

重新思考在无约束场景中检测显著与伪装目标

周张俊^{1*} 李一平^{1*} 钟春林^{1*} 黄佳诺¹ 裴佳伦² 李华³ 唐赫^{1†}

¹ 华中科技大学 软件学院, ² 香港中文大学 计算机科学与工程学院

³ 海南大学 计算机科学与技术学院

摘要

人类视觉系统在感知显著目标与伪装目标时存在不同的神经机制,但现有模型在区分这两类任务时仍面临困难。具体而言,显著目标检测 (*Salient Object Detection, SOD*) 模型常常将伪装目标误判为显著目标,而伪装目标检测 (*Camouflaged Object Detection, COD*) 模型则容易将显著目标误判为伪装目标。我们推测这一问题主要源于两个方面: (i) 当前 *SOD* 与 *COD* 数据集的特定标注范式,以及 (ii) 现有模型缺乏显式的属性关系建模机制。现有 *SOD/COD* 数据集普遍采用互斥性约束,默认场景中仅包含显著或伪装目标,这与真实世界的复杂场景存在较大偏差。此外,当前 *SOD/COD* 方法大多针对这种高度受限的数据集设计,缺乏对显著与伪装目标关系的显式建模。为推动非受限显著与伪装目标检测的发展,本文构建了一个大规模数据集 **USC12K**,该数据集具有全面的标注,并覆盖了显著与伪装目标所有可能共存的四类场景。为显式建模显著与伪装目标之间的关系,我们提出了新模型 **USCNet**,该模型引入了两种不同的提示查询机制,分别用于建模样本间与样本内的属性关系。此外,我们设计了新指标 **CSCS**,用于评估模型区分显著与伪装目标的能力。实验结果表明,我们的方法在所有场景下均取得了当前最优性能。详见: <https://github.com/ssecv/USCNet>。

1. Introduction

注意力机制是人类的关键认知功能之一 [46]。在现实场景中,人们往往会被显著目标所吸引,而忽略伪装

表 1. 当 *SOD* 与 *COD* 模型跨任务应用时,普遍存在误判现象。左: *SOD* 模型在 *COD* 数据集上的推理结果, *VSCode* 使用 *SOD* 提示。右: *COD* 模型在 *SOD* 数据集上的推理结果, *VSCode* 使用 *COD* 提示。评价指标为 F_β^ω , 其中 EV 表示期望值 (Expected Value)。

| SOD Models | COD Datasets | | EV | COD Models | SOD Datasets | | EV |
|-------------|--------------|--------|----|--------------|--------------|--------|----|
| | COD10K | NC4K | | | DUTS | HKU-IS | |
| ICON [77] | 0.6384 | 0.7522 | 0 | SINet-V2 [7] | 0.7412 | 0.7691 | 0 |
| F3Net [59] | 0.4327 | 0.6229 | 0 | PFNet [35] | 0.7361 | 0.7657 | 0 |
| VSCode [31] | 0.7145 | 0.8054 | 0 | VSCode [31] | 0.8614 | 0.8733 | 0 |

目标。从人类视觉识别系统的角度来看,显著目标和伪装目标代表着相对立的概念。显著目标检测 (*Salient Object Detection, SOD*) 旨在检测图像中人类视觉系统认为最显著或最吸引注意力的目标,而伪装目标检测 (*Camouflaged Object Detection, COD*) 旨在检测难以感知或与周围环境相融合的目标 [18]。在计算机视觉中,这两类任务通常通过二值掩码来表示模型的输出。前者模拟了人类聚焦显著目标的能力,而后者则模拟了人类发现伪装目标的能力。它们在诸多领域展现出重要潜力,例如医学图像分析中的异常检测 [53]、自动驾驶中的障碍物识别、军事侦察中的伪装检测 [23] 以及环境监测中的野生动物追踪 [51]。

现有的 *SOD* 和 *COD* 方法在显著目标与伪装目标的检测方面已经取得了显著进展。一些统一的方法 [18, 29, 31, 75] 通过联合训练 *SOD* 和 *COD* 数据集,提高了模型在这两类任务中的泛化能力。然而,我们观察到一个反直觉的现象:尽管显著目标和伪装目标在概念上是对立的, *SOD* 模型却常常将伪装目标误判为显著目标,而 *COD* 模型则将显著目标误判为伪装目标。这种误检是不可取的,并且违背了这两类任务的设计初衷。

为了进一步验证这一现象,我们选取了五个采用

*同等贡献。

†通讯作者: 唐赫, hetang@hust.edu.cn。

其原始预训练权重的模型，在四个数据集上进行推理。如表 1 所示，SOD 模型（例如 ICON [77]）在 COD 数据集 COD10K [8] 上的推理中，意外地获得了较高的 F_β^ω 分数 0.6384；而 COD 模型（例如 SINet-V2 [7]）在 SOD 数据集 DUTS [55] 上也取得了 0.7412 的分数。统一模型 VSCode [31] 在使用 COD 提示时，在 SOD 数据集 HKU-IS 上达到了 0.8733 的分数。这些方法的误检分数均显著高于期望值 0。更多示例可参见 [附录 §1](#)。

我们认为首要原因在于现有 SOD 与 COD 数据集的特定标注范式。尽管现有的 SOD 和 COD 数据集通过针对性设计极大地推动了各自领域的发展 [9, 11]，但其标注方式施加了严格约束，默认场景中仅包含显著目标或伪装目标，并且仅提供单一类型的标注。例如，在 COD 数据集中，某些场景中实际存在显著目标，但由于其并非伪装目标，这些目标被标注为背景。由此，COD 模型对显著目标失去敏感性，难以准确刻画显著性特征空间的边界。SOD 模型也存在类似问题。这种特定的标注范式在统一模型中表现得更为突出：统一模型旨在同时处理两类任务，但具有相同视觉特征的目标在不同数据集中可能会被赋予相互矛盾的标签。具体而言，显著目标在 COD 数据集中被标注为背景，而在 SOD 数据集中则标注为显著目标；伪装目标也存在同样的不一致性。这类标注差异在理论上可能导致多任务学习过程中信息丢失，从而削弱模型的泛化能力。

在本文中，我们的目标是推动无约束的显著与伪装目标检测的发展，使模型能够在所有可能的逻辑场景中同时检测显著与伪装目标。为此，我们构建了一个新的数据集，命名为 **USC12K**。该数据集包含 12,000 张具有完整标注的图像，涵盖了显著目标与伪装目标存在的四种逻辑场景。除了对现有数据集进行修正与优化外，我们还从互联网收集并人工标注了 2,617 张同时包含显著与伪装目标的图像，以及 1,436 张既不包含显著目标也不包含伪装目标的图像，以保证四种场景的分布平衡。我们在 USC12K 上评估了 22 种与 SOD 和 COD 任务相关的方法，建立了一个面向无约束显著与伪装目标检测的基线，旨在推动该领域的进一步研究。

我们认为第二个原因在于缺乏适应无约束条件下的属性关系建模机制。现有的 SOD 和 COD 模型在单任务性能上已取得显著成功。然而，这些模型主要针对受限的单场景数据集而设计，其基于单一属性的检测策略难以适应同时包含两种对立视觉模式的目标检测。

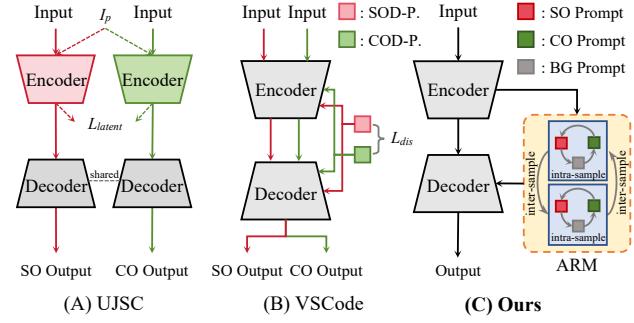


图 1. 统一模型的架构比较。SOD-P 和 COD-P 分别表示 SOD 与 COD 任务的提示。SO prompt、CO prompt 和 BG prompt 分别表示显著、伪装和背景的提示。

