

# 基于自适应分割与背景检索的伪装目标检测

Bowen Yin<sup>1</sup>, Xuying Zhang<sup>1</sup>, Li Liu<sup>2\*</sup>, Ming-Ming Cheng<sup>1</sup>, Yongxiang Liu<sup>2</sup>, and Qibin Hou<sup>1\*</sup>

<sup>1</sup> VCIP, CS, Nankai University, Tianjin, China

<sup>2</sup> Academy of Advanced Technology Research of Hunan, Changsha, China

**摘要** 最近的研究证实了局部细节在伪装目标检测中的重要性。然而，如何通过背景线索检测目标物体周围的细节仍缺乏深入研究。在本文中，我们考虑到这一点，提出了一种用于伪装目标检测的新型学习框架，称为 AdaptCOD。具体而言，我们的方法将检测过程解耦为三个部分，即定位、分割和检索。我们设计了一种上下文自适应分割策略，以动态选择合理的上下文区域进行局部分割，并设计了一个背景检索模块来进一步优化伪装目标的边界。尽管方法简单，但我们的方法甚至能使一个简单的 COD 模型实现出色的性能。大量实验表明，AdaptCOD 在三个广泛使用的伪装目标检测基准测试中超越了所有现有的最先进方法。代码公布在：<https://github.com/HVision-NKU/AdaptCOD>。

**Keywords:** 伪装目标检测 · 自适应分割 · 背景检索。

## 1 引言

伪装目标检测 (COD) 旨在从视觉相似的环境中识别伪装目标，是计算机视觉领域的一项新兴研究。它还推动了不同领域中广泛且有价值的实际应用，例如生物学研究中的物种发现 [14]、工业制造中的表面缺陷检测 [54, 28]，以及医学诊断中的息肉分割 [12]。在伪装目标检测 (COD) 场景中 [4, 25]，目标物体通常隐藏在复杂杂乱的环境中，这使得模型难以准确地对其进行定位和分割。此外，诸多因素（如背景极其相似、目标尺寸多样、目标边缘不规则等）也加剧了该任务的难度。图 1 展示了一些伪装目标的示例。



图 1. 图中展示了我们的 AdaptCOD 与两种最新的伪装目标检测先进方法（如 SegMaR [24] 和 DTINet [35]）的视觉对比。可以看到，尽管近期的先进方法能够定位目标，但在区分极其相似的遮挡区域和背景环境时仍存在困难。

近年来，主流的伪装目标检测 (COD) 方法主要基于 Transformer [35, 30, 57]、卷积神经网络 [24, 45, 76]，或两者结合。值得注意的是，大多数方法侧重于设计新颖的网络架构以提升分割性能，而伪装目标周围上下文信息的作用却鲜少被探究。作为一个例外，SegMaR [24] 指出，无需通过观察整幅图像来检测伪装目标。该方法提出了一种多阶段模型，该模型利用迭代扩展优化策略，仅通过观察关注区域来逐步优化伪装目标。尽管 SegMaR 性能良好，但其仅专注于研究如何迭代优化目标周围的紧密区域，却忽略了分析如何合理利用上下文信息和背景来改善分割结果。

基于上述分析，本文首先探讨为了更好地检测伪装目标，上下文区域应多大为宜。如图 2 所示，我们从伪装目标检测数据集中随机选取一幅图像，逐步增加其分辨率，并以目标中心为基准将其裁剪为固定尺

\* Corresponding authors.



zoom ratio	×1	×2	×3	×5	×10	×Max
$S_m$	0.845	0.873	0.888	0.918	0.891	0.879
$wF_m$	0.741	0.779	0.798	0.851	0.812	0.795

图 2. 伪装目标检测模型在随机选取的不同缩放比例图像上的性能表现。该模型在标准训练集上完成充分训练，且参数固定。×5 表示我们使用 5 倍缩放比例增大图像尺寸，然后以目标中心为基准将结果裁剪为固定尺寸。×Max 表示最大缩放比例。

寸。我们发现，随着伪装目标尺寸的增大，分割性能起初会提升，但随后会下降。这一现象表明，适量的背景上下文信息对于检测目标物体至关重要，而如何判断背景区域的选择会对伪装目标检测 (COD) 模型的性能产生显著影响。

受这些观察结果的启发，我们提出了一种新颖的“定位 - 分割 - 优化”的学习框架 (称为 AdaptCOD) 用于伪装目标检测，如图 3 所示。首先，将整幅图像输入一个简单网络，以生成伪装目标的粗预测图。基于该粗预测图，我们设计了一种上下文自适应分割策略，将整个图像划分为候选前景区域和剩余背景区域。然后，将候选前景区域再次输入该简单网络，以捕捉伪装目标周围的更多细节。为进一步优化分割结果，我们提出一种背景检索模块，利用背景区域作为查询来纠正错误预测。

为验证 AdaptCOD 的有效性，我们在三个主流伪装目标检测 (COD) 基准数据集 (NC4K [38]、COD10K [10] 和 CAMO [27]) 上进行了广泛的消融实验。值得注意的是，与近期前沿方法 (如 EINet [30] 和 TPRNet [72]) 相比，我们的方法在所有基准数据集中均创下了新的先进 (SOTA) 记录。具体而言，在 COD10K 测试数据集上，我们的方法取得了 0.819 的加权 F 值 [39] 和 0.892 的 S 值 [8]，而次优模型 TransCoop [59] 的对应结果分别为 0.779 和 0.863。此外，可视化结果也表明，AdaptCOD 相比现有 COD 方法具有显著优势。

本文的主要贡献总结如下：

- 揭示了伪装目标检测 (COD) 方法利用适当的上下文信息以生成高质量分割结果的重要性。
- 提出了一种新颖的“定位 - 分割 - 优化” (locate-segment-polish) 框架 (名为 AdaptCOD) 用于伪装目标检测研究，该框架包含上下文自适应分割策略和背景检索模块。
- 证明了 AdaptCOD 可显著提升现有 COD 方法的性能。此外，借助该框架，即使是简单网络也能实现优异的性能。

## 2 相关工作

### 2.1 伪装目标检测

伪装目标检测是一项具有挑战性的任务，因为目标与背景之间存在细微差异。为解决这一问题，研究人员先后提出了一系列工作 [44, 23, 76, 45, 71, 24, 41, 11]，以提高分割结果的准确性。具体而言，这些方法根据所采用的策略可分为三类：1) 多尺度特征聚合：ZoomNet [45] 以三种尺度处理输入图像，并统一不同尺度下的特定尺度外观特征。PreyNet [71] 根据捕食过程将伪装目标的检测过程分为初始检测和捕食者学习。2) 多阶段优化策略：在 SINetV2 [10] 中，通过设计邻域连接解码器和组逆注意力机制，进一步提升伪装目标的分割精度。3) 多任务训练：SLSR [38] 通过引入伪装排序或从显著目标到伪装目标的迁移学习，提出了一种联合学习框架。

近年来，源自自然语言处理 (NLP) 领域的 Transformer 模型 [56, 7] 在一系列视觉任务中取得了显著成功，包括图像分类 [51, 15, 64, 5]、目标检测 [66]、语义分割 [60, 31]、超分辨率重建 [77] 以及多模态学习 [73, 61]。这些成果也启发了诸多研究将该架构应用于伪装目标检测 (COD) 任务。UGTR [65] 利

用 Transformer 框架下的概率表征模型来学习伪装目标的不确定性。DTINet [35] 提出了一种双任务交互 Transformer，可同时分割伪装目标及其细节边界。EINet [30] 设计了一个包含邻域连接和跳跃连接的解码器，以逐步优化伪装目标。这些方法大多聚焦于新型网络架构的设计。尽管性能良好，但它们大多忽略了伪装目标周围上下文信息的重要性。

## 2.2 扩展策略

为获取伪装目标的细粒度前景预测，需要感知目标周围足够的细节信息 [26, 53, 78, 3]。近年来，一类具有代表性的研究通过从全图中以更细尺度采样若干子区域来实现这一目标。例如，TASN [75] 设计了一种基于注意力的采样器，在保持图像原始尺寸的同时增加关注区域的比例。随后，PA-KRN [63] 利用该采样方法，基于身体注意力图以更高分辨率突出前景目标区域。最近，SegMaR [24] 根据注意力图对目标周围区域进行迭代扩展，以实现伪装目标（尤其是小目标）的精准分割。这些工作为目标的精确定位和分割带来了积极提升。然而，它们仅依赖注意力图扩展目标，忽略了上下文和背景信息的利用。与这些方法不同，我们的框架根据感知到的全局上下文自适应划分图像区域，能够以更低的计算量实现更优性能。

## 2.3 判别性内容检索

分割具有精细细节的伪装目标的核心在于在混淆场景中检索出存在差异的位置 [74, 32, 49, 37, 70]。大量研究 [17, 13, 43, 69] 致力于挖掘更具判别性的上下文信息，以提升其方法的检测能力。例如，PFNet [41] 提出了一种干扰挖掘策略，通过分别处理前景和背景特征来消除错误预测。DDS [34] 提出了多样化的深度监督机制以增强边缘感知能力。BGNet [53] 和 BSA Net [78] 则利用边界特征对伪装目标进行更精细的分割。OPNet [42] 提出了一种双焦点模块，通过融合局部和全局表征实现伪装目标的精确定位。这些方法隐含地利用了全图的判别性上下文信息，但在捕捉伪装目标的细节方面仍存在困难。相比之下，本文通过将背景表征作为判别线索，显式优化目标周围的背景区域，能够实现更优的分割结果。

