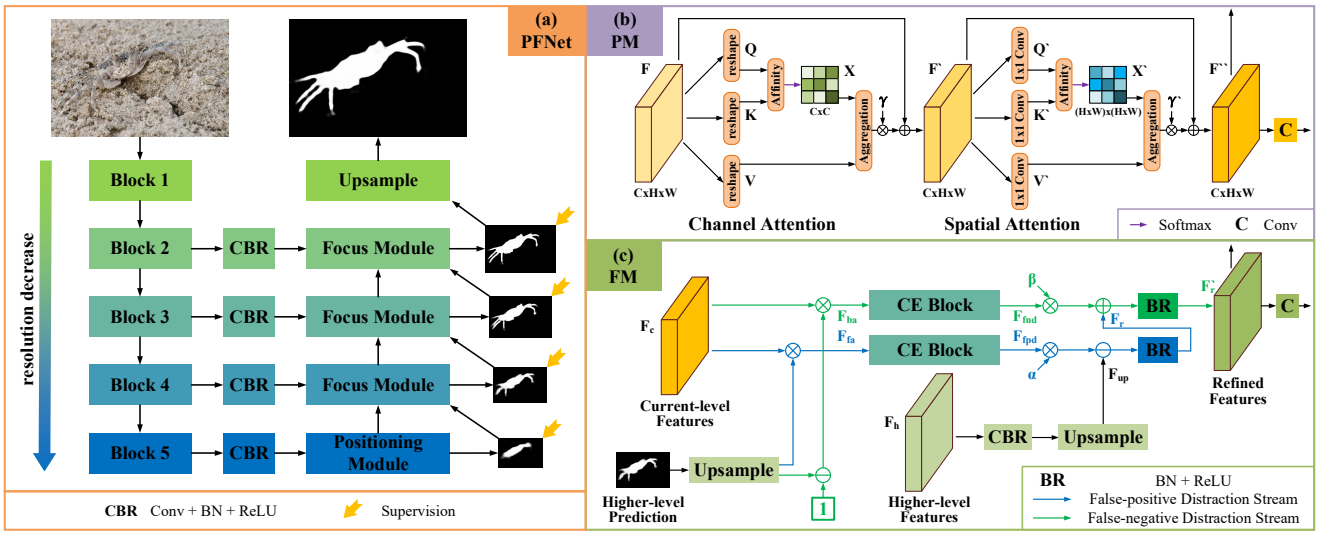


图 2: （a）本文的定位与聚焦网络（PFNet）及其两个主要组成模块：（b）定位模块（PM）和（c）聚焦模块（FM）。

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/背景的注意特征中探索上下文，用于上下文的推理和分心区域的发现。实验验证了本文方法的有效性。

3. 方法

生物学相关研究 [15] 指出，捕食过程可分为三个阶段：检测、识别和捕获。受到捕食的前两个阶段的启发，本文设计了一个由两个关键模块，即定位模块（PM）和聚焦模块（FM）组成的定位和聚焦网络（PFNet）。PM 被设计成模仿捕食中的检测过程，从全局角的度定位潜在的物体，而 FM 则被用于执行捕食中的识别过程，通过聚焦歧义区域来细化初始的分割结果。

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，本文首先将其送入 ResNet-50 [18] 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。然后，在最深层特征上应用定位模块（PM）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. 定位模块

图2 （b）展示了精心设计的定位模块（PM）的详细结构。给定输入的最深层次特征，PM 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。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范围的依赖关系，从全局角度增强最深层次特征的语义表示。

具体来说，给定输入特性 $F \in \mathbb{R}^{C\times H\times W}$，其中 C、H 和 W 分别表示通道数量、特征的高度和宽度，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 $X \in \mathbb{R}^{C\times C}$：

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

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 $F'\in\mathbb{R}^{C\times H\times W}$：

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

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范围语义依赖关系，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 $X'\in\mathbb{R}^{N\times N}$：

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

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 $F''\in\mathbb{R}^{C\times H\times W}$：