尽管统一模型通过共享网络组件实现了对两类属性目标的检测，但在无约束的显著与伪装目标检测中，其处理流程仍然是次优的。具体而言，这些模型通常采用分别学习显著与伪装目标检测的方式，而这种方式是针对以往受限数据集设计的。尽管部分模型尝试通过对比学习等技术在两个属性之间建立联系，例如图 1 所示，UJSC [18] 通过在额外数据集 I_p 上最小化潜在空间损失 L_{latent} 来拉开显著编码器与伪装编码器特征空间分布的距离；VSCode [31] 通过最小化判别损失 L_{dis} 来降低显著与伪装提示之间的相似性。然而，这些方法仍然缺乏在网络内部对显著与伪装目标关系的显式建模，损失函数主要建模的是样本间的显著与伪装关系，而无法捕捉单个样本内部的属性关系。因此，这些模型在无约束场景下难以有效学习二者的深层共性与差异。

为在无约束场景下显式建模显著与伪装目标之间的关系，我们提出了一种具有统一优化流程的网络，命名为 **Unconstrained Salient and Camouflaged Net (USC-Net)**，其具有以下两点优势：1) USCNet 基于 SAM 构建属性提示框架，引入带有适配器的 SAM 编码器，以充分利用 SAM 的场景泛化能力，并提升在复杂场景中的适应性；2) 提出了一个属性关系建模（Aspect Relation Modeling, ARM）模块，用于结合两类互补的提示查询来建模显著与伪装之间的关系：其中，Inter-SPQ 通过建模跨样本的属性关系来学习通用特征并捕获全局信息；而 Intra-SPQ 则聚焦于单个样本的上下文信息，以建模样本内部的属性关系，从而学习样本内部显著与伪装目标的具体联系。

此外，现有评价指标无法有效量化模型在显著目标与伪装目标之间的混淆情况，因为它们主要评估的是前景与背景的分离程度，例如加权 F-measure [34]。为

弥补这一不足，我们设计了一个新的指标名为显伪混淆分数 (Camouflage-Saliency Confusion Score, CSCS)，用于评估模型区分显著与伪装目标的能力。

综上，我们的主要贡献如下：

- 构建了一个大规模的无约束显著与伪装目标检测数据集 **USC12K**，并提供了完整标注。据调查，这是首个在显著与伪装目标的存在性上不受约束的数据集。
- 提出了一个基于 SAM 的模型 **USCNet**，引入 ARM 模块以学习属性关系，并设计了两种查询机制，用于建模样本间与样本内的显著与伪装关系。
- 提出了一个新的评价指标 **CSCS**，用于评估模型对于显著目标与伪装目标的混淆情况。
- 在 USC12K 数据集上评估了 22 种相关方法，建立了一个全面的基线。我们的模型在所有指标和所有场景中均取得了最先进的性能。

2. 相关工作

2.1. 显著和伪装目标检测

经典的 SOD 与 COD。 近年来，SOD 模型主要致力于更好地检测图像中的显著目标，方法大致分为基于注意力 [26, 40, 45, 74]、基于多层特征 [10, 14, 38, 58] 以及基于递归的方法 [5, 25, 56]。显著性检测 [27, 41, 72, 77] 主要关注在保持结构信息的同时实现准确预测。相比之下，COD 方法 [7, 12, 35, 37, 39] 更强调边缘与纹理感知，主要可分为基于多层特征 [49, 68, 71, 73] 和边缘联合学习 [12, 52, 70] 两类。总体而言，SOD 与 COD 模型均针对各自任务独立设计，缺乏对二者关系的建模。

统一方法。 近年来，一些工作 [18, 29, 31, 75] 已经开始尝试统一 SOD 与 COD 两个任务，并尝试在两者之间建立联系。VSCode [31] 通过最小化 SOD 与 COD 任务提示的余弦相似度来学习可区分的显著提示和伪装提示。UJSC [18] 引入额外的 PASCAL VOC 2007 数据集，以增强显著特征提取器与伪装特征提取器之间的区分能力。Spider [75] 通过挖掘全局上下文中前景/背景相关的语义线索来区分目标的不同属性。EVP [29] 学习任务特定的视觉提示，以区分显著与伪装目标。然而，现有统一模型由于缺乏对显著与伪装之间关系的直接建模，仍然难以有效区分二者。针对这一局限，我们引入了 ARM 模块，通过样本间与样本内交互显式建模显著与伪装目标之间的关系，从而提升模型在无约束场景下的区分能力。

2.2. SAM 应用

Segment Anything Model (SAM) [16] 在利用大型视觉模型进行场景分割方面取得了重要进展。当前利用 SAM 的工作 [2, 33, 64] 展示了其在下游任务中的适应性，尤其是在传统分割模型表现不佳的领域，如 EfficientSAM [64] 和 MedSAM [33]。近期，SAM2 [48] 的发布增强了原始 SAM 对视频内容的处理能力，同时在各类下游应用的图像分割中展示了更高的分割精度和推理效率 [22, 42, 43, 66]。

部分将 SAM 应用于 SOD 和 COD 的工作与我们的研究密切相关。MDSAM [11] 是基于 SAM 的多尺度细节增强显著目标检测模型，旨在提升 SOD 任务的性能和泛化能力。SAM-Adapter [2] 和 SAM2-Adapter [1] 提供了一种参数高效微调的方式，通过引入任务特定知识来增强 SAM 和 SAM2 在下游任务中的表现。基于 SAM 强大的泛化能力，我们尝试探索其在无约束场景下检测显著与伪装目标的潜力。

3. USC12K 数据集

现有 COD 数据集，如 COD10K [8]、CAMO [17] 和 NC4K [32]，主要包含仅有伪装目标的场景。类似地，SOD 数据集，如 DUTS [55] 和 HKU-IS [20]，则聚焦于仅含显著目标的场景。同时包含显著与伪装目标的样本极为稀少，即使存在也只对其中一类目标进行了标注，这很大程度上阻碍了无约束显著与伪装目标检测的发展。为此，我们提出了 **USC12K**，一个涵盖更全面、更复杂场景的无约束显著与伪装目标检测数据集。该数据集包含三类场景：同时存在显著与伪装目标的场景、仅包含单类目标的场景，以及不包含任何目标的场景。

3.1. 数据收集

在确保样本平衡的前提下，我们从 8 个不同来源收集了 12,000 张图像，并在人工筛选后将其划分为四类场景：(A) 仅含显著目标的场景：从 SOD 数据集 DUTS 和 HKU-IS 中选取的 3,000 张仅包含显著目标的图像；(B) 仅含伪装目标的场景：从 COD 数据集 COD10K 和 CAMO 中选取的 3,000 张仅包含伪装目标的图像；(C) 同时包含显著与伪装目标的场景：来自 COD 数据集 COD10K、CAMO 和 NC4K 的 342 张图像，以及来自 LSUI [44] 和 AWA2 [61] 的 41 张图像，再加上从互联网收集的 2,617 张图像，总计 3,000 张；(D) 不含显著与伪