## 3 模型框架

在本节中，我们首先简要介绍所提出的 AdaptCOD 的总体架构，然后依次阐述其核心组成部分，即全局定位、局部分割和背景检索。最后，我们给出本文使用的训练损失函数。

### 3.1 总览

我们提出的 AdaptCOD 整体架构如图 3 所示。整个流程可分为三个部分：全局定位、局部分割和背景检索。第一部分定位旨在通过全局上下文建模粗略定位伪装目标。输入图像被送入共享网络以获得粗预测，如图 3 左上角所示，基于粗预测，我们使用自适应分割策略生成分割掩码，将图像分为前景区域和背景区域。在分割部分，我们将前景区域划分为细粒度补丁，并利用共享网络提取具有精细细节的特征以进行精确分割。检索部分将分割掩码和前两部分的特征图作为输入。我们利用提出的背景检索模块，通过度量局部特征与背景特征的距离来优化局部特征，然后使用优化后的特征生成最终预测。

### 3.2 全局定位

在全局定位部分，我们将空间尺寸为  $H \times W$  的输入图像送入基于编码器 - 解码器架构的共享网络。所采用的解码器结构基于特征金字塔网络 (FPN) [33]，与主流伪装目标检测方法 [10, 67] 类似，但仅由基础卷积层组成。具体而言，编码器通过四个阶段提取多尺度视觉特征。我们使用  $1 \times 1$  卷积将这些特征的通道数压缩至  $C = 128$ ，以在效率和性能之间实现更好的权衡。生成的特征  $\{E_i\}_{i=1}^4$  的空间尺寸为  $\frac{H}{2^{i+1}} \times \frac{W}{2^{i+1}}$ ，然后被送入解码器生成分割图。特别地，最后三个阶段的特征被输入进入一个卷积块，在其中进行融合并通过两个卷积层处理，生成包含更全面图像理解的特征  $D_1^{loc}$ 。该过程可定义如下：

$$D_1^{loc} = \mathcal{F}_{\text{conv}5 \times 5}(\mathcal{F}_{\text{conv}3 \times 3}([\mathcal{F}_{\text{resize}}(E_2), \mathcal{F}_{\text{resize}}(E_3), E_4])), \quad (1)$$

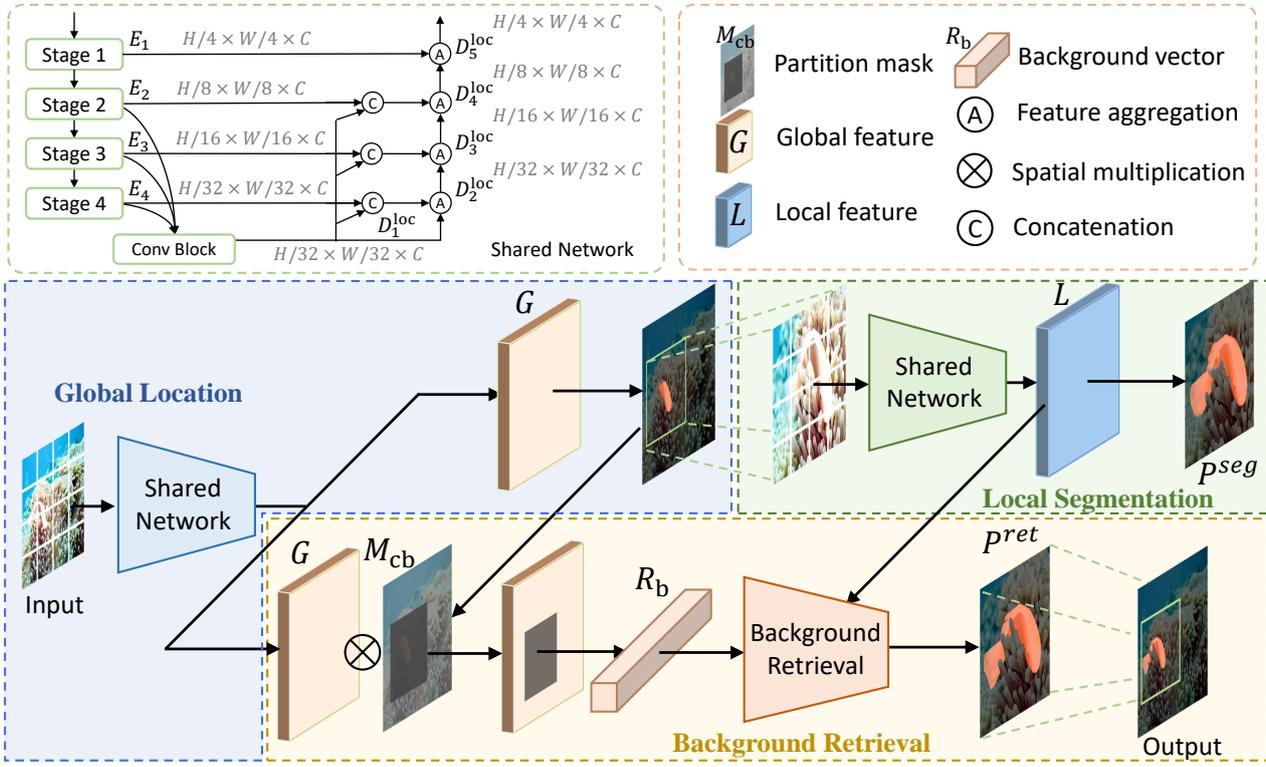


图3. AdaptCOD 的整体架构。首先，将整幅图像输入共享网络进行粗预测以定位目标。共享网络采用预训练的 PVTV2-b4 [58] 作为主干网络，与 HitNet [19] 和 EINet [30] 一致。然后，自适应地将图像划分为背景和候选前景区域，并将后者输入共享网络以获取更多局部细节。最后，利用背景检索策略进一步增强模型的判别能力。特征聚合过程如公式 3 所示。

其中， $\mathcal{F}_{\text{resize}}$  表示与  $E_4$  进行分辨率匹配的双线性插值函数， $\mathcal{F}_{\text{conv}k \times k}$  表示核大小为  $k \times k$  的卷积操作。融合特征  $D_1^{\text{loc}} \in \mathbb{R}^{H/32 \times W/32 \times C}$  用于按如下方式增强特征  $\{E_i\}_{i=2}^4$ ：

$$E_i^{\text{loc}} = \mathcal{F}_{\text{conv}3 \times 3}([\mathcal{F}_{\text{resize}}(D_1^{\text{loc}}), E_i]), i \in \{2, 3, 4\}, \quad (2)$$

其中， $E_i^{\text{loc}}$  为增强后的视觉特征，且  $E_1^{\text{loc}} = E_1$ 。遵循特征金字塔网络 (FPN) [33]，我们也采用自顶向下的方式对多阶段视觉特征进行解码，如图 3 所示。自顶向下结构中的特征聚合过程公式化如下：

$$D_j^{\text{loc}} = \mathcal{F}_{\text{conv}3 \times 3}(\mathcal{F}_{\text{resize}}(D_{j-1}^{\text{loc}}) \cdot E_{6-j}^{\text{loc}} + E_{6-j}^{\text{loc}}), j \in [2, 5]. \quad (3)$$

解码器特征  $D_i^{\text{loc}}$  可通过  $3 \times 3$  卷积操作后接双线性插值操作转换为预测结果  $P_i^{\text{loc}}$ 。

考虑到最后一个解码器特  $D_5^{\text{loc}}$  包含整幅图像丰富的全局上下文信息，我们将其作为全局特征，记为  $G$ 。同时，对应的预测结果  $P_5^{\text{loc}}$  也会在局部分割部分中使用。

### 3.3 局部分割

如第 1 节所示，如何调整伪装目标周围的上下文区域对模型性能有很大影响。在此，我们提出一种自适应分割策略，可自动为局部分割选择合理的上下文区域。

给定粗预测结果  $P_5^{\text{loc}} \in [0, 1]^{H \times W}$ ，我们可以轻松获得包含预测目标的边界框。我们将边界框的高度和宽度分别记为  $h$  和  $w$ 。我们可以通过如下公式得到高度为  $h_e$  和宽度为  $w_e$  的新矩形区域：

$$\begin{cases} h_e &= h \cdot (1 + \gamma \times r_t), \\ w_e &= w \cdot (1 + \gamma \times r_t), \\ r_t &= \sum P_5^{\text{loc}} / (h \times w), \end{cases} \quad (4)$$

其中， $\gamma$  是控制上下文自适应扩展幅度的固定缩放参数， $r_t$  表示预测前景区域与初始边界框的面积比。该策略可解释如下：给定一个固定尺寸的边界框，当目标物体的比例  $r_t$  越大时，背景上下文的区域往往越小。在这种情况下，应使用边界框外的更多背景区域来进行精确分割。

为了提供更易理解的解释，一些扩展示例如图 4 所示。第一行中，初始边界框中的背景区域较小，因此我们的策略以相对较大的比例对其进行扩展。第二行的情况则正好相反。因此，我们的上下文自适应分割策略可以动态地将包含目标物体的边界框扩展到合适的大小。