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

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。

最后，本文可以通过在 F'' 上应用卷积核为 7×7、填充为 3 的卷积来得到伪装物体的初始位置图。F'' 和初始位置图将由后续的聚焦模块（FMs）逐步细化。

3.3. 聚焦模块

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。

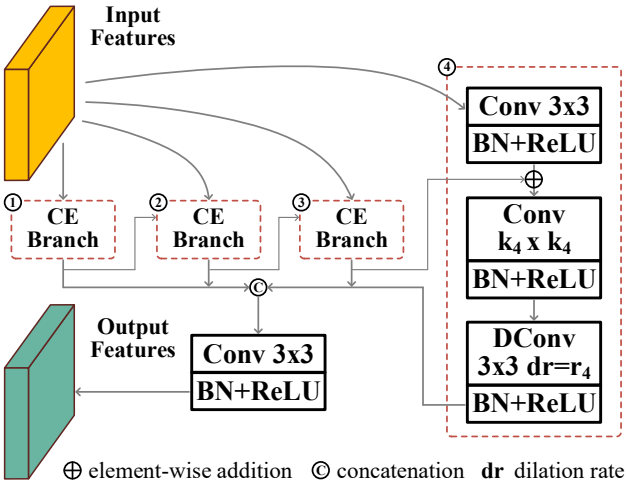


图 3: 本文的上下文探索（CE）模块的结构。

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，即比较歧义区域和自信区域的模式，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，以发现与自信前景（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，本文将这两种类型的特征送入两个并行的上下文探索（CE）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3所示，CE 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，CE 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：

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

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 CE 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 [4, 56]，语义分割 [21] 和视觉跟踪 [73]。现有的工作利用假阳性分心 [21, 56, 73] 或假阴性分心 [4] 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 [70] 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（即 [20, 25, 72]）的预测结果与真值之间的的差异得到的，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，本文相信所提出的分心挖掘策略可以为其他视觉任务提供启发。

3.4. 损失函数

PFNet 中有四个输出结果，一个来自定位模块 PM，三个来自聚焦模块 FM。对于定位模块，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} [44]，即 $\mathcal{L}_{pm} = \ell_{bce} + \ell_{iou}$，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} [53] 和加权 IoU 损失 ℓ_{wiou} [53]，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：

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

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

4. 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 [46]，CAMO [26] 以及 COD10K [12]。CHAMELEON [46] 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 [26] 包含 1,250 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 [12] 是目前最大的基准数据集，它包括从多个摄影网站下载的 5,066 张伪装图片（其中 3,040 张用于训练，2,026 张用于测试），涵盖 5 个大类和 69 个子类。本文仿照之前的工作 [12]，使用 CAMO [26] 和 COD10K [12] 的训练集作为训练集（4,040 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）[10]，自适应 E 度量（E^{ad}_ϕ）[11]，加权 F 度量（F^w_β）[34]，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 [10] 一样被设置为 0.5。E_ϕ 同时评估像素级匹配和图像

Methods	Pub.'Year	CHAMELEON (76 images)				CAMO-Test (250 images)				COD10K-Test (2,026 images)			
		$S_\alpha\uparrow$	$E_\phi^{ad}\uparrow$	$F_\beta^w\uparrow$	$M\downarrow$	$S_\alpha\uparrow$	$E_\phi^{ad}\uparrow$	$F_\beta^w\uparrow$	$M\downarrow$	$S_\alpha\uparrow$	$E_\phi^{ad}\uparrow$	$F_\beta^w\uparrow$	$M\downarrow$
FPN$^\circ$ [30]	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$^\bullet$ [66]	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* [17]	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++§ [71]	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$^\triangle$ [20]	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† [32]	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$^\triangle$ [72]	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* [2]	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* [22]	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† [44]	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† [55]	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† [69]	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† [67]	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† [53]	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† [5]	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§ [13]	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† [40]	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* [12]	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
PFNet*	Ours	**0.882**	**0.942**	**0.810**	**0.033**	**0.782**	**0.852**	**0.695**	**0.085**	**0.800**	**0.868**	**0.660**	**0.040**

表 1: 本文提出的方法与相关的 18 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^w （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，$\star$：实例分割方法，$\triangle$：阴影检测方法，§：医学图像分割方法，†：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。评测代码：https://github.com/DengPingFan/CODToolbox.