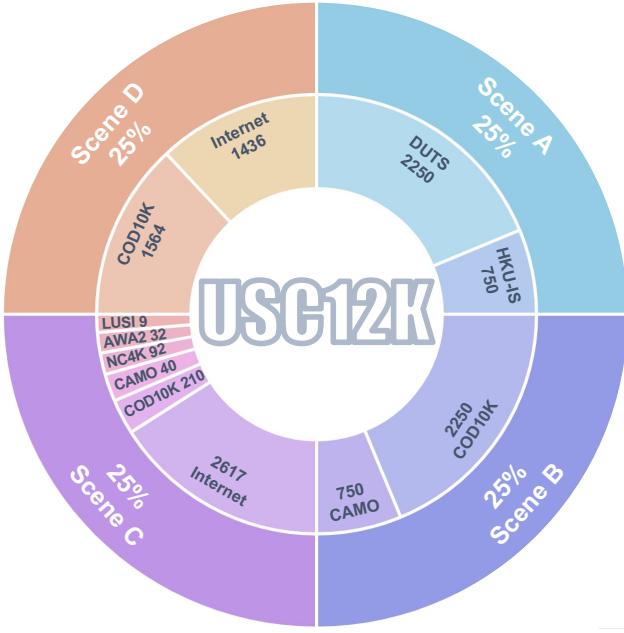


图 2. 四种场景的数据来源及分布情况。

装目标的场景，作为背景：来自 COD10K 的 1,564 张图像，以及从互联网收集的 1,436 张图像，总计 3,000 张。每类场景均经过人工标注人员的严格审核，以确保场景分类的正确性。最终，我们得到总计 12,000 张图像，其中训练集包含 8,400 张图像，测试集包含 3,600 张图像。数据来源如图 2 所示。

3.2. 数据标注

对于已完成场景分类的数据，我们对场景 A 和 B 保留原始标签。对于场景 C，我们先使用 SAM [16] 进行粗略标注，再通过人工修正来补充掩码标签，包括单类图像和从网络收集的数据。在类别标注过程中，我们将整个数据集中掩码之外的像素设置为 0。我们首先使用 CLIP [47] 获得初步粗分类结果，然后进行人工核查和精修。除 COD10K [8] 收集的图像已经包含伪装目标类别标签外，其他所有目标均需进行类别分类。部分 USC12K 数据集中不同场景的示例见图 3。随后，我们为每张图像分配类别标签，覆盖 9 个大类和 179 个子类。数据集的类别分布见图 3，详见附录 §6。

3.3. 数据分析

为了更深入地分析 SOD 与 COD 数据集，我们将 USC12K 与其他 13 个相关数据集进行了比较，包括：(1) 九个 SOD 数据集：SOD [36]、PASCAL-

表 2. 现存数据集的数据统计分析

| Dataset | #Ann. IMG | Class | Scene A | Scene B | Scene C | Scene D |
|---------------------|--------------|------------|-------------|-------------|-------------|-------------|
| SOD [36] | 300 | - | 300 | X | X | X |
| PASCAL-S [21] | 850 | - | 850 | X | X | X |
| ECSSD [65] | 1000 | - | 1000 | X | X | X |
| HKU-IS [20] | 4447 | - | 4447 | X | X | X |
| MSRA-B [28] | 5000 | - | 5000 | X | X | X |
| DUT-OMRON [67] | 5168 | - | 5168 | X | X | X |
| MSRA10K [4] | 10000 | - | 10000 | X | X | X |
| DUTS [55] | 15572 | - | 15572 | X | X | X |
| SOC [6] | 3000 | 80 | 3000 | X | X | X |
| CAMO [17] | 1250 | 8 | X | 1250 | X | X |
| CHAMELEON [50] | 76 | - | X | 76 | X | X |
| NC4K [32] | 4121 | - | X | 4121 | X | X |
| COD10K [8] | 7000 | 78 | X | 5066 | X | 1934 |
| USC12K(Ours) | 12000 | 179 | 3000 | 3000 | 3000 | 3000 |

S [21]、ECSSD [65]、HKU-IS [20]、MSRA-B [28]、DUT-OMRON [67]、MSRA10K [4]、DUTS [55] 和 SOC [6]；(2) 四个 COD 数据集：CAMO [17]、CHAMELEON [50]、COD10K [8] 和 NC4K [32]。表 2 给出了这些数据集的详细信息。可以看出，除 COD10K 外，所有 SOD 数据集仅包含显著目标，而几乎所有 COD 数据集仅包含伪装目标，这些数据集覆盖的场景相对有限。值得注意的是，尽管 COD 数据集 COD10K 包含一些显著目标图像以及不含任何目标的图像，但这些图像缺少标签，且未参与训练过程。相比之下，我们提出的 USC12K 数据集对场景不做限制，并为显著、伪装及背景三类提供了完整标注，同时保证了各场景的均衡分布。

4. USCNet 模型

概述。如图 4 所示，所提出的 USCNet 的主要组件包括：(1) 带适配器层的 SAM 图像编码器，用于提取目标特征表示；(2) 属性关系建模 (Aspect Relation Modeling, ARM) 模块，用于生成三类属性提示（显著、伪装、背景），在模块内部实现交互，并引入两种不同的提示查询机制，以在样本间和样本内两个层面建模属性关系；(3) 冻结的 SAM 掩码解码器，根据不同属性提示预测最终的显著、伪装和背景掩码。

4.1. 带适配器的 SAM 编码器

SAM [16] 包含图像编码器、提示编码器和掩码解码器。在 USCNet 中，我们利用 SAM 的提示架构来识别三类属性：显著、伪装和背景。属性提示由设计的 ARM 模块生成，无需人工提示。每个属性提示映射到一个独立的二值掩码。借鉴 SAM-Adapter [2]，USCNet 在 SAM

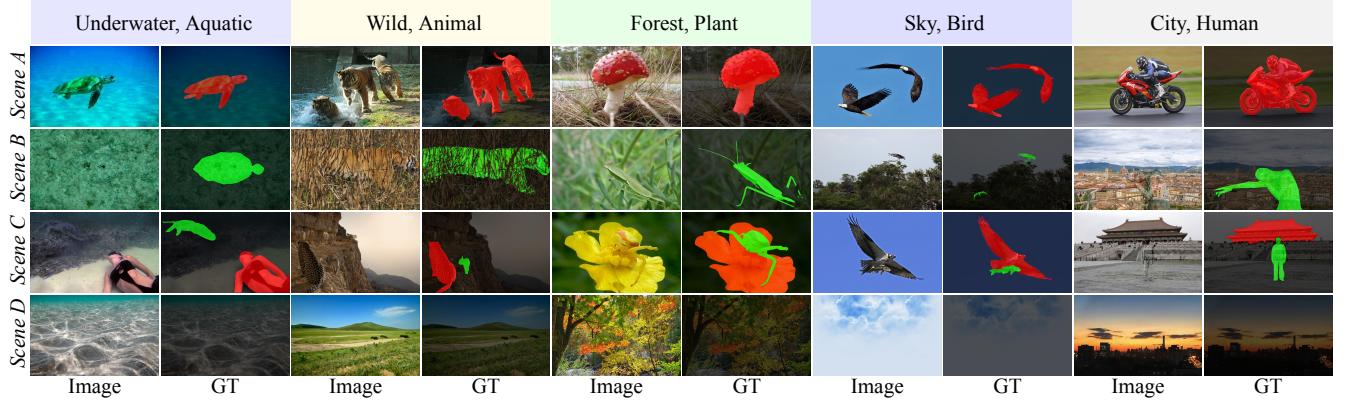


图 3. USC12K 数据集示例图像：场景 A：仅包含伪装目标；场景 B：仅包含显著目标；场景 C：显著目标与伪装目标同时存在；场景 D：背景，均不存在显著或伪装目标。更多示例可见附录 §6。

编码器的每一层集成了适配器，并采用参数高效微调的方法，如图 4 所示。通过该方法，USCNet 将 SOD 和 COD 的任务特定知识与大模型获得的通用知识融合，从而更好地适应无约束场景。

4.2. 属性关系建模 (Aspect Relation Modeling)