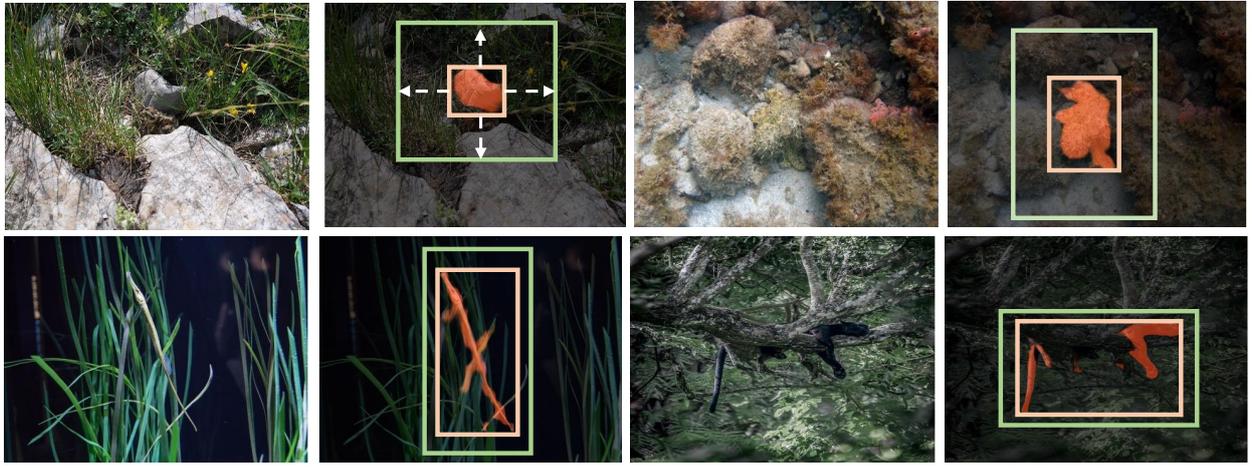


图4. 上下文自适应扩展策略示意图。内部矩形表示预测目标的边界框，外部矩形则是通过上下文自适应扩展策略从内部矩形扩展而来。

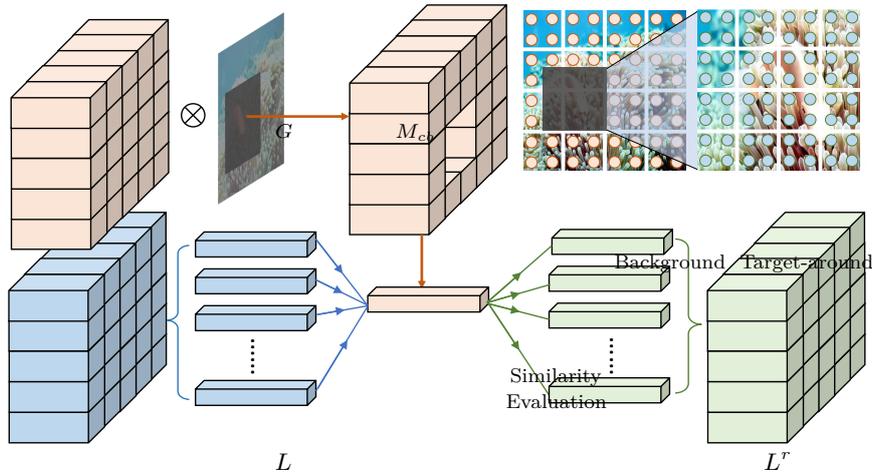


图5. 背景检索模块示意图。我们首先使用分割掩码从全局特征中排除前景周围的线索，以获取纯背景信息。然后，通过度量局部特征与背景表示之间的相似性，在局部特征中检索背景点。

基于扩展后的边界框，我们将整个图像划分为与边界框内区域对应的候选前景  $R_{cf}$  和边界框外的背景区域  $R_{cb}$ 。背景掩码  $M_{cb} \in \{0, 1\}^{H \times W}$  可通过如下方式轻松获得：

$$M_{cb}(i, j) = \begin{cases} 0, & (i, j) \in R_{cf} \\ 1, & (i, j) \in R_{cb}, \end{cases} \quad (5)$$

其将用于背景检索，下一小节将对此进行说明。

对于剩余的前景区域，我们从原始图像中裁剪出该区域作为局部图像，并将其尺寸调整为空间大小  $H \times W$ 。然后，我们将局部图像输入共享网络，以生成具有精细细节的分割结果  $\{P_i^{seg}\}_{i=1}^5$ 。遵循与全局定位相同的原则，我们选择最后阶段的解码器特征（记为  $D_5^{seg}$ ）作为局部特征  $L$ ，其形状为  $H/4 \times W/4 \times C$ 。

### 3.4 背景检索

由于伪装物体与背景上下文之间的细微差异，即使使用了局部分割部分，清晰检测伪装物体的边界仍然具有挑战性。我们观察到，一些背景区域仍被错误地预测为前景，这会降低模型性能。为了进一步优化物体边界周围的区域，我们提出通过背景检索模块将伪装目标与其邻近的背景环境区分开来。

其基本思想如图5所示。在该模块中，我们首先使用全局特征  $G \in \mathbb{R}^{H/4 \times W/4 \times C}$  和背景掩码  $M_{cb} \in \{0, 1\}^{H \times W}$ ，通过掩码平均池化（MAP）函数生成索引向量  $R^b \in \mathbb{R}^{1 \times C}$ 。该过程的公式化表达如下：

$$R^b = \frac{\sum (\mathcal{F}_{\text{resize}}(M_{cb}) \odot G)}{\sum \mathcal{F}_{\text{resize}}(M_{cb})}, \quad (6)$$

其中， $\odot$  表示逐元素乘法运算， $\sum(\cdot)$  表示沿空间维度累加特征值。

给定索引向量  $R^b$ ，我们使用它从局部特征中检索背景区域。这可以通过测量  $R^b$  与局部特征  $L \in \mathbb{R}^{H/4 \times W/4 \times C}$  中每个位置的余弦相似性距离来简单实现。然后，通过以下方式可获得结果特征  $L^r \in \mathbb{R}^{H/4 \times W/4 \times C}$ ：

$$L^r(i, j) = L(i, j) \left(1 - \frac{R^b L^T(i, j)}{\|R^b\| \|L(i, j)\|}\right) / 2, \quad (7)$$

其中， $\|\cdot\|$  表示  $L_2$  范数函数。最后，我们从  $L^r$  中通过一个  $3 \times 3$  卷积层（后接双线性插值函数）获得与候选前景  $R_{cf}$  空间形状相同的精细预测  $P^{ret}$ ，并将其作为最终预测粘贴回其原始位置，如图 3 所示。

### 3.5 训练目标

整个训练过程持续 75 个 epoch。具体来说，我们首先在标准 COD 训练集上训练 AdaptCOD 模型以完成定位任务。遵循文献 [18, 62] 的方法，我们在每个特征层级添加了侧监督。给定定位部分生成的预测结果  $P^{loc} = \{P_i^{loc}\}_{i=1}^5$  及其对应的真实标签  $T$ ，前 60 个训练周期的损失函数可定义为：

$$\mathcal{L}(P^{loc}, T) = \sum_{i=1}^5 \mathcal{L}_{bce}(P_i^{loc}, T) + \mathcal{L}_{iou}(P_i^{loc}, T), \quad (8)$$

其中， $\mathcal{L}_{bce}$  和  $\mathcal{L}_{iou}$  分别表示二元交叉熵损失 [6] 和交并比损失 [40]。

之后，我们对整个框架进行联合训练。给定来自三个部分的预测图  $\{P_i^{loc}\}_{i=1}^5$ 、 $\{P_i^{seg}\}_{i=1}^5$  和  $P^{ret}$ ，最后 15 个训练周期的损失函数可表示为：

$$\mathcal{L} = \mathcal{L}(P^{loc}, T) + \mathcal{L}(P^{seg}, T_l) + \mathcal{L}(P^{ret}, T_l), \quad (9)$$

其中， $T_l$  表示扩展边界框内的真实标签。

## 4 实验

### 4.1 实验设置

**数据集：**我们在伪装目标检测（COD）研究中最常用的三个基准数据集上评估所提出的 AdaptCOD 方法，即 CAMO [27]、COD10K [10] 和 NC4k [38]。CAMO 数据集包含 2500 个样本，其中 1250 张图像包含伪装目标，剩余 1250 张图像不含伪装目标。COD10K 数据集由 5066 张包含伪装目标的图像、3000 张背景图像和 1934 张非伪装图像组成。NC4K 是一个包含 4121 张图像的大规模数据集，所有图像均包含伪装目标。遵循先前研究 [11, 24, 45]，我们使用 CAMO 的 1000 张图像和 COD10K 的 3040 张图像进行训练，其余图像用于测试。

**评估指标：**遵循先前研究 [45, 24, 76] 中的协议，我们采用四种广泛使用的指标对预测结果进行评估，包括结构度量 ( $S_m$ ) [8]、平均绝对误差 (M) [47]、加权 F 值 ( $wF$ ) [39] 和自适应 E 度量 ( $\alpha E$ ) [9]。具体来说，平均绝对误差表示预测图与真实标签之间的绝对差异。结构度量评估预测和真实标签之间区域感知和目标感知的结构相似性。加权 F 值提供了召回率和精确率的全面衡量。自适应 E 度量评估元素级相似性和图像级统计量。此外，我们还绘制了精确率 - 召回率 (PR) 曲线和  $F_\beta$  - 阈值 ( $F_\beta$ ) 曲线，以进一步评估伪装目标检测模型的性能。