级统计信息，其被证明与人类的视觉感知相关 [11]。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 [10,11] 证明加权的 F_β（即 F_β^w [34]）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。这里采用 PyTorch [41] 实现本文的模型。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 内存）。模型训练阶段，输入图像的大小被调整为 416×416，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 [18] 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 16，并通过 poly 策略 [33] 调整学习率，其中基础学习率为 0.001，幂次为 0.9。网络经过 45 轮后收敛，耗时仅需约 76 分钟。测试阶段，本文首先将图像的大小调整为 416×416 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）[24]）来进一步增强最终输出。416×416 图像的推断仅需 0.014 秒（约 72 帧/秒）。

比较方法。为了证明本文的 PFNet 的有效性，本文将其与 18 个最新方法进行了比较：目标检测方法 FPN [30]，语义分割方法 PSPNet [66]，实例分割方法 Mask RCNN [17]、HTC [2] 以及 MSRCNN [22]，阴影检测方法 DSC [20] 以及 BDRAR [72]，医学图像分割方法 UNet++ [71] 以及 PraNet [13]，显著性目标检测方法 PiCANet [32]、BASNet [44]、CPD [55]、PFANet [69]、EGNet [67]、F3Net [53]、GCPANet [5] 以及 MINet-R [40] 和伪装物体分割方法 SINet [12]。为了公平比较，上述方法的所有预测图要么由公共网站提供，要么通过运行用开放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。