为了刻画无约束场景下显著与伪装目标之间的复杂关系，我们引入了属性关系建模 (Aspect Relation Modeling, ARM) 模块。

我们的核心观点是，在无约束场景中对显著与伪装目标关系进行建模，需要考虑两个维度的交互：(i) 样本间关系建模 (Inter-sample relationship modeling)：显著与伪装目标在不同样本中通常表现出不同的特征。对其进行建模有助于识别通用的判别特征，如大小、位置、颜色和纹理，这些特征对于区分两者至关重要。(ii) 样本内关系建模 (Intra-sample relationship modeling)：在显著与伪装目标共存的复杂场景中，仅依赖样本间建模可能不足。例如，当显著目标与伪装目标在颜色或类别上相似时，样本内的上下文信息对于准确区分二者变得必不可少。基于此，我们提出 ARM 模块，协同整合样本间提示查询 (Inter-Sample Prompt Query, Inter-SPQ) 与样本内提示查询 (Intra-Sample Prompt Query, Intra-SPQ)，分别用于不同样本间和样本内部判别性地查询属性特定提示。Inter-SPQ 用于提取样本之间的通用特征，捕捉所有样本普遍共有的属性；而 Intra-SPQ 则专注于提取单个样本内部的特征，强调每个特定样本中固有的独特上下文信息。

Inter-SPQ 由一组可学习的查询嵌入组成： $Q_{S_r} \in \mathbb{R}^{N \times C}$ 、 $Q_{C_r} \in \mathbb{R}^{N \times C}$ 和 $Q_{B_r} \in \mathbb{R}^{N \times C}$ ，其中 N 表示

查询数量， C 表示查询的维度。Inter-SPQ 在推理过程中保持不变，与具体样本无关。

Intra-SPQ 的生成过程如下：首先，从编码器中提取特征 F ，并通过由两层 3×3 卷积组成的注意力头 (attention head) 生成注意力图。该注意力图以真实标签为监督，引导其关注图像中的显著和伪装区域。随后，对注意力图的每个通道应用 sigmoid 函数处理。处理后的注意力图与原始特征 F 进行逐元素相乘，以提取属性特定特征。这些特征随后通过线性层进行下采样，从而生成 Intra-SPQ，确保其维度与 Inter-SPQ 一致。Intra-SPQ 会根据样本变化而在推理时动态更新。其生成过程可表示为：

$$[Q_{S_a}, Q_{C_a}, Q_{B_a}] = \text{Linear}(\sigma(\Phi_{AH}(F)) \otimes F), \quad (1)$$

其中 Q_{S_a} 、 Q_{C_a} 和 Q_{B_a} 分别表示显著、伪装和背景的 Intra-SPQ， σ 表示 sigmoid 函数， Φ_{AH} 表示注意力头， \otimes 表示逐元素相乘。

Intra-SPQ 捕捉特定图像的特征信息，而 Inter-SPQ 则辨别属性之间的基本差异。通过将两者进行相加整合，ARM 获得了处理无约束场景的能力，无论图像中是否存在显著或伪装目标。随后，我们利用自注意力 (self-attention, SA) 建立查询间的关系，并通过查询-to-图像 (Query-to-Image, Q2I) 注意力与图像嵌入 F 进行交互，最终生成属性提示：

$$P = \{P_S \in \mathbb{R}^{N \times C}, P_C \in \mathbb{R}^{N \times C}, P_B \in \mathbb{R}^{N \times C}\}.$$

该过程可表示为：

$$P = \text{MLP}(Q2I(SA(\text{Intra-SPQ} + \text{Inter-SPQ}), F)), \quad (2)$$

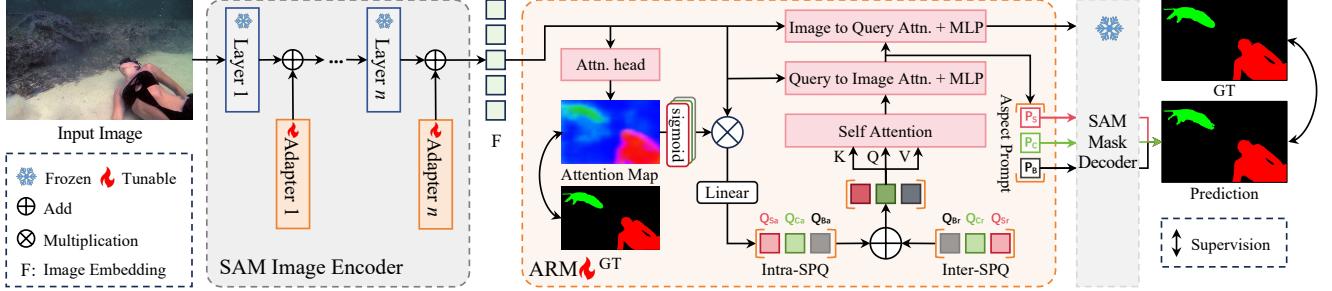


图 4. 我们 USCNet 的网络架构。USCNet 包括：带适配器的 SAM 图像编码器、ARM 模块，以及冻结的 SAM 掩码解码器。

其中 $Q2I$ 表示从查询-to-图像嵌入 F 的交叉注意力，使模型能够根据查询关注输入中的相关部分。此外，我们使用图像到查询 (Image-to-Query, $I2Q$) 注意力来关注与属性相关的特征。

通过向 SAM 预训练掩码解码器输入三类属性提示，可分别生成三张掩码 M_S 、 M_C 和 M_B ，对应显著、伪装和背景的预测结果。该过程可表示为：

$$[M_S, M_C, M_B] = MaskDe([P_S, P_C, P_B], F), \quad (3)$$

其中 $MaskDe$ 表示冻结的 SAM 掩码解码器。最终，通过对三张掩码应用 softmax 函数，生成最终预测结果。

4.3. 损失函数

我们使用真值标签对最终预测和注意力图进行监督。USCNet 的总损失函数定义如下：

$$\mathcal{L}_{Total} = \lambda_p \mathcal{L}_{pred.} + \lambda_a \mathcal{L}_{att.}, \quad (4)$$

其中， $\mathcal{L}_{pred.}$ 表示预测掩码的损失， $\mathcal{L}_{att.}$ 表示 ARM 模块中注意力图的损失。二者均采用 Focal Loss [24] 计算：

$$\mathcal{L}_{focal} = - \sum_{i=1}^C \alpha_{t_i} (1 - p_{t_i})^\gamma \log(p_{t_i}), \quad (5)$$

其中 p_{t_i} 为目标类别 t_i 的预测概率， α_{t_i} 为类别 t_i 的权重因子（根据像素数量比例经验地设置为背景、显著、伪装像素为 1:4:6）， γ 默认为 2，用于考虑不同目标的难易程度， C 为类别数量。

此外， λ_p 和 λ_a 分别经验地设为 1 和 0.5，用以平衡总损失。

5. USC12K 基准

我们在 USC12K 上评估了 22 个相关方法，以建立一个全面的基准。所有模型均在 USC12K 的训练集

(8,400 张图像) 上进行训练，并在 USC12K 的测试集 (3,600 张图像) 上进行测试。

评估指标. USC12K 基准涵盖三个不同的属性：显著 (saliency)、伪装 (camouflage) 和背景 (background)。我们采用三种常用的语义分割指标：mAcc↑、IoU↑ 和 mIoU↑ [30, 63]。受 [19] 启发，我们还采用 AUC↑、SI-AUC↑、 F_m^β ↑、SI- F_m^β ↑、 F_{max}^β ↑、SI- F_{max}^β ↑、 E_m ↑ 来评估模型对不同尺寸目标的检测能力。

此外，为了评估模型区分显著目标与伪装目标的能力，我们提出了一种新的指标：伪装-显著混淆得分 (Camouflage-Saliency Confusion Score, CSCS↓)，其计算公式为：

$$CSCS = \frac{1}{2} \left(\frac{\mathcal{P}_{CS}}{\mathcal{P}_{BS} + \mathcal{P}_{SS} + \mathcal{P}_{CS}} + \frac{\mathcal{P}_{SC}}{\mathcal{P}_{BC} + \mathcal{P}_{SC} + \mathcal{P}_{CC}} \right), \quad (6)$$