**超参数细节：**遵循 HitNet [19, 30]，我们采用在 ImageNet 数据集上预训练的 PVTv2 - b4 [58] 作为主干网络，并将自适应扩展策略中的  $\gamma$  设置为 0.4。在训练阶段，我们将批量大小设置为 16，使用随机梯度下降 (SGD) 优化器，并将初始学习率设置为  $1e-2$ ，学习率通过余弦退火方法 [36] 逐渐降低。在推理阶段，首先将完整图像调整大小为  $384 \times 384$ ，并输入到 AdaptCOD 的共享网络中进行伪装目标定位。然后，在分割部分，从原始图像中裁剪出上下文合理区域并调整大小为  $384 \times 384$ ，将该局部图像再次输入到共享网络以生成局部特征。最后，通过背景检索消除误预测来优化这些特征。本文所有实验均使用 PyTorch 库 [46] 在 NVIDIA GeForce RTX 3090 GPU 上进行。

**计算成本与速度：**如前所述，训练周期总数设置为 75 个，其中 60 个用于初始定位，15 个用于进一步优化。整个过程耗时约 5.8 小时，包括 3.3 小时的训练和 2.5 小时的微调。我们的 AdaptCOD 模型包含 6640 万参数和 67.5G 次乘加运算 (Macps)，推理速度为 30.4 FPS。图 10 展示了我们的方法与其他方法在计算成本和速度上的对比。



图 6. 使用 AdaptCOD 中不同组件的预测结果可视化。可视化样本随机选自 NC4K 和 COD10K 数据集。

Setting	Parts		NC4K (4,121)				COD10K-Test (2,026)				CAMO-Test (250)			
	LS	BR	$S_m \uparrow$	$\alpha E \uparrow$	$wF \uparrow$	$M \downarrow$	$S_m \uparrow$	$\alpha E \uparrow$	$wF \uparrow$	$M \downarrow$	$S_m \uparrow$	$\alpha E \uparrow$	$wF \uparrow$	$M \downarrow$
1			0.870	0.920	0.810	0.038	0.843	0.909	0.787	0.029	0.839	0.901	0.791	0.059
2	✓		0.885	0.930	0.839	0.035	0.860	0.924	0.799	0.025	0.866	0.926	0.827	0.048
3		✓	0.892	0.933	0.852	0.032	0.887	0.929	0.815	0.024	0.877	0.927	0.840	0.046
4	✓	✓	<b>0.906</b>	<b>0.942</b>	<b>0.860</b>	<b>0.029</b>	<b>0.892</b>	<b>0.938</b>	<b>0.819</b>	<b>0.021</b>	<b>0.886</b>	<b>0.932</b>	<b>0.844</b>	<b>0.043</b>

表 1. 对所提出的三部分框架的消融研究表明，所有部分均对整体性能有贡献。其中，LS 表示局部分割部分，BR 表示背景检索部分。

Settings	Expansion Manner	NC4K (4,121)				COD10K-Test (2,026)				CAMO-Test (250)			
		$S_m \uparrow$	$\alpha E \uparrow$	$wF \uparrow$	$M \downarrow$	$S_m \uparrow$	$\alpha E \uparrow$	$wF \uparrow$	$M \downarrow$	$S_m \uparrow$	$\alpha E \uparrow$	$wF \uparrow$	$M \downarrow$
1	Fixed $\alpha = 0$	0.883	0.924	0.825	0.037	0.851	0.897	0.734	0.028	0.862	0.906	0.812	0.052
2	Fixed $\alpha = 1/4$	0.890	0.933	0.845	0.031	0.882	0.930	0.811	0.026	0.871	0.916	0.830	0.048
3	Fixed $\alpha = 1/2$	0.889	0.927	0.828	0.034	0.863	0.896	0.733	0.027	0.870	0.909	0.823	0.051
4	Fixed $\alpha = 1$	0.878	0.925	0.829	0.035	0.850	0.891	0.730	0.029	0.866	0.901	0.816	0.055
5	Context Adaptive	<b>0.906</b>	<b>0.942</b>	<b>0.860</b>	<b>0.029</b>	<b>0.892</b>	<b>0.938</b>	<b>0.819</b>	<b>0.021</b>	<b>0.886</b>	<b>0.932</b>	<b>0.844</b>	<b>0.043</b>

表 2. 目标边界框扩展方式的消融研究。Fixed: 固定边界框的扩展率 Adaptive: 通过我们提出的上下文自适应扩展策略扩展边界框

## 4.2 消融实验

**各组件的重要性:** 我们首先分析三个组件的影响。如表 1 所示，实验结果表明，分割和检索部分均有助于提升 AdaptCOD 的性能。具体而言，配备背景检索策略后，分割阶段在所有基准数据集上的  $S_m$  和  $wF$  指标平均分别提升了 2.1% 和 3.2%。此外，三个组件的结合可进一步提升性能，这表明局部分割和背景检索的重要性。相应的可视化结果如图 6 所示。可以看到，仅使用定位组件的预测结果能够粗略定位目标物体；使用分割组件可进一步改善分割结果，但无法处理伪装目标边界附近的部分细节；在检索组件中，潜在的背景像素通过视觉特征得以区分，从而生成更准确的预测结果。

**上下文自适应扩展的有效性:** 为验证所提出扩展策略的有效性，我们对 AdaptCOD 采用不同扩展率进行微调以作对比。给定固定扩展率  $\alpha$ ，边界框尺寸可扩展为  $(h + 2\alpha h, w + 2\alpha w)$ 。不同扩展策略在所有基准数据集上的实验结果如表 2 所示。值得注意的是，我们的上下文自适应划分策略取得了最佳性能，而固定扩展率  $\alpha = 1/4$  的策略次之。

这些结果表明，过多或过少的背景上下文均会干扰分割预测的准确性，同时也验证了我们边界框扩展策略的合理性。

**扩展策略的进一步分析:** 为进一步验证上下文自适应扩展策略在伪装目标检测中的有效性，我们将自适应扩展率与不同样本上的最佳扩展率进行了对比。具体而言，我们在 COD10K 测试集的每张图像上等距采样的  $\alpha = 0, 0.01, 0.02, \dots, 0.99, 1.00$  进行测试，并选择使  $wF$  值最高的  $\alpha$  作为该图像的最佳扩展率。图 7(a) 展示了样本的自适应扩展率与理想扩展率的平滑线图。可以看出，最佳扩展率曲线与自适应扩展率曲线呈正相关。因此，若进行适当缩放，我们的上下文自适应扩展策略甚至可成为理想的扩展策略。此外，该曲线还解释了为何表 2 中固定  $\alpha = 1/4$  的策略能取得次优性能——因为  $1/4$  最接近最佳扩展率曲线。

图 7(b) 分别以手工扩展率和自适应扩展率为纵轴和横轴，采用最小二乘法拟合两者关系，并将垂直误差小于 0.05 的点标注在两条对角线之间。显然，超过 95% 的点位于两线之间，表明自适应扩展率与最优扩展率之间存在显著正相关。这些结果进一步证明了我们方法的有效性。

**与 SegMaR 对比:** 在此，我们将自适应划分采样器与 SegMaR [24] 中使用的基于注意力的采样器进行对比。首先，基于注意力的采样器采用各向异性采样，可能导致图像失真。这类采样器最初为图像分类任务设计，该任务中失真不影响类别判断，但在伪装目标检测这一密集预测任务中，采样图像的预测结果需还原以匹配原始图像，可能导致空洞结果。此外，这种采样方法还会在原始图像与采样图像之间产生明显的

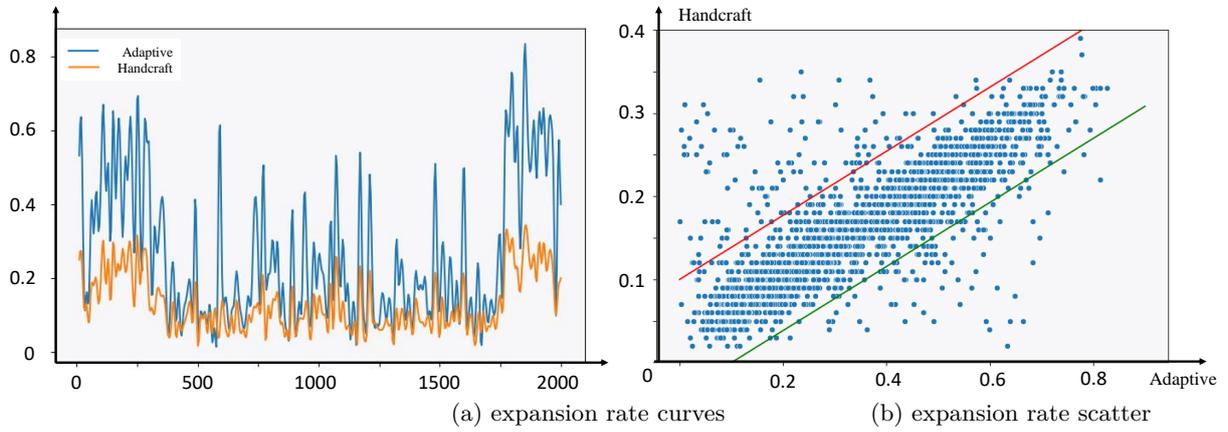


图7. 手动扩展率与自适应扩展率的曲线及散点图。(a) COD10K 测试集上理想值与自适应值的扩展率曲线对比。其中, Handcraft: 特定图像的最佳扩展率; Adaptive: 我们的上下文自适应扩展率  $r_t$ 。(b) 以手动扩展率为横轴、自适应扩展率为纵轴的散点图。