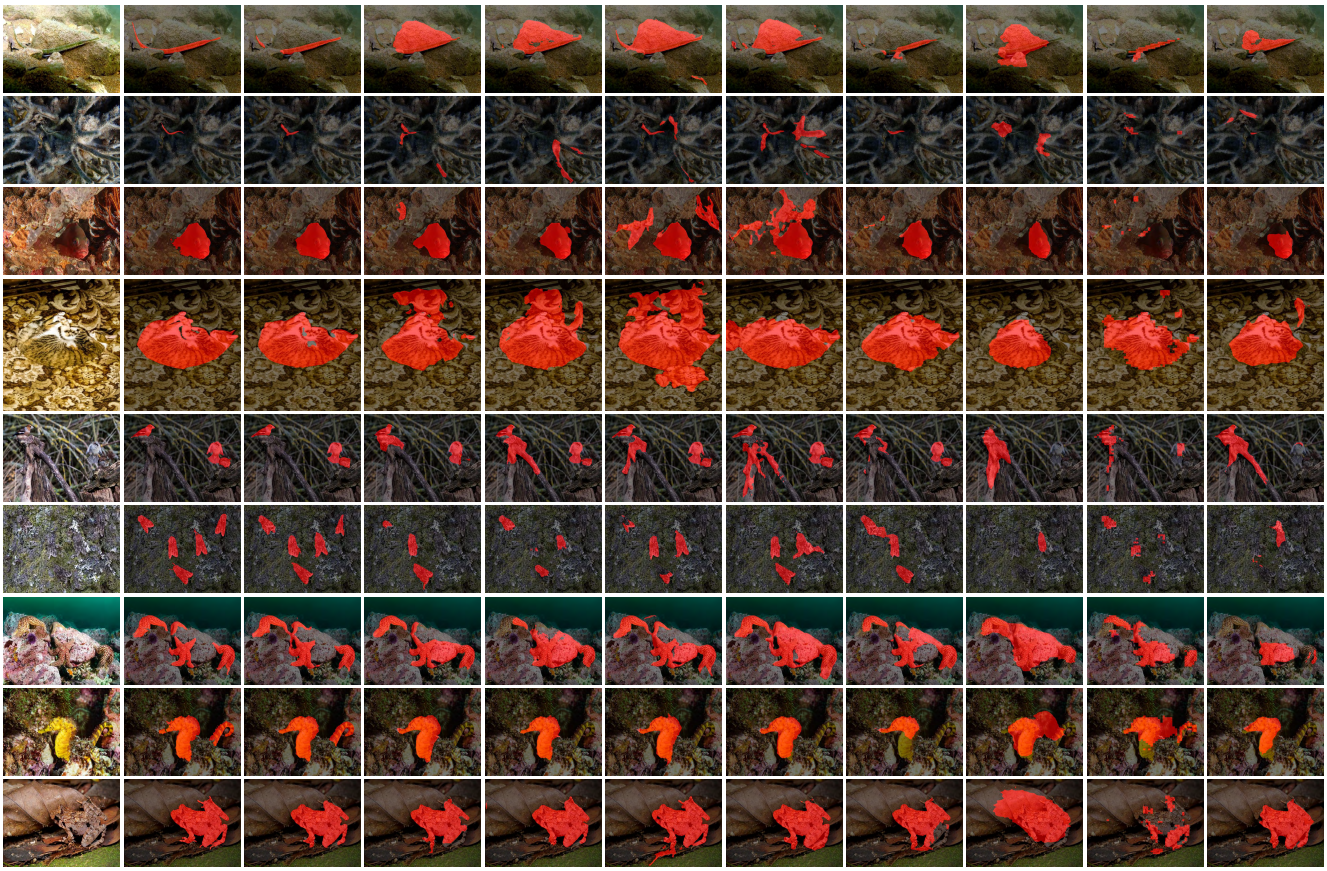


图 4: 本文模型与最新方法的视觉比较。显然，本文的方法能够更准确地分割不同环境下的各种伪装物体。

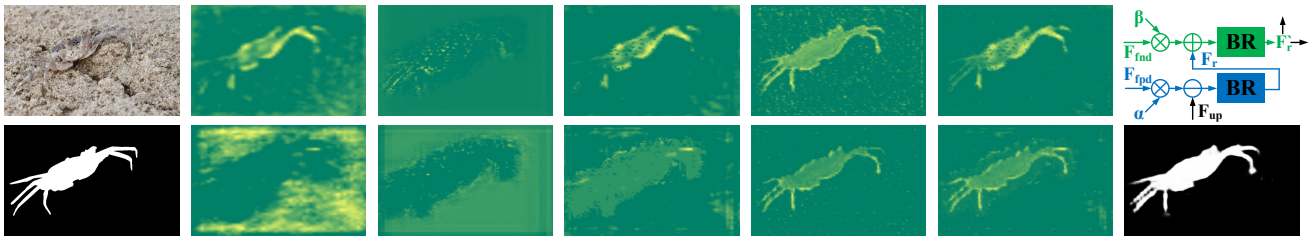
4.2. 和最新方法的比较

表1报告了 PFNet 与其他 18 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 [12] 相比，本文的方法将 F_β^w 在 CHAMELEON [46]、CAMO [26]、以及 COD10K [12] 数据集上分别提升了 7.0%、8.9%、以及 10.9%。值得注意的是，本文的方法也比 SINet 更快（即 72 帧/秒对比 51 帧/秒）。

此外，图4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）。这主要是因为定位模块能够通过挖掘长范围的语义依赖关系，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，受益于分心发现过程中的多尺度上下文探索，本文的方法能够获取到精细的分心信息，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



(a) Image & GT (b) F_{up} (c) F_{fpd} (d) F_r (e) F_{fnd} (f) $F_r^`$ (g) Prediction

图 5: 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。

Networks		COD10K-Test (2,026 images)			
		$S_\alpha\uparrow$	$E_\phi^{ad}\uparrow$	$F_\beta^w\uparrow$	$M\downarrow$
(a)	B	0.779	0.803	0.591	0.051
(b)	B + CA	0.788	0.819	0.618	0.046
(c)	B + SA	0.791	0.826	0.624	0.046
(d)	B + PM	0.792	0.835	0.631	0.045
(e)	B + FPD	0.790	0.844	0.632	0.043
(f)	B + FND	0.790	0.837	0.628	0.043
(g)	B + FM *w/o* A	0.796	0.843	0.639	0.042
(h)	B + FM	0.797	0.860	0.649	0.041
(i)	B + PM + FPD	0.796	0.854	0.645	0.042
(j)	B + PM + FND	0.796	0.847	0.644	0.043
(k)	B + PM + FM *w/o* A	0.796	0.851	0.647	0.042
(l)	PFNet	**0.800**	**0.868**	**0.660**	**0.040**

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扮演着重要的角色，并对分割性能做出贡献。

聚焦模块（FM））的有效性，并将结果报告在表2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。

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^w 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性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的特征图进行了可视化（参见图5）。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。

5. 结论

在这项工作中，本文致力于迎接挑战，以实现准确的伪装物体分割。本文开发了一种新颖的分心挖掘策略来发现和去除分心。通过在本文的生物启发的框架（**P**ositioning and **F**ocus **Net**work，PFNet）中采用分心挖掘策略，展示出本文的方法在三个基准数据集上实现了最优的性能。在未来，本文计划探索该方法在其他领域应用（例如：息肉分割和 COVID-19 肺部感染检测）的潜力，并继续改善模型的性能，使之能够分割视频中的伪装物体。

参考文献

[1] Radhakrishna Achanta, Sheila Hemami, Francisco Estrada, and Sabine Susstrunk. Frequency-tuned salient region detection. In *CVPR*, 2009.