其中 $\mathbb{P} = \{\mathcal{P}_{\lambda\theta} | \lambda \in \Theta, \theta \in \Theta\}$ ， $\Theta = \{B, C, S\}$ ，B、C 和 S 分别表示背景、伪装和显著。 \mathcal{P}_{CS} 表示伪装区域被预测为显著的区域， \mathcal{P}_{SC} 表示显著区域被预测为伪装的区域；两者均属于混淆区域。CSCS 越低，表示模型区分显著与伪装目标的能力越强。

对比方法. 我们与 22 个相关模型进行比较，包括 (I) SOD 模型：GateNet [76], F3Net [59], MSFNet [72], VST [27], EDN [60], PGNet [62], ICON [77], MD-SAM [11]；(II) COD 模型：SINet-V2 [7], PFNet [35], ZoomNet [37], FEDER [12], ICEG [13], PRNet [15], CamoDiffusion [3], CamoFormer [69], PGT [57], SAM-Adapter [2] 和 SAM2-Adapter [1]；(III) 统一方法：VS-Code [31], Spider [75] 和 EVP [29]。

技术细节. 所有模型均在 USC12K 训练集上重新训练，输入分辨率为 352×352 。数据增强包括水平翻转和随机裁剪。实验在一块 NVIDIA L40 GPU 上进行。各模型的训练参数数量详见表 3。我们采用 SAM2-Adapter [1] 中的 SAM2 hiera-large 版本。优化器为 AdamW，并采

表3. 与22种相关方法在显著目标检测(SOD)和伪装目标检测(COD)上的定量比较。IoUs↑: 显著目标的IoU分数。IoU_C↑: 伪装目标的IoU分数。表现最好的两个分数分别用**红色**和**绿色**高亮显示。表中所有指标均以百分比(%)表示。我们使用mIoU↑、mAcc↑和CSCS↓来评估模型在所有场景下的表现。

| Task | Model | Venue | Update Para.(M) | Scene A | | Scene B | | Scene C | | Overall Scenes | | | |
|---------------|-------------------|----------|-----------------|------------------|------------------|------------------|------------------|------------------|------------------|----------------|-------|-------|--|
| | | | | IoU _S | IoU _C | IoU _S | IoU _C | IoU _S | IoU _C | mIoU | mAcc | CSCS | |
| SOD | GateNet [76] | ECCV'20 | 128 | 68.32 | 54.26 | 66.85 | 35.03 | 65.08 | 44.17 | 68.27 | 78.07 | 11.30 | |
| | F3Net [59] | AAAI'20 | 26 | 70.05 | 52.62 | 67.20 | 36.38 | 66.12 | 44.81 | 68.80 | 77.86 | 9.36 | |
| | MSFNet [72] | MM'21 | 28 | 70.14 | 54.78 | 69.92 | 36.64 | 66.69 | 45.89 | 69.40 | 79.77 | 9.90 | |
| | VST [27] | ICCV'21 | 43 | 68.14 | 49.82 | 61.61 | 22.56 | 63.18 | 38.45 | 65.55 | 74.77 | 11.30 | |
| | EDN [60] | TIP'22 | 43 | 71.59 | 57.94 | 69.37 | 37.70 | 68.00 | 48.27 | 70.70 | 80.60 | 9.23 | |
| | PGNet [62] | CVPR'22 | 73 | 74.69 | 57.31 | 71.94 | 37.21 | 70.72 | 48.78 | 71.82 | 80.76 | 7.71 | |
| | ICON [77] | TPAMI'22 | 32 | 68.09 | 50.57 | 67.48 | 30.65 | 65.86 | 45.53 | 68.99 | 79.53 | 10.24 | |
| COD | MDSAM [11] | MM'24 | 14 | 72.96 | 56.16 | 67.21 | 36.06 | 69.67 | 49.05 | 71.67 | 82.92 | 10.21 | |
| | SINet-V2 [7] | TPAMI'21 | 27 | 72.96 | 56.16 | 67.21 | 36.06 | 69.50 | 47.47 | 70.20 | 79.58 | 8.83 | |
| | PFNet [35] | CVPR'21 | 47 | 69.07 | 52.83 | 67.20 | 32.81 | 65.73 | 43.76 | 68.30 | 78.00 | 10.04 | |
| | ZoomNet [37] | CVPR'22 | 33 | 74.11 | 51.12 | 66.79 | 29.69 | 66.43 | 43.28 | 68.35 | 77.72 | 8.88 | |
| | FDER [12] | CVPR'23 | 44 | 74.35 | 58.04 | 67.66 | 32.26 | 68.65 | 46.46 | 70.32 | 81.27 | 10.01 | |
| | PRNet [15] | TCSVT'24 | 13 | 76.10 | 61.54 | 60.10 | 32.16 | 68.68 | 50.88 | 71.87 | 82.89 | 8.40 | |
| | ICEG [13] | ICLR'24 | 100 | 73.67 | 68.38 | 68.43 | 44.33 | 69.22 | 58.71 | 74.68 | 83.53 | 8.16 | |
| | CamoDiffusion [3] | AAAI'24 | 72 | 75.01 | 59.39 | 53.49 | 45.03 | 63.49 | 52.80 | 70.70 | 77.73 | 7.73 | |
| | CamoFormer [69] | TPAMI'24 | 71 | 75.88 | 66.19 | 73.33 | 44.14 | 71.86 | 56.09 | 74.81 | 84.17 | 7.57 | |
| | PGT [57] | CVIU'24 | 68 | 72.75 | 61.51 | 70.01 | 41.21 | 71.46 | 56.83 | 75.03 | 83.35 | 9.09 | |
| Unified | SAM-Adapter [2] | ICCVW'23 | 4.11 | 78.90 | 67.69 | 68.19 | 27.73 | 70.66 | 52.69 | 73.38 | 83.35 | 10.28 | |
| | SAM2-Adapter [1] | arXiv'24 | 4.36 | 78.75 | 70.28 | 69.01 | 38.20 | 71.42 | 56.71 | 74.98 | 84.74 | 9.12 | |
| | EVP [29] | CVPR'23 | 4.95 | 75.85 | 59.81 | 71.41 | 37.64 | 70.30 | 50.36 | 72.16 | 79.96 | 8.67 | |
| | VSCode [31] | CVPR'24 | 60 | 76.04 | 60.31 | 72.58 | 39.46 | 71.08 | 55.14 | 74.17 | 84.01 | 8.17 | |
| USCNet (Ours) | Spider [75] | ICML'24 | 175 | 76.81 | 63.54 | 72.87 | 42.39 | 71.34 | 56.68 | 74.92 | 85.79 | 7.86 | |
| | USCNet (Ours) | - | 4.04 | 79.70 | 74.99 | 74.80 | 45.73 | 75.57 | 61.34 | 78.03 | 87.92 | 7.49 | |

表4. 在USC12K上训练后，误判现象基本被消除。左: SOD模型在COD数据集上的推理结果。VSCode使用SOD提示。右: COD模型在SOD数据集上的推理结果。VSCode使用COD提示。评估指标为 F_{β}^{ω} 。EV表示期望值(Expected Value)。

| SOD Models | COD Datasets | | EV | COD Models | SOD Datasets | | EV |
|-------------|--------------|--------|----|--------------|--------------|--------|----|
| | COD10K | NC4K | | | DUTS | HKU-IS | |
| ICON [77] | 0.0146 | 0.0834 | 0 | SINet-V2 [7] | 0.0708 | 0.0443 | 0 |
| F3Net [59] | 0.0129 | 0.0787 | 0 | PFNet [35] | 0.1152 | 0.0874 | 0 |
| VSCode [31] | 0.0097 | 0.0626 | 0 | VSCode [31] | 0.0537 | 0.0391 | 0 |