Sample	Computation			COD10K (2,026)			NC4K (4,121)		
	Param (M)	Macs (G)	FPS	$S_m \uparrow$	$wF \uparrow$	$M \downarrow$	$S_m \uparrow$	$wF \uparrow$	$M \downarrow$
attention-based sampler	132.8	67.5	19.8	0.886	0.808	0.024	0.897	0.851	<b>0.029</b>
<b>Adaptive partition (Ours)</b>	66.4	67.5	30.4	<b>0.892</b>	<b>0.819</b>	<b>0.021</b>	<b>0.906</b>	<b>0.860</b>	<b>0.029</b>

表3. 自适应划分策略与基于注意力采样器的对比。

领域间隙, 因此需保留各部分的模型。而我们的采样器仅需一个模型, 更适用于 COD 任务。如表3所示, 我们的自适应划分策略使用一半参数且速度更快, 同时能实现更好的性能。

**不同场景下背景检索的讨论:** 我们将 COD10K 测试集划分为四个子集: SL (单个大目标)、SS (单个小目标)、M (多个目标) 和 MS (多个小目标), 分别包含 579、1,259、188、119 个样本。表4表明, 我们的背景检索过程在所有四种场景下均一致提升性能, 但检索模块在不同场景下的有效性存在差异。对于单个小目标场景, 检索模块带来的提升最大, 而对于多个小伪装目标场景, 提升相对较低。

**超参数  $\gamma$  的选择:** 在公式4中,  $\gamma$  是一个固定缩放参数, 用于控制上下文自适应扩展的幅度。对于大多数场景中的伪装目标, 通过经验我们将  $\gamma$  设置为  $\sqrt{2} - 1 \approx 0.4$ , 使扩展后的边界框面积接近粗略预测目标边界框面积的  $2 \times$  倍。我们进一步通过实验验证该超参数选择的合理性。此处, 我们研究调整超参数  $\gamma$  时模型性能的变化情况。我们在 0.4 附近选取 7 个不同值, 即  $\{0, 0.2, 0.3, 0.4, 0.5, 0.6, 0.8\}$ , 实验结果如表5所示。当  $\gamma = 0.4$  或  $0.5$  时, 模型性能最佳, 而增大或减小  $\gamma$  均会带来负面影响。这些结果表明, 适当大小的背景上下文对模型有益, 建议用户将  $\gamma$  设置为 0.4。

**失败样例:** 我们的 AdaptCOD 旨在通过背景检索策略优化粗略的分割结果。如图8所示, 对于初始预测存在错误区域或遗漏部分目标的场景, 我们的方法也可能失效。然而, 正如表6所示, 我们的方法相比近期最先进方法仍有显著提升, 因此仍具有重要意义。

### 4.3 与其他方法的对比

**与最先进方法的对比:** 我们将 AdaptCOD 与近期 22 个先进方法进行对比, 包括 OPNet [42], FEDER [16], FSPNet [20], ZoomNet [45], FDNet [76], SegMaR [24], DGNet [22], SINetV2 [10], C<sup>2</sup>FNet [52], TPRNet [72], UJSC [29], PFNet [41], MGL-R [68], SLSR [38], SINet [11], PraNet [12], COS-T [57], HitNet [19], EINet [30], DTINet [35], TransCoop [59] and UGTR [65]。为确保对比公平, 预测图直接由作者提供或通过其训练好的模型生成, 并使用相同代码进行评估。如表6所示, 我们的 AdaptCOD 在无需任何额外训练数据或后处理技巧的情况下, 所有指标均优于其他方法。特别是在  $S_m$  和  $wF$  指标上, 相比次优方法 TransCoop [59] 平均提升 2.4%。与近期涉及增强技术的方法 SegMaR [24] 和 ZoomNet [45] 相比, 我们的方法在所有数据集上也展现出显著的性能提升。

Methods	Single Large (579)			Single Small (1,259)			Multi Objects (188)			Multi Small (119)		
	$S_m \uparrow$	$wF \uparrow$	$M \downarrow$	$S_m \uparrow$	$wF \uparrow$	$M \downarrow$	$S_m \uparrow$	$wF \uparrow$	$M \downarrow$	$S_m \uparrow$	$wF \uparrow$	$M \downarrow$
AdaptCOD w/o BR	0.909	0.891	0.019	0.847	0.749	0.026	0.816	0.723	0.031	0.790	0.653	0.033
AdaptCOD	<b>0.923</b>	<b>0.901</b>	<b>0.015</b>	<b>0.888</b>	<b>0.798</b>	<b>0.017</b>	<b>0.827</b>	<b>0.731</b>	<b>0.026</b>	<b>0.796</b>	<b>0.661</b>	<b>0.030</b>

表 4. 背景检索模块在不同场景下的有效性。BR 表示背景检索模块。

Settings	Selection of $\gamma$	NC4K (4,121)				COD10K-Test (2,026)				CAMO-Test (250)			
		$S_m \uparrow$	$\alpha E \uparrow$	$wF \uparrow$	$M \downarrow$	$S_m \uparrow$	$\alpha E \uparrow$	$wF \uparrow$	$M \downarrow$	$S_m \uparrow$	$\alpha E \uparrow$	$wF \uparrow$	$M \downarrow$
1	$\gamma = 0$	0.883	0.924	0.825	0.037	0.851	0.897	0.734	0.028	0.862	0.906	0.812	0.052
2	$\gamma = 0.2$	0.892	0.934	0.833	0.034	0.872	0.917	0.799	0.028	0.873	0.916	0.825	0.048
3	$\gamma = 0.3$	0.899	0.937	0.846	0.033	0.880	0.925	0.807	0.024	0.874	0.920	0.831	0.046
4	$\gamma = 0.4$	<b>0.906</b>	<b>0.942</b>	<b>0.860</b>	<b>0.029</b>	0.892	0.938	<b>0.819</b>	<b>0.021</b>	<b>0.886</b>	<b>0.932</b>	<b>0.844</b>	<b>0.043</b>
5	$\gamma = 0.5$	0.904	0.941	0.857	0.030	<b>0.893</b>	<b>0.939</b>	0.818	0.022	0.883	0.930	<b>0.844</b>	<b>0.043</b>
6	$\gamma = 0.6$	0.900	0.938	0.850	0.031	0.884	0.930	0.810	0.022	0.870	0.919	0.829	0.045
7	$\gamma = 0.8$	0.895	0.933	0.844	0.033	0.878	0.922	0.805	0.025	0.868	0.913	0.821	0.047

表 5. AdaptCOD 使用不同的超参数  $\gamma$ 。

**PR 和  $F_\beta$  曲线:** 我们还提供了所提出的 AdaptCOD 与先前前沿方法的 PR 曲线和  $F_\beta$  曲线。如图 9 所示, 属于我们 AdaptCOD 的红色实线大幅高于其他方法的曲线, 这进一步证明了 AdaptCOD 中自适应划分策略和背景检索模块的有效性。

**性能与计算效率的权衡:** 基于最常用的 COD 基准数据集 COD10K, 我们展示了 AdaptCOD 与其他方法在性能和效率之间的关系。如图 10 所示, 在相近的计算成本下, 我们的 AdaptCOD 显著超越了先前的先进方法。

**定性对比:** 图 11 展示了我们的 AdaptCOD 与近期先进方法在挑战性场景下的可视化对比。所选样本可分为两部分: 第一部分包含小伪装目标 (如第 6、10、12 行), 这类目标难以检测; 第二部分中伪装目标的图案与周围环境极其相似 (如第 1、2 行), 极易造成混淆。从图中可见, 其他先进方法的分割结果存在明显的漏检、误判和模糊问题, 而我们的 AdaptCOD 借助上下文自适应划分策略和背景检索模块实现了精准分割。此外, 与其他方法相比, 我们的预测结果目标区域更清晰、完整。这些可视化结果直观体现了所提方法相较于其他方法的优越性能。

#### 4.4 在其他模型上的应用

考虑到所提出的 AdaptCOD 是一种适用于伪装目标检测研究的通用方案, 我们还尝试将该方案应用于其他 COD 模型。我们唯一需要做的是使用训练良好的参数对所选模型进行微调。在实验中, 我们将框架中的共享模型替换为近期的先进模型 ZoomNet [45] 和 HitNet [19]。如表 7 所示, AdaptCOD 的引入为基础 ZoomNet 和 HitNet 模型在所有基准测试中带来了显著且一致的性能提升。这些结果验证了我们的工作对其他替代 COD 方法的强泛化能力。

我们的 AdaptCOD 方法还适用于医学影像中的息肉分割任务。在结肠镜诊断中, 由于息肉与其高度相似的周围环境之间的边界对比度低, 结直肠癌的早期检测具有挑战性。如表 8 所示, 将 AdaptCOD 应用于息肉分割方法后, 在所有数据集上均产生了更好的性能。这些实验结果进一步表明了 AdaptCOD 的泛化能力。