[2] Kai Chen, Jiangmiao Pang, Jiaqi Wang, Yu Xiong, Xiaoxiao Li, Shuyang Sun, Wansen Feng, Ziwei Liu, Jianping Shi, Wanli Ouyang, et al. Hybrid task cascade for instance segmentation. In *CVPR*, 2019.

[3] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE TPAMI*, 2017.

[4] Shuhan Chen, Xiuli Tan, Ben Wang, and Xuelong Hu. Reverse attention for salient object detection. In *ECCV*, 2018.

[5] Zuyao Chen, Qianqian Xu, Runmin Cong, and Qingming Huang. Global context-aware progressive aggregation network for salient object detection. In *AAAI*, 2020.

[6] Ming-Ming Cheng, Niloy J Mitra, Xiaolei Huang, Philip HS Torr, and Shi-Min Hu. Global contrast based salient region detection. *IEEE TPAMI*, 2014.

[7] Hugh Bamford Cott. Adaptive coloration in animals. *Methuen & Co. Ltd*, 1940.

[8] Henghui Ding, Xudong Jiang, Bing Shuai, Ai Qun Liu, and Gang Wang. Context contrasted feature and gated multi-scale aggregation for scene segmentation. In *CVPR*, 2018.

[9] Deng-Ping Fan, Ming-Ming Cheng, Jiang-Jiang Liu, Shang-Hua Gao, Qibin Hou, and Ali Borji. Salient objects in clutter: Bringing salient object detection to the foreground. In *ECCV*, 2018.

[10] Deng-Ping Fan, Ming-Ming Cheng, Yun Liu, Tao Li, and Ali Borji. Structure-measure: A new way to evaluate foreground maps. In *ICCV*, 2017.

[11] Deng-Ping Fan, Ge-Peng Ji, Xuebin Qin, and Ming-Ming Cheng. Cognitive vision inspired object segmentation metric and loss function. *SSI*, 2021.

[12] Deng-Ping Fan, Ge-Peng Ji, Guolei Sun, Ming-Ming Cheng, Jianbing Shen, and Ling Shao. Camouflaged object detection. In *CVPR*, 2020.

[13] Deng-Ping Fan, Ge-Peng Ji, Tao Zhou, Geng Chen, Huazhu Fu, Jianbing Shen, and Ling Shao. Pranet: Parallel reverse attention network for polyp segmentation. In *MICCAI*, 2020.

[14] Deng-Ping Fan, Tao Zhou, Ge-Peng Ji, Yi Zhou, Geng Chen, Huazhu Fu, Jianbing Shen, and Ling Shao. Inf-net: Automatic covid-19 lung infection segmentation from ct images. *IEEE TMI*, 2020.

[15] Joanna R Hall, Innes C Cuthill, Roland J Baddeley, Adam J Shohet, and Nicholas E Scott-Samuel. Camouflage, detection and identification of moving targets. *Proceedings of The Royal Society B: Biological Sciences*, 2013.

[16] Xiaofeng Han, Chuong Nguyen, Shaodi You, and Jianfeng Lu. Single image water hazard detection using fcn with reflection attention units. In *ECCV*, 2018.

[17] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *ICCV*, 2017.

[18] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, 2016.

[19] Qibin Hou, Ming-Ming Cheng, Xiaowei Hu, Ali Borji, Zhuowen Tu, and Philip Torr. Deeply supervised salient object detection with short connections. *IEEE TPAMI*, 2019.

[20] Xiaowei Hu, Lei Zhu, Chi-Wing Fu, Jing Qin, and Pheng-Ann Heng. Direction-aware spatial context features for shadow detection. In *CVPR*, 2018.

[21] Qin Huang, Chunyang Xia, Chi-Hao Wu, Siyang Li, Ye Wang, Yuhang Song, and C.-C. Jay Kuo. Semantic segmentation with reverse attention. In *BMVC*, 2017.

[22] Zhaojin Huang, Lichao Huang, Yongchao Gong, Chang Huang, and Xinggang Wang. Mask scoring r-cnn. In *CVPR*, 2019.

[23] Alexander Kirillov, Kaiming He, Ross Girshick, Carsten Rother, and Piotr Dollar. Panoptic segmentation. In *CVPR*, 2019.

[24] Philipp Krähenbühl and Vladlen Koltun. Efficient inference in fully connected crfs with gaussian edge potentials. In *NeurIPS*, 2011.

[25] Hieu Le, Tomas F. Yago Vicente, Vu Nguyen, Minh Hoai, and Dimitris Samaras. A+d net: Training a shadow detector with adversarial shadow attenuation. In *ECCV*, 2018.