表5. 六种方法在SOD数据集(DUTS、HKU-IS)和COD数据集(NC4K、COD10K)上的泛化性能。更多结果见附录§4。

| Model | DUTS | | HKU-IS | | NC4K | | COD10K | |
|--------------|---|---|---|---|---|---|---|---|
| | $F_{\beta}^{\max} \uparrow E_{\phi}^m \uparrow$ |
| ICON [77] | .679 | .785 | .814 | .874 | .540 | .715 | .631 | .752 |
| F3Net [59] | .703 | .794 | .832 | .881 | .576 | .744 | .661 | .773 |
| SINet-V2 [7] | .732 | .821 | .838 | .884 | .609 | .763 | .662 | .769 |
| PFNet [35] | .691 | .790 | .818 | .876 | .556 | .730 | .660 | .769 |
| VSCode [31] | .724 | .812 | .834 | .885 | .626 | .787 | .684 | .783 |
| USCNet | .784 | .852 | .844 | .886 | .794 | .877 | .743 | .869 |

用预热和线性衰减策略。初始学习率设置为0.0001，批量大小为24，最大训练轮数为90。所有方法技术细节可见附录§5。

5.1. 定量评估

在USC12K上的性能. 表3给出了各对比模型在USC12K基准上的性能。将单属性场景(指场景A和场景B)中的模型结果与多属性场景(指场景C)中的结果进行比较，可以发现所有模型在场景C中的得分均低于场景A和场景B。这表明显著目标和伪装目标同时存在会增加模型同时识别两者的难度。我们的方法在所有场景中取得了更大的领先优势，证明当面对更具挑战性的场景时，我们的模型具有更强的适应能力。此外，USCNet在所有场景中均优于其他对比方法。其他指标的结果，包括AUC↑、SI-AUC↑、 F_m^{β} ↑、SI- F_m^{β} ↑、 F_{\max}^{β} ↑、SI- F_{\max}^{β} ↑、 E_m ↑，可见表3以及附录§3。

误检性能. 为探究使用USC12K数据集训练对模型误检率的影响，我们在COD和SOD数据集上测试SOD和COD模型。表4的结果表明，与表1相比，使用USC12K训练显著降低了模型的误检得分。

泛化性能. 我们在六个广泛使用的数据集上评估模型的性能，包括COD数据集COD10K、NC4K和CAMO-TE，以及SOD数据集DUT-TE、HKU-IS和DUT-OMRON。结果表明我们的模型具有更强的泛化能力。相关结果如表5所示，并可见于附录§4。

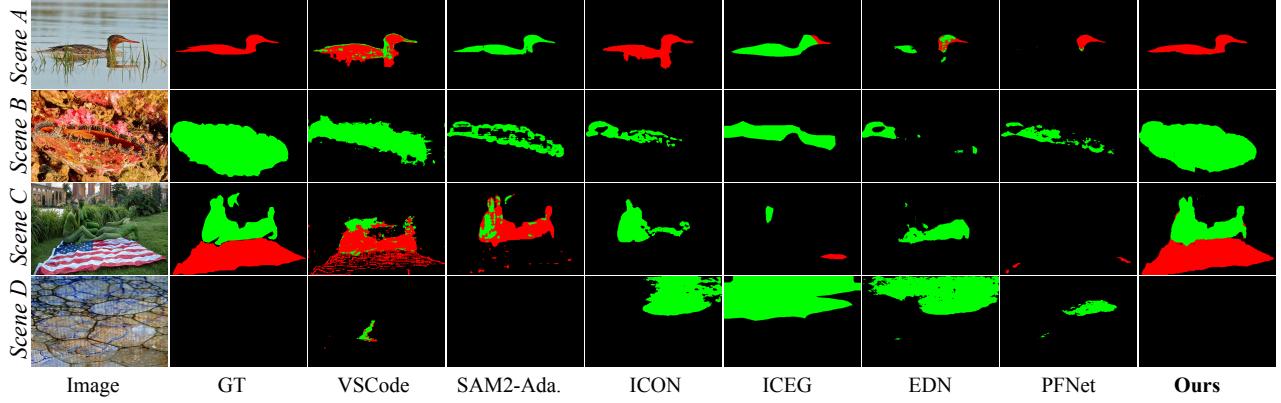


图 5. USCNet 与六种模型在所有场景下的定性比较。更多可视化结果可见附录 §7。

表 6. ARM 模块中不同组件的有效性。Intra-S.: Intra-SPQ。

| Encoder | Decoder | Intra-S. | Inter-S. | Q2I | I2Q | Para. | IoU_S | IoU_C | mIoU | mAcc | CSCS |
|---------|---------|----------|----------|-----|-----|-------|----------------|----------------|--------------|--------------|-------------|
| Frozen | Tuning | ✗ | ✗ | ✗ | ✗ | 4.22 | 66.42 | 44.02 | 68.78 | 77.65 | 11.58 |
| Tuning | Tuning | ✗ | ✗ | ✗ | ✗ | 4.36 | 71.42 | 56.71 | 74.98 | 84.74 | 9.12 |
| Tuning | Frozen | ✗ | ✓ | ✓ | ✓ | 3.44 | 71.68 | 57.53 | 75.31 | 85.15 | 9.07 |
| Tuning | Frozen | ✓ | ✗ | ✓ | ✓ | 4.03 | 74.32 | 58.91 | 76.96 | 85.80 | 7.98 |
| Tuning | Frozen | ✓ | ✓ | ✗ | ✗ | 0.75 | 70.97 | 56.56 | 74.77 | 84.43 | 9.85 |
| Tuning | Frozen | ✓ | ✓ | ✓ | ✗ | 2.40 | 73.08 | 58.45 | 76.73 | 85.63 | 8.52 |
| Tuning | Frozen | ✓ | ✓ | ✓ | ✓ | 4.04 | 75.57 | 61.34 | 78.03 | 87.92 | 7.49 |

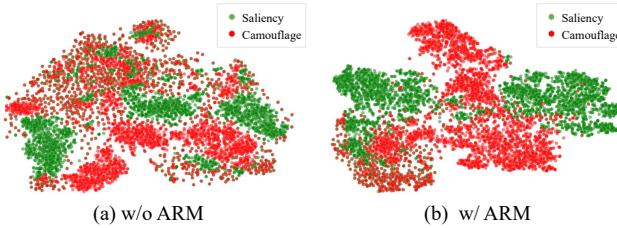


图 6. 我们模型在 USC12K 测试集图像中对显著目标与伪装目标的嵌入可视化，分别展示不使用与使用 ARM 模块的结果。维度通过 t-SNE [54] 降维。

5.2. 定性评估

在图 5 中，我们将模型的定性结果与六个对比模型进行了比较。我们的方法在所有场景中对显著目标和伪装目标表现出更优的检测能力。得益于 ARM 模块，我们的方法在图 5 的场景 C 中能够更好地区分同一图像中的显著目标和伪装目标。在场景 D 中，SOD 和 COD 方法在遇到背景时容易混淆，导致性能较差且鲁棒性不稳定，而我们的模型表现更优。

5.3. 消融实验

ARM 中不同组件的有效性. 如表 6 所示，对所有场景的消融实验验证了 ARM 模块各组件的有效性。从第 3、

4 和 7 行可以看出，Intra-SPQ 和 Inter-SPQ 均能提升模型性能，其中 Intra-SPQ 与 Inter-SPQ 联合使用时有更大的性能提升，达到最佳效果。第 5、6 和 7 行显示，Q2I 和 I2Q 也有助于区分显著与伪装目标，分别使 CSCS 降低 2.36% 和 1.08%。此外，相较于原始 SAM2 (见第 1 行) 和 SAM2-Adapter (见第 2 行)，通过引入 ARM 模块和冻结的 mask 解码器，所提出的 USCNet 在 USC12K 上以更高效的微调方式在所有指标上均取得更优表现。

ARM 对特征嵌入分离的影响. 我们通过可视化显著与伪装目标的嵌入进一步验证 ARM 的有效性。如图 6 所示，无 ARM 模块的嵌入分布较为混杂，显著目标与伪装目标的嵌入难以区分。而加入 ARM 模块的模型生成了更加清晰、聚类良好的表示，更好地分离了显著目标与伪装目标的嵌入。更多消融实验可见附录 §8。