## 5 结论和未来展望

**结论:** 在本文中, 我们提出了一种用于伪装目标检测的新型学习方案, 称为 AdaptCOD, 该方案将预测过程分为三个部分, 即定位、分割和检索。我们设计了一种上下文自适应划分策略, 以动态选择目标物体周

Method	NC4K (4,121)				COD10K-Test (2,026)				CAMO-Test (250)			
	$S_m \uparrow$	$\alpha E \uparrow$	$wF \uparrow$	$M \downarrow$	$S_m \uparrow$	$\alpha E \uparrow$	$wF \uparrow$	$M \downarrow$	$S_m \uparrow$	$\alpha E \uparrow$	$wF \uparrow$	$M \downarrow$
<b>PraNet</b> <sub>2020</sub> [12]	0.822	0.871	0.724	0.059	0.789	0.839	0.629	0.045	0.769	0.833	0.663	0.094
<b>SINet</b> <sub>2020</sub> [11]	0.808	0.883	0.723	0.058	0.776	0.867	0.631	0.043	0.745	0.825	0.644	0.092
<b>SLSR</b> <sub>2021</sub> [38]	0.840	0.902	0.766	0.048	0.804	0.882	0.673	0.037	0.787	0.855	0.696	0.080
<b>MGL-R</b> <sub>2021</sub> [68]	0.833	0.893	0.739	0.053	0.814	0.865	0.666	0.035	0.782	0.847	0.695	0.085
<b>PFNet</b> <sub>2021</sub> [41]	0.829	0.892	0.745	0.053	0.800	0.868	0.660	0.040	0.782	0.852	0.695	0.085
<b>UJSC</b> <sub>2021</sub> [29]	0.842	0.907	0.771	0.047	0.809	0.891	0.684	0.035	0.800	0.853	0.728	0.073
<b>C<sup>2</sup>FNet</b> <sub>2021</sub> [52]	0.838	0.898	0.762	0.049	0.813	0.886	0.686	0.036	0.796	0.864	0.719	0.080
<b>COS-T</b> <sub>2021</sub> [57]	0.825	0.881	0.730	0.055	0.790	0.901	0.693	0.035	0.813	0.896	0.776	0.060
<b>UGTR</b> <sub>2021</sub> [65]	0.839	0.886	0.746	0.052	0.817	0.850	0.666	0.036	0.784	0.859	0.794	0.086
<b>SINetV2</b> <sub>2022</sub> [10]	0.847	0.898	0.770	0.048	0.815	0.863	0.680	0.037	0.820	0.875	0.743	0.070
<b>DGNet</b> <sub>2022</sub> [22]	0.857	0.907	0.784	0.042	0.822	0.877	0.693	0.033	0.839	0.901	0.769	0.057
<b>SegMaR</b> <sub>2022</sub> [24]	0.841	0.905	0.781	0.046	0.833	0.895	0.724	0.033	0.815	0.872	0.742	0.071
<b>ZoomNet</b> <sub>2022</sub> [45]	0.853	0.907	0.784	0.043	0.838	0.893	0.729	0.029	0.820	0.883	0.752	0.066
<b>FDNet</b> <sub>2022</sub> [76]	0.834	0.895	0.750	0.052	0.837	0.897	0.731	0.030	0.844	0.903	0.778	0.062
<b>TPRNet</b> <sub>2022</sub> [72]	0.854	0.903	0.790	0.047	0.829	0.892	0.725	0.034	0.814	0.870	0.781	0.076
<b>DTINet</b> <sub>2022</sub> [35]	0.863	0.915	0.792	0.041	0.824	0.893	0.695	0.034	0.857	0.912	0.796	0.050
<b>EINet</b> <sub>2022</sub> [30]	0.880	0.931	0.826	0.035	0.849	0.918	0.746	0.027	0.870	0.925	0.822	0.048
<b>HitNet</b> <sub>2022</sub> [19]	0.870	0.921	0.825	0.039	0.868	0.932	0.798	0.024	0.844	0.902	0.801	0.057
<b>TransCoop</b> <sub>2022</sub> [59]	0.883	0.927	0.837	0.033	0.863	0.919	0.779	0.024	0.870	0.923	0.832	0.047
<b>FEDER</b> <sub>2023</sub> [16]	0.859	0.915	0.832	0.041	0.837	0.913	0.756	0.028	0.827	0.893	0.801	0.064
<b>FSPNet</b> <sub>2023</sub> [20]	0.879	0.915	0.816	0.035	0.851	0.895	0.735	0.026	0.856	0.899	0.799	0.064
<b>OPNet</b> <sub>2023</sub> [42]	0.883	0.932	0.838	0.034	0.857	0.919	0.767	0.026	0.858	0.915	0.817	0.050
<b>AdaptCOD (Ours)</b>	<b>0.906</b>	<b>0.942</b>	<b>0.860</b>	<b>0.029</b>	<b>0.892</b>	<b>0.938</b>	<b>0.819</b>	<b>0.021</b>	<b>0.886</b>	<b>0.932</b>	<b>0.844</b>	<b>0.043</b>

表 6. 我们的 AdaptCOD 与最先进方法的对比. 如图所示, 我们的 AdaptCOD 相比先前方法有显著优势.  $\uparrow$ : 数值越高越好,  $\downarrow$ : 数值越低越好.

Method	NC4K (4,121)				COD10K-Test (2,026)				CAMO-Test (250)			
	$S_m \uparrow$	$\alpha E \uparrow$	$wF \uparrow$	$M \downarrow$	$S_m \uparrow$	$\alpha E \uparrow$	$wF \uparrow$	$M \downarrow$	$S_m \uparrow$	$\alpha E \uparrow$	$wF \uparrow$	$M \downarrow$
<b>ZoomNet</b> <sub>2022</sub> [45]	0.853	0.907	0.784	0.043	0.838	0.893	0.729	0.029	0.820	0.883	0.752	0.066
w/ Seg part (Ours)	0.858	0.912	0.795	0.040	0.845	0.897	0.738	0.028	0.824	0.886	0.755	0.065
w/ AdaptCOD (Ours)	<b>0.868</b>	<b>0.920</b>	<b>0.801</b>	<b>0.038</b>	<b>0.851</b>	<b>0.903</b>	<b>0.745</b>	<b>0.026</b>	<b>0.832</b>	<b>0.894</b>	<b>0.769</b>	<b>0.063</b>
<b>HitNet</b> <sub>2022</sub> [19]	0.870	0.921	0.825	0.039	0.868	0.932	0.798	0.024	0.844	0.902	0.801	0.057
w/ Seg part (Ours)	0.876	0.925	0.836	0.037	0.873	0.937	0.805	0.023	0.849	0.905	0.811	0.053
w/ AdaptCOD (Ours)	<b>0.884</b>	<b>0.930</b>	<b>0.841</b>	<b>0.035</b>	<b>0.880</b>	<b>0.940</b>	<b>0.812</b>	<b>0.022</b>	<b>0.860</b>	<b>0.912</b>	<b>0.824</b>	<b>0.050</b>

表 7. 所提框架 AdaptCOD 在先进 COD 模型 ZoomNet [45] 和 HitNet [19] 中的应用.w/ Seg part: 配备 AdaptCOD 中分割部分的模型. w/ AdaptCOD: 配备 AdaptCOD 中分割部分和背景检索部分的模型.

Method	CVC300 [2]		CVC-ClinicDB [1]		Kvasir [21]		ColonDB [55]		ETIS [50]	
	DICE $\uparrow$	mIoU $\uparrow$	DICE $\uparrow$	mIoU $\uparrow$	DICE $\uparrow$	mIoU $\uparrow$	DICE $\uparrow$	mIoU $\uparrow$	DICE $\uparrow$	mIoU $\uparrow$
<b>CASCADE</b> <sub>2023</sub> [48]	90.47	83.79	94.34	89.98	92.58	87.76	82.54	<b>74.53</b>	80.07	72.58
w / Seg part (Ours)	90.71	83.88	94.79	90.55	92.87	87.80	<b>82.58</b>	74.50	82.09	73.55
w / AdaptCOD (Ours)	<b>91.05</b>	<b>83.98</b>	<b>95.16</b>	<b>90.95</b>	<b>92.90</b>	<b>87.86</b>	82.56	74.42	<b>83.49</b>	<b>75.26</b>

表 8. 我们的 AdaptCOD 在先进息肉分割模型 CASCADE [48] 中的应用. w/ Seg part: 配备 AdaptCOD 中分割部分的 CASCADE 模型.



图8. 失败样例. 定位部分对目标的初始预测完全错误是导致失败案例的主要原因。

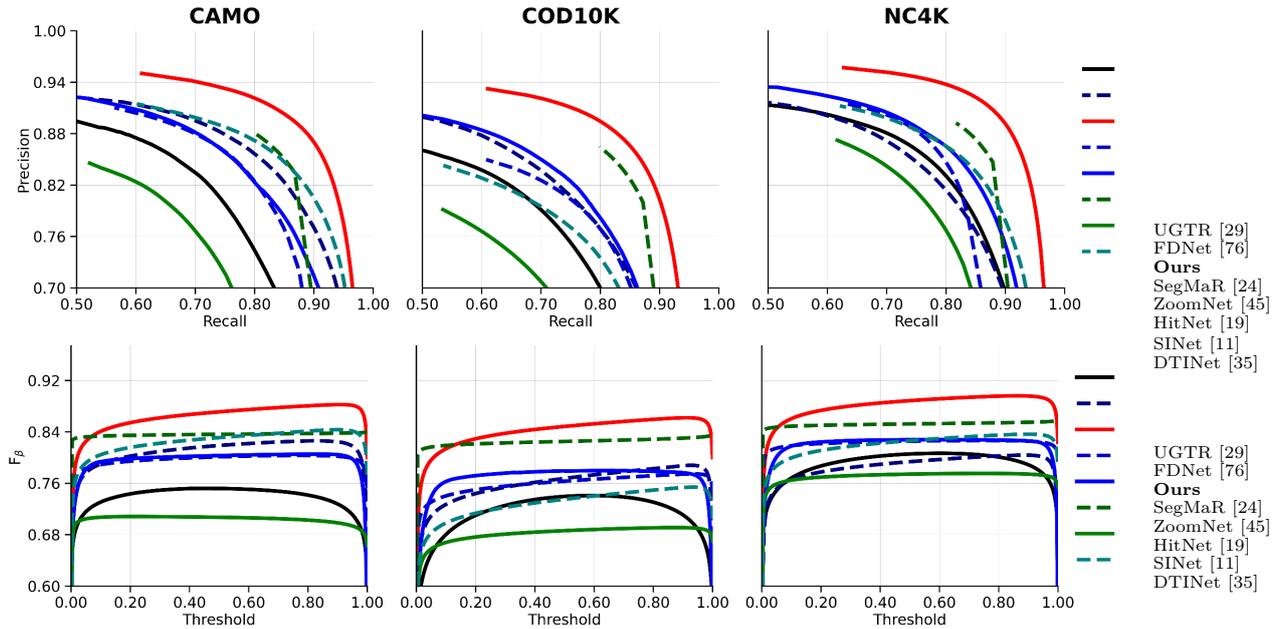


图9. 所提出的 AdaptCOD 与目前最先进的算法在三个 COD 数据集上的 PR 曲线和  $F_\beta$  曲线。

围的适当上下文区域。还开发了一个背景检索模块，以进一步优化预测的伪装目标。实验结果表明了所提方法的有效性。我们还展示了现有 COD 方法也可以从所提方法中受益，体现了其泛化能力。