[26] Trung-Nghia Le, Tam V Nguyen, Zhongliang Nie, Minh-Triet Tran, and Akihiro Sugimoto. Anabranch network for camouflaged object segmentation. *CVIU*, 2019.

[27] Gayoung Lee, Yu-Wing Tai, and Junmo Kim. Deep saliency with encoded low level distance map and high level features. In *CVPR*, 2016.

[28] Jianqin Yin Yanbin Han Wendi Hou Jinping Li. Detection of the mobile object with camouflage color under dynamic background based on optical flow. *Procedia Engineering*, 2011.

[29] Li-Jia Li, Richard Socher, and Li Fei-Fei. Towards total scene understanding: Classification, annotation and segmentation in an automatic framework. In *CVPR*, 2009.

[30] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection. In *CVPR*, 2017.

[31] Li Liu, Wanli Ouyang, Xiaogang Wang, Paul Fieguth, Jie Chen, Xinwang Liu, and Matti Pietikinen. Deep

learning for generic object detection: A survey. *IJCV*, 2018.

[32] Nian Liu, Junwei Han, and Ming-Hsuan Yang. Picanet: Learning pixel-wise contextual attention for saliency detection. In *CVPR*, 2018.

[33] Wei Liu, Andrew Rabinovich, and Alexander C. Berg. Parsenet: Looking wider to see better. *arXiv:1506.04579*, 2015.

[34] Ran Margolin, Lihi Zelnik-Manor, and Ayellet Tal. How to evaluate foreground maps? In *CVPR*, 2014.

[35] Haiyang Mei, Bo Dong, Wen Dong, Pieter Peers, Xin Yang, Qiang Zhang, and Xiaopeng Wei. Depth-aware mirror segmentation. In *CVPR*, 2021.

[36] Haiyang Mei, Gepeng Ji, Ziqi Wei, Xin Yang, Xiaopeng Wei, and Dengping Fan. Camouflaged object segmentation with distraction mining. In *CVPR*, 2021.

[37] Haiyang Mei, Yuanyuan Liu, Ziqi Wei, Li Zhu, Yuxin Wang, Dongsheng Zhou, Qiang Zhang, and Xin Yang. Exploring dense context for salient object detection. *IEEE TCSVT*, 2021.

[38] Haiyang Mei, Xin Yang, Yang Wang, Yuanyuan Liu, Shengfeng He, Qiang Zhang, Xiaopeng Wei, and Rynson W.H. Lau. Don't hit me! glass detection in real-world scenes. In *CVPR*, 2020.

[39] Yuxin Pan, Yiwang Chen, Qiang Fu, Ping Zhang, and Xin Xu. Study on the camouflaged target detection method based on 3d convexity. *Modern Applied Science*, 2011.

[40] Youwei Pang, Xiaoqi Zhao, Lihe Zhang, and Huchuan Lu. Multi-scale interactive network for salient object detection. In *CVPR*, 2020.

[41] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. In *NeurIPS*, 2019.

[42] Chao Peng, Xiangyu Zhang, Gang Yu, Guiming Luo, and Jian Sun. Large kernel matters —improve semantic segmentation by global convolutional network. In *CVPR*, 2017.

[43] Xuebin Qin, Zichen Zhang, Chenyang Huang, Chao Gao, Masood Dehghan, and Martin Jagersand. Basnet: Boundary-aware salient object detection. In *CVPR*, 2019.

[44] Xuebin Qin, Zichen Zhang, Chenyang Huang, Chao Gao, Masood Dehghan, and Martin Jagersand. Basnet: Boundary-aware salient object detection. In *CVPR*, 2019.

[45] P. Sengottuvelan, A. Wahi, and A. Shanmugam. Performance of decamouflaging through exploratory image analysis. In *ETET*, 2008.

[46] P Skurowski, H Abdulameer, J Błaszczyk, T Depta, A Kornacki, and P Kozieł. Animal camouflage analysis: Chameleon database. *Unpublished Manuscript*, 2018.

[47] Martin Stevens and Sami Merilaita. Animal camouflage: current issues and new perspectives. *Philosophical Transactions of the Royal Society B*, 2009.

[48] Jinming Su, Jia Li, Yu Zhang, Changqun Xia, and Yonghong Tian. Selectivity or invariance: Boundary-aware salient object detection. In *ICCV*, 2019.