6. 结论

我们分析了当前 SOD 和 COD 方法中显著与伪装目标的误检问题，并指出了现有数据集和模型的局限性。为推进无约束的显著与伪装目标检测，我们构建了一个大规模数据集 USC12K。基于此，我们设计了统一的 USCNet 流水线，显式建模样本间 (inter-sample) 与样本内 (intra-sample) 的特征关系。此外，我们提出了一种新的评价指标 CSCS，用于评估模型在显著与伪装目标识别上的混淆情况。大量实验表明，所构建的数据集能够缓解模型的误检问题，而我们的方法在 USC12K 基准上超越了现有相关模型，并在六个 SOD 与 COD 数据集上展现了更好的泛化能力。我们相信 USC12K 基准将推动 SOD 与 COD 领域的进一步研究，帮助模型更好地捕捉与人类视觉系统一致的显著和伪装模式。

7. 致谢

本工作得到湖北省自然科学基金资助（项目编号：2024AFB545）。计算工作在华中科技大学高性能计算平台（HPC Platform）完成。

参考文献

- [1] Tianrun Chen, Ankang Lu, Lanyun Zhu, Chaotao Ding, Chunnan Yu, Deyi Ji, Zejian Li, Lingyun Sun, Papa Mao, and Ying Zang. Sam2-adapter: Evaluating & adapting segment anything 2 in downstream tasks: Camouflage, shadow, medical image segmentation, and more. *arXiv preprint arXiv:2408.04579*, 2024. 3, 6, 7
- [2] Tianrun Chen, Lanyun Zhu, Chaotao Ding, Runlong Cao, Yan Wang, Shangzhan Zhang, Zejian Li, Lingyun Sun, Ying Zang, and Papa Mao. Sam-adapter: Adapting segment anything in underperformed scenes. In *ICCVW*, pages 3359–3367, 2023. 3, 4, 6, 7
- [3] Zhongxi Chen, Ke Sun, and Xianming Lin. Camodiffusion: Camouflaged object detection via conditional diffusion models. In *AAAI*, volume 38, pages 1272–1280, 2024. 6, 7
- [4] Ming-Ming Cheng, Niloy J. Mitra, Xiaolei Huang, Philip H. S. Torr, and Shi-Min Hu. Global contrast based salient region detection. *IEEE TPAMI*, 37(3):569–582, 2015. 4
- [5] Zijun Deng, Xiaowei Hu, Lei Zhu, Xuemiao Xu, Jing Qin, Guoqiang Han, and Pheng-Ann Heng. R3net: Recurrent residual refinement network for saliency detection. In *AAAI*, volume 684690, 2018. 3
- [6] 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. 4
- [7] Deng-Ping Fan, Ge-Peng Ji, Ming-Ming Cheng, and Ling Shao. Concealed object detection. *IEEE TPAMI*, 44(10):6024–6042, 2021. 1, 2, 3, 6, 7
- [8] Deng-Ping Fan, Ge-Peng Ji, Guolei Sun, Ming-Ming Cheng, Jianbing Shen, and Ling Shao. Camouflaged object detection. In *CVPR*, pages 2777–2787, 2020. 2, 3, 4
- [9] Deng-Ping Fan, Ge-Peng Ji, Peng Xu, Ming-Ming Cheng, Christos Sakaridis, and Luc Van Gool. Advances in deep concealed scene understanding. *Visual Intelligence*, 1(1):16, 2023. 2
- [10] Chaowei Fang, Haibin Tian, Dingwen Zhang, Qiang Zhang, Jungong Han, and Junwei Han. Densely nested top-down flows for salient object detection. *Science China Information Sciences*, 65(8):182103, 2022. 3
- [11] Shixuan Gao, Pingping Zhang, Tianyu Yan, and Huchuan Lu. Multi-scale and detail-enhanced segment anything model for salient object detection. *arXiv preprint arXiv:2408.04326*, 2024. 2, 3, 6, 7
- [12] Chunming He, Kai Li, Yachao Zhang, Longxiang Tang, Yulun Zhang, Zhenhua Guo, and Xiu Li. Camouflaged object detection with feature decomposition and edge reconstruction. In *CVPR*, pages 22046–22055, 2023. 3, 6, 7
- [13] Chunming He, Kai Li, Yachao Zhang, Yulun Zhang, Chenyu You, Zhenhua Guo, Xiu Li, Martin Danelljan, and Fisher Yu. Strategic preys make acute predators: Enhancing camouflaged object detectors by generating camouflaged objects. In *ICLR*, 2024. 6, 7
- [14] Qibin Hou, Ming-Ming Cheng, Xiaowei Hu, Ali Borji, Zhuowen Tu, and Philip HS Torr. Deeply supervised salient object detection with short connections. In *CVPR*, pages 3203–3212, 2017. 3
- [15] Xihang Hu, Xiaoli Zhang, Fasheng Wang, Jing Sun, and Fuming Sun. Efficient camouflaged object detection network based on global localization perception and local guidance refinement. *IEEE TCSVT*, 2024. 6, 7
- [16] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. In *ICCV*, pages 4015–4026, 2023. 3, 4
- [17] Trung-Nghia Le, Tam V Nguyen, Zhongliang Nie, Minh-Triet Tran, and Akihiro Sugimoto. Anabanch network for camouflaged object segmentation. *CVIU*, 184:45–56, 2019. 3, 4
- [18] Aixuan Li, Jing Zhang, Yunqiu Lv, Bowen Liu, Tong Zhang, and Yuchao Dai. Uncertainty-aware joint salient object and camouflaged object detection. In *CVPR*, pages 10071–10081, 2021. 1, 2, 3
- [19] Feiran Li, Qianqian Xu, Shilong Bao, Zhiyong Yang, Runmin Cong, Xiaochun Cao, and Qingming Huang. Size-invariance matters: Rethinking metrics and losses for imbalanced multi-object salient object detection. In *ICML*, pages 28989–29021, 2024. 6
- [20] Guanbin Li and Yizhou Yu. Visual saliency based on multi-scale deep features. In *CVPR*, 2015. 3, 4
- [21] Yin Li, Xiaodi Hou, Christof Koch, James M Rehg, and Alan L Yuille. The secrets of salient object segmentation. In *CVPR*, 2014. 4
- [22] Shijie Lian and Hua Li. Evaluation of segment anything