**未来展望:** 尽管所提方法表现出色，但未来仍有一些值得探索的方向：

1) 定位增强: 由于 AdaptCOD 框架通过逐步优化分割结果实现目标检测，伪装目标的精确定位对性能影响显著，尤其是在极复杂场景中。

2) 背景检索方式优化: 当前 AdaptCOD 通过度量背景索引向量与候选前景位置的相似度来检索背景区域，以细化局部细节。我们认为，引入更高效的度量方法（如基于图神经网络的关系建模或动态特征对齐）可进一步提升伪装目标边界附近的分割质量，减少背景误检或目标边缘模糊问题。

## 参考文献

- Bernal, J., Sánchez, F.J., Fernández-Esparrach, G., Gil, D., Rodríguez, C., Vilariño, F.: Wm-dova maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians. *CMIG* **43**, 99–111 (2015)
- Bernal, J., Sánchez, J., Vilarino, F.: Towards automatic polyp detection with a polyp appearance model. *Pattern Recognition* **45**(9), 3166–3182 (2012)
- Chen, G., Liu, S.J., Sun, Y.J., Ji, G.P., Wu, Y.F., Zhou, T.: Camouflaged object detection via context-aware cross-level fusion. *IEEE Transactions on Circuits and Systems for Video Technology* **32**(10), 6981–6993 (2022)
- Cheng, X., Xiong, H., Fan, D.P., Zhong, Y., Harandi, M., Drummond, T., Ge, Z.: Implicit motion handling for video camouflaged object detection. In: *CVPR* (2022)

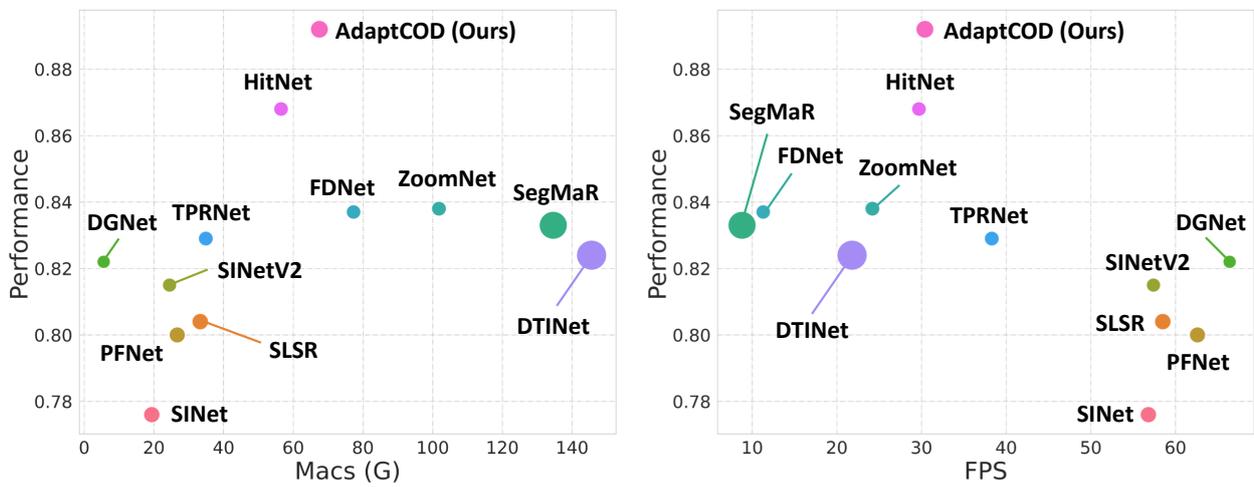


图 10. 模型性能与效率的关系. 我们在 COD10K 数据集上报告了不同方法的 S-measure 性能, 其中彩色散点的大小越大, 表示模型参数越多.

- Chu, X., Tian, Z., Wang, Y., Zhang, B., Ren, H., Wei, X., Xia, H., Shen, C.: Twins: Revisiting the design of spatial attention in vision transformers. In: NeurIPS (2021)
- De Boer, P.T., Kroese, D.P., Mannor, S., Rubinstein, R.Y.: A tutorial on the cross-entropy method. *Ann. Oper. Res.* **134**(1), 19–67 (2005)
- Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding. In: NAACL (2018)
- Fan, D.P., Cheng, M.M., Liu, Y., Li, T., Borji, A.: Structure-measure: A new way to evaluate foreground maps. In: IEEE ICCV (2017)
- Fan, D.P., Gong, C., Cao, Y., Ren, B., Cheng, M.M., Borji, A.: Enhanced-alignment measure for binary foreground map evaluation. In: IJCAI (2018)
- Fan, D.P., Ji, G.P., Cheng, M.M., Shao, L.: Concealed object detection. *IEEE TPAMI* **44**(10), 6024–6042 (2022)
- Fan, D.P., Ji, G.P., Sun, G., Cheng, M.M., Shen, J., Shao, L.: Camouflaged object detection. In: IEEE CVPR (2020)
- Fan, D.P., Ji, G.P., Zhou, T., Chen, G., Fu, H., Shen, J., Shao, L.: Pranel: Parallel reverse attention network for polyp segmentation. In: MICCAI (2020)
- Fan, D.P., Zhang, J., Xu, G., Cheng, M.M., Shao, L.: Salient objects in clutter. *IEEE TPAMI* **45**(2), 2344–2366 (2023)
- Pérez-de la Fuente, R., Delclòs, X., Peñalver, E., Speranza, M., Wierzos, J., Ascaso, C., Engel, M.S.: Early evolution and ecology of camouflage in insects. *Proceedings of the National Academy of Sciences* **109**(52), 21414–21419 (2012)
- Han, K., Xiao, A., Wu, E., Guo, J., Xu, C., Wang, Y.: Transformer in transformer. In: NeurIPS (2021)
- He, C., Li, K., Zhang, Y., Tang, L., Zhang, Y., Guo, Z., Li, X.: Camouflaged object detection with feature decomposition and edge reconstruction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 22046–22055 (2023)
- He, R., Dong, Q., Lin, J., Lau, R.W.: Weakly-supervised camouflaged object detection with scribble annotations. In: AAAI (2023)
- Hou, Q., Cheng, M.M., Hu, X., Borji, A., Tu, Z., Torr, P.: Deeply supervised salient object detection with short connections. *IEEE TPAMI* **41**(4), 815–828 (2019)
- Hu, X., Wang, S., Qin, X., Dai, H., Ren, W., Luo, D., Tai, Y., Shao, L.: High-resolution iterative feedback network for camouflaged object detection. In: AAAI (2023)
- Huang, Z., Dai, H., Xiang, T.Z., Wang, S., Chen, H.X., Qin, J., Xiong, H.: Feature shrinkage pyramid for camouflaged object detection with transformers. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5557–5566 (2023)
- Jha, D., Smedsrud, P.H., Riegler, M.A., Halvorsen, P., Lange, T.d., Johansen, D., Johansen, H.D.: Kvasir-seg: A segmented polyp dataset. In: MMM (2020)
- Ji, G.P., Fan, D.P., Chou, Y.C., Dai, D., Liniger, A., Van Gool, L.: Deep gradient learning for efficient camouflaged object detection. *MIR* (2022), doi: <https://doi.org/10.1007/s11633-022-1365-9>



图 11. COD10K 测试集挑战性样本的定性对比.