[49] Gerald Handerson Thayer and Abbott Handerson 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.

[50] Tom Troscianko, Christopher P Benton, P. George Lovell, David J Tolhurst, and Zygmunt Pizlo. Camouflage and visual perception. *Philosophical Transactions of the Royal Society B*, 2009.

[51] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *NeurIPS*, 2017.

[52] Wenguan Wang, Jianbing Shen, Ming-Ming Cheng, and Ling Shao. An iterative and cooperative top-down and bottom-up inference network for salient object detection. In *CVPR*, 2019.

[53] Jun Wei, Shuhui Wang, and Qingming Huang. F3net: Fusion, feedback and focus for salient object detection. In *AAAI*, 2020.

[54] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. Cbam: Convolutional block attention module. In *ECCV*, 2018.

[55] Zhe Wu, Li Su, and Qingming Huang. Cascaded partial decoder for fast and accurate salient object detection. In *CVPR*, 2019.

[56] Huaxin Xiao, Jiashi Feng, Yunchao Wei, Maojun Zhang, and Shuicheng Yan. Deep salient object detection with dense connections and distraction diagnosis. *IEEE TMM*, 2018.

[57] Enze Xie, Wenjia Wang, Wenhai Wang, Mingyu Ding, Chunhua Shen, and Ping Luo. Segmenting transparent objects in the wild. In *ECCV*, 2020.

[58] Jinnan Yan, Trung-Nghia Le, Khanh-Duy Nguyen, Minh-Triet Tran, Thanh-Toan Do, and Tam V. Nguyen. Mirrornet: Bio-inspired adversarial attack for camouflaged object segmentation. *arXiv:2007.12881*, 2020.

[59] Xin Yang, Haiyang Mei, Ke Xu, Xiaopeng Wei, Baocai Yin, and Rynson W.H. Lau. Where is my mirror? In *ICCV*, 2019.

[60] Xin Yang, Haiyang Mei, Jiqing Zhang, Ke Xu, Baocai Yin, Qiang Zhang, and Xiaopeng Wei. Drfn: Deep recurrent fusion network for single-image super-resolution with large factors. *IEEE TMM*, 2019.

[61] Alper Yilmaz, Omar Javed, and Mubarak Shah. Object tracking: A survey. *ACM Computing Surveys*, 2006.

[62] Jiqing Zhang, Chengjiang Long, Yuxin Wang, Xin Yang, Haiyang Mei, and Baocai Yin. Multi-context and enhanced reconstruction network for single image super resolution. In *ICME*, 2020.

[63] Pingping Zhang, Dong Wang, Huchuan Lu, Hongyu Wang, and Xiang Ruan. Amulet: Aggregating multi-level convolutional features for salient object detection. In *ICCV*, 2017.

[64] Xiaoning Zhang, Tiantian Wang, Jinqing Qi, Huchuan Lu, and Gang Wang. Progressive attention guided recurrent network for salient object detection. In *CVPR*, 2018.

[65] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. Pyramid scene parsing network. In *CVPR*, 2017.

[66] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. Pyramid scene parsing network. In *CVPR*, 2017.

[67] Jia-Xing Zhao, Jiang-Jiang Liu, Deng-Ping Fan, Yang Cao, Jufeng Yang, and Ming-Ming Cheng. Egnet: Edge guidance network for salient object detection. In *ICCV*, 2019.

[68] Ting Zhao and Xiangqian Wu. Pyramid feature attention network for saliency detection. In *CVPR*, 2019.

[69] Ting Zhao and Xiangqian Wu. Pyramid feature attention network for saliency detection. In *CVPR*, 2019.

[70] Quanlong Zheng, Xiaotian Qiao, Ying Cao, and Rynson WH Lau. Distraction-aware shadow detection. In *CVPR*, 2019.

[71] Zongwei Zhou, Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: A nested u-net architecture for medical image segmentation. *DLMIA*, 2018.

[72] Lei Zhu, Zijun Deng, Xiaowei Hu, Chi-Wing Fu, Xuemiao Xu, Jing Qin, and Pheng-Ann Heng. Bidirectional feature pyramid network with recurrent attention residual modules for shadow detection. In *ECCV*, 2018.

[73] Zheng Zhu, Qiang Wang, Bo Li, Wei Wu, Junjie Yan, and Weiming Hu. Distractor-aware siamese networks for visual object tracking. In *ECCV*, 2018.