- model 2: The role of sam2 in the underwater environment. *arXiv preprint arXiv:2408.02924*, 2024. 3
- [23] Chiuhsiang Joe Lin and Yogi Tri Prasetyo. A metaheuristic-based approach to optimizing color design for military camouflage using particle swarm optimization. *Color Research & Application*, 44(5):740–748, 2019. 1
- [24] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *ICCV*, pages 2980–2988, 2017. 6
- [25] Nian Liu and Junwei Han. Dhsnet: Deep hierarchical saliency network for salient object detection. In *CVPR*, pages 678–686, 2016. 3
- [26] Nian Liu, Junwei Han, and Ming-Hsuan Yang. Picanet: Learning pixel-wise contextual attention for saliency detection. In *CVPR*, pages 3089–3098, 2018. 3
- [27] Nian Liu, Ni Zhang, Kaiyuan Wan, Ling Shao, and Junwei Han. Visual saliency transformer. In *ICCV*, pages 4722–4732, 2021. 3, 6, 7
- [28] Tie Liu, Zejian Yuan, Jian Sun, Jingdong Wang, Nanning Zheng, Xiaou Tang, and Heung-Yeung Shum. Learning to detect a salient object. *IEEE TPAMI*, 33(2):353–367, 2011. 4
- [29] Weihuang Liu, Xi Shen, Chi-Man Pun, and Xiaodong Cun. Explicit visual prompting for low-level structure segmentations. In *CVPR*, pages 19434–19445, 2023. 1, 3, 6, 7
- [30] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *CVPR*, pages 3431–3440, 2015. 6
- [31] Ziyang Luo, Nian Liu, Wangbo Zhao, Xuguang Yang, Dingwen Zhang, Deng-Ping Fan, Fahad Khan, and Junwei Han. Vscode: General visual salient and camouflaged object detection with 2d prompt learning. In *CVPR*, pages 17169–17180, 2024. 1, 2, 3, 6, 7
- [32] Yunqiu Lv, Jing Zhang, Yuchao Dai, Aixuan Li, Bowen Liu, Nick Barnes, and Deng-Ping Fan. Simultaneously localize, segment and rank the camouflaged objects. In *CVPR*, pages 11591–11601, 2021. 3, 4
- [33] Jun Ma, Yuting He, Feifei Li, Lin Han, Chenyu You, and Bo Wang. Segment anything in medical images. *Nature Communications*, 15(1):654, 2024. 3
- [34] Ran Margolin, Lihia Zelnik-Manor, and Ayelet Tal. How to evaluate foreground maps? In *CVPR*, pages 248–255, 2014. 2
- [35] Haiyang Mei, Ge-Peng Ji, Ziqi Wei, Xin Yang, Xiaopeng Wei, and Deng-Ping Fan. Camouflaged object segmentation with distraction mining. In *CVPR*, pages 8772–8781, 2021. 1, 3, 6, 7
- [36] Vida Movahedi and James H Elder. Design and perceptual validation of performance measures for salient object segmentation. In *CVPRW*, 2010. 4
- [37] Youwei Pang, Xiaoqi Zhao, Tian-Zhu Xiang, Lihe Zhang, and Huchuan Lu. Zoom in and out: A mixed-scale triplet network for camouflaged object detection. In *CVPR*, pages 2160–2170, 2022. 3, 6, 7
- [38] Youwei Pang, Xiaoqi Zhao, Lihe Zhang, and Huchuan Lu. Multi-scale interactive network for salient object detection. In *CVPR*, pages 9413–9422, 2020. 3
- [39] Jialun Pei, Tianyang Cheng, Deng-Ping Fan, He Tang, Chuanbo Chen, and Luc Van Gool. Osformer: One-stage camouflaged instance segmentation with transformers. In *ECCV*, pages 19–37, 2022. 3
- [40] Jialun Pei, Tianyang Cheng, He Tang, and Chuanbo Chen. Transformer-based efficient salient instance segmentation networks with orientative query. *IEEE TMM*, 25:1964–1978, 2022. 3
- [41] Jialun Pei, Tao Jiang, He Tang, Nian Liu, Yueming Jin, Deng-Ping Fan, and Pheng-Ann Heng. Calibnet: Dual-branch cross-modal calibration for rgb-d salient instance segmentation. *IEEE TIP*, 2024. 3
- [42] Jialun Pei, Zhangjun Zhou, Diandian Guo, Zhixi Li, Jing Qin, Bo Du, and Pheng-Ann Heng. Synergistic bleeding region and point detection in laparoscopic surgical videos. *arXiv preprint arXiv:2503.22174*, 2025. 3
- [43] Jialun Pei, Zhangjun Zhou, and Tiantian Zhang. Evaluation study on sam 2 for class-agnostic instance-level segmentation. *arXiv preprint arXiv:2409.02567*, 2024. 3
- [44] Lintao Peng, Chunli Zhu, and Liheng Bian. U-shape transformer for underwater image enhancement. *IEEE TIP*, 2023. 3
- [45] Yongri Piao, Wei Ji, Jingjing Li, Miao Zhang, and Huchuan Lu. Depth-induced multi-scale recurrent attention network for saliency detection. In *ICCV*, pages 7254–7263, 2019. 3
- [46] Michael I Posner, Steven E Petersen, et al. The attention system of the human brain. *Annual review of neuroscience*, 13(1):25–42, 1990. 1
- [47] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *ICML*, pages 8748–8763, 2021. 4
- [48] Nikhila Ravi, Valentin Gabeur, Yuan-Ting Hu, Ronghang Hu, Chaitanya Ryali, Tengyu Ma, Haitham Khedr, Roman

- Rädle, Chloe Rolland, Laura Gustafson, et al. Sam 2: Segment anything in images and videos. In *ICLR*, 2025. 3
- [49] Jingjing Ren, Xiaowei Hu, Lei Zhu, Xuemiao Xu, Yangyang Xu, Weiming Wang, Zijun Deng, and Pheng-Ann Heng. Deep texture-aware features for camouflaged object detection. *IEEE TCSVT*, 33(3):1157–1167, 2021. 3
- [50] Przemysław Skurowski, Hassan Abdulameer, Jakub Błaszczyk, Tomasz Depta, Adam Kornacki, and Przemysław Kozięć. Animal camouflage analysis: Chameleon database. *Unpublished Manuscript*, 2018. 4
- [51] Martin Stevens and Sami Merilaita. Animal camouflage: current issues and new perspectives. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1516):423–427, 2009. 1
- [52] Yujia Sun, Shuo Wang, Chenglizhao Chen, and Tian-Zhu Xiang. Boundary-guided camouflaged object detection. *arXiv preprint arXiv:2207.00794*, 2022. 3
- [53] Longxiang Tang, Kai Li, Chunming He, Yulun Zhang, and Xiu Li. Source-free domain adaptive fundus image segmentation with class-balanced mean teacher. In *MICCAI*, pages 684–694, 2023. 1
- [54] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(11), 2008. 8
- [55] Lijun Wang, Huchuan Lu, Yifan Wang, Mengyang Feng, Dong Wang, Baocai Yin, and Xiang Ruan. Learning to detect salient objects with image-level supervision. In *CVPR*, 2017. 2, 3, 4
- [56] Linzhao Wang, Lijun Wang, Huchuan Lu, Pingping Zhang, and Xiang Ruan. Salient object detection with recurrent fully convolutional networks. *IEEE TPAMI*, 41(7):1734–1746, 2018. 3
- [57] Rui Wang, Caijuan Shi, Changyu Duan, Weixiang Gao, Hongli Zhu, Yunchao Wei, and Meiqin Liu. Camouflaged object segmentation with prior via two-stage training. *CVIU*, 246:104061, 2024. 6, 7
- [58] Tiantian Wang, Ali Borji, Lihe Zhang, Pingping Zhang, and Huchuan Lu. A stagewise refinement model for detecting salient objects in images. In *ICCV*, pages 4019–4028, 2017. 3
- [59] Jun Wei, Shuhui Wang, and Qingming Huang. F³ net: fusion, feedback and focus for salient object detection. In *AAAI*, volume 34, pages 12321–12328, 2020. 1, 6, 7
- [60] Yu-Huan Wu, Yun Liu, Le Zhang, Ming-Ming Cheng, and Bo Ren. Edn: Salient object detection via extremely-downsampled network. *IEEE TIP*, 31:3125–3136, 2022. 6, 7
- [61] Yongqin Xian, Christoph H Lampert, Bernt Schiele, and Zeynep Akata. Zero-shot learning—a comprehensive evaluation of the good, the bad and the ugly. *IEEE TPAMI*, 41(9):2251–2265, 2018. 3
- [62] Chenxi Xie, Changqun Xia, Mingcan Ma, Zhirui Zhao, Xiaowu Chen, and Jia Li. Pyramid grafting network for one-stage high resolution saliency detection. In *CVPR*, pages 11717–11726, 2022. 6, 7
- [63] Enze Xie, Wenhui Wang, Zhiding Yu, Anima Anandkumar, Jose M Alvarez, and Ping Luo. Segformer: Simple and efficient design for semantic segmentation with transformers. *NeurIPS*, 34:12077–12090, 2021. 6
- [64] Yunyang Xiong, Bala Varadarajan, Lemeng