23. Ji, G.P., Zhu, L., Zhuge, M., Fu, K.: Fast camouflaged object detection via edge-based reversible re-calibration network. *PR* **123**, 108414 (2022)
24. Jia, Q., Yao, S., Liu, Y., Fan, X., Liu, R., Luo, Z.: Segment, magnify and reiterate: Detecting camouflaged objects the hard way. In: *IEEE CVPR* (2022)
25. Lamdouar, H., Yang, C., Xie, W., Zisserman, A.: Betrayed by motion: Camouflaged object discovery via motion segmentation. In: *ACCV* (2020)
26. Le, T.N., Cao, Y., Nguyen, T.C., Le, M.Q., Nguyen, K.D., Do, T.T., Tran, M.T., Nguyen, T.V.: Camouflaged instance segmentation in-the-wild: Dataset, method, and benchmark suite. *IEEE TIP* **31**, 287–300 (2022)
27. Le, T.N., Nguyen, T.V., Nie, Z., Tran, M.T., Sugimoto, A.: Anabranched network for camouflaged object segmentation. *CVIU* **184**, 45–56 (2019)
28. Le, X., Mei, J., Zhang, H., Zhou, B., Xi, J.: A learning-based approach for surface defect detection using small image datasets. *Neurocomputing* (2020)
29. Li, A., Zhang, J., Lv, Y., Liu, B., Zhang, T., Dai, Y.: Uncertainty-aware joint salient object and camouflaged object detection. In: *IEEE CVPR* (2021)
30. Li, C., Jiao, G.: Einet: camouflaged object detection with pyramid vision transformer. *JEI* **31**(5), 053002 (2022)
31. Li, Z.Y., Gao, S., Cheng, M.M.: Exploring feature self-relation for self-supervised transformer. *arXiv preprint arXiv:2206.05184* (2022)
32. Lin, J., Tan, X., Xu, K., Ma, L., Lau, R.W.: Frequency-aware camouflaged object detection. *ACM TMCCA* **19**(2), 1–16 (2023)
33. Lin, T.Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S.: Feature pyramid networks for object detection. In: *IEEE CVPR* (2017)
34. Liu, Y., Cheng, M.M., Fan, D.P., Zhang, L., Bian, J.W., Tao, D.: Semantic edge detection with diverse deep supervision. *International Journal of Computer Vision* **130**(1), 179–198 (2022)
35. Liu, Z., Zhang, Z., Wu, W.: Boosting camouflaged object detection with dual-task interactive transformer. *ICPR* (2022)
36. Loshchilov, I., Hutter, F.: Sgdr: Stochastic gradient descent with warm restarts. *ICLR* (2017)
37. Lv, Y., Zhang, J., Dai, Y., Li, A., Barnes, N., Fan, D.P.: Towards deeper understanding of camouflaged object detection. *IEEE TCSVT* (2023)
38. Lv, Y., Zhang, J., Dai, Y., Li, A., Liu, B., Barnes, N., Fan, D.P.: Simultaneously localize, segment and rank the camouflaged objects. In: *IEEE CVPR* (2021)
39. Margolin, R., Zelnik-Manor, L., Tal, A.: How to evaluate foreground maps? In: *IEEE CVPR* (2014)
40. Mátyus, G., Luo, W., Urtasun, R.: Deeproadmapper: Extracting road topology from aerial images. In: *IEEE ICCV* (2017)
41. Mei, H., Ji, G.P., Wei, Z., Yang, X., Wei, X., Fan, D.P.: Camouflaged object segmentation with distraction mining. In: *IEEE CVPR* (2021)
42. Mei, H., Xu, K., Zhou, Y., Wang, Y., Piao, H., Wei, X., Yang, X.: Camouflaged object segmentation with omni perception. *International Journal of Computer Vision* pp. 1–16 (2023)
43. Mei, H., Yang, X., Zhou, Y., Ji, G.P., Wei, X., Fan, D.P.: Distraction-aware camouflaged object segmentation. *SCIS* (2023)
44. Mondal, A., Ghosh, S., Ghosh, A.: Partially camouflaged object tracking using modified probabilistic neural network and fuzzy energy based active contour. *International Journal of Computer Vision* **122**, 116–148 (2017)
45. Pang, Y., Zhao, X., Xiang, T.Z., Zhang, L., Lu, H.: Zoom in and out: A mixed-scale triplet network for camouflaged object detection. In: *IEEE CVPR* (2022)
46. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al.: Pytorch: An imperative style, high-performance deep learning library. In: *NeurIPS* (2019)
47. Perazzi, F., Krähenbühl, P., Pritch, Y., Hornung, A.: Saliency filters: Contrast based filtering for salient region detection. In: *IEEE CVPR* (2012)
48. Rahman, M.M., Marculescu, R.: Medical image segmentation via cascaded attention decoding. In: *IEEE WACV* (2023)

49. Ren, J., Hu, X., Zhu, L., Xu, X., Xu, Y., Wang, W., Deng, Z., Heng, P.A.: Deep texture-aware features for camouflaged object detection. *IEEE TCSVT* **33**(3), 1157–1167 (2023)
50. Silva, J., Histace, A., Romain, O., Dray, X., Granado, B.: Toward embedded detection of polyps in wce images for early diagnosis of colorectal cancer. *International journal of computer assisted radiology and surgery* **9**, 283–293 (2014)
51. Srinivas, A., Lin, T.Y., Parmar, N., Shlens, J., Abbeel, P., Vaswani, A.: Bottleneck transformers for visual recognition. In: *IEEE CVPR* (2021)
52. Sun, Y., Chen, G., Zhou, T., Zhang, Y., Liu, N.: Context-aware cross-level fusion network for camouflaged object detection. In: *IJCAI* (2021)
53. Sun, Y., Wang, S., Chen, C., Xiang, T.Z.: Boundary-guided camouflaged object detection. In: *IJCAI* (2022)
54. Tabernik, D., Šela, S., Skvarč, J., Skočaj, D.: Segmentation-based deep-learning approach for surface-defect detection. *J. Intell. Manuf.* **31**(3), 759–776 (2020)
55. Tajbakhsh, N., Gurudu, S.R., Liang, J.: Automated polyp detection in colonoscopy videos using shape and context information. *IEEE TMI* **35**(2), 630–644 (2015)
56. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. In: *NeurIPS* (2017)
57. Wang, H., Wang, X., Sun, F., Song, Y.: Camouflaged object segmentation with transformer. In: *ICCSIP* (2021)
58. Wang, W., Xie, E., Li, X., Fan, D.P., Song, K., Liang, D., Lu, T., Luo, P., Shao, L.: Pvt v2: Improved baselines with pyramid vision transformer. *CVMJ* **8**(3), 415–424 (2022)
59. Wu, F., Li, X., Zhang, Y., Hu, K.: Transcoop: Cooperation of transformers and cnns for camouflaged object segmentation. In: *IEEE ICME* (2022)
60. Wu, H., Xiao, B., Codella, N., Liu, M., Dai, X., Yuan, L., Zhang, L.: Cvt: Introducing convolutions to vision transformers. In: *IEEE ICCV* (2021)
61. Wu, M., Zhang, X., Sun, X., Zhou, Y., Chen, C., Gu, J., Sun, X., Ji, R.: Difnet: Boosting visual information flow for image captioning. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 18020–18029 (2022)
62. Xie, S., Tu, Z.: Holistically-nested edge detection. In: *IEEE ICCV* (2015)
63. Xu, B., Liang, H., Liang, R., Chen, P.: Locate globally, segment locally: A progressive architecture with knowledge review network for salient object detection. In: *AAAI* (2021)
64. Xu, W., Xu, Y., Chang, T., Tu, Z.: Co-scale conv-attentional image transformers. In: *IEEE ICCV* (2021)
65. Yang, F., Zhai, Q., Li, X., Huang, R., Luo, A., Cheng, H., Fan, D.P.: Uncertainty-guided transformer reasoning for camouflaged object detection. In: *IEEE ICCV* (2021)
66. Yang, J., Li, C., Zhang, P., Dai, X., Xiao, B., Yuan, L., Gao, J.: Focal attention for long-range interactions in vision transformers. In: *NeurIPS* (2021)
67. Yin, B., Zhang, X., Hou, Q., Sun, B.Y., Fan, D.P., Van Gool, L.: Camoformer: Masked separable attention for camouflaged object detection. *arXiv preprint arXiv:2212.06570* (2022)
68. Zhai, Q., Li, X., Yang, F., Chen, C., Cheng, H., Fan, D.P.: Mutual graph learning for camouflaged object detection. In: *CVPR* (2021)
69. Zhai, Q., Li, X., Yang, F., Jiao, Z., Luo, P., Cheng, H., Liu, Z.: Mgl: Mutual graph learning for camouflaged object detection. *IEEE TIP* **32**, 1897–1910 (2023)
70. Zhai, W., Cao, Y., Xie, H., Zha, Z.J.: Deep texton-coherence network for camouflaged object detection. *IEEE TMM* (2022)
71. Zhang, M., Xu, S., Piao, Y., Shi, D., Lin, S., Lu, H.: Preynet: Preying on camouflaged objects. In: *ACM MM* (2022)
72. Zhang, Q., Ge, Y., Zhang, C., Bi, H.: Tprnet: camouflaged object detection via transformer-induced progressive refinement network. *TVCJ* pp. 1–15 (2022)
73. Zhang, X., Sun, X., Luo, Y., Ji, J., Zhou, Y., Wu, Y., Huang, F., Ji, R.: Rstnet: Captioning with adaptive attention on visual and non-visual words. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. pp. 15465–15474 (2021)
74. Zheng, D., Zheng, X., Yang, L.T., Gao, Y., Zhu, C., Ruan, Y.: Mfn: Multi-view feature fusion network for camouflaged object detection. In: *WACV* (2023)

75. Zheng, H., Fu, J., Zha, Z.J., Luo, J.: Looking for the devil in the details: Learning trilinear attention sampling network for fine-grained image recognition. In: IEEE CVPR (2019)
76. Zhong, Y., Li, B., Tang, L., Kuang, S., Wu, S., Ding, S.: Detecting camouflaged object in frequency domain. In: IEEE CVPR (2022)
77. Zhou, Y., Li, Z., Guo, C.L., Bai, S., Cheng, M.M., Hou, Q.: Srformer: Permuted self-attention for single image super-resolution. In: ICCV (2023)
78. Zhu, H., Li, P., Xie, H., Yan, X., Liang, D., Chen, D., Wei, M., Qin, J.: I can find you! boundary-guided separated attention network for camouflaged object detection. In: AAAI (2022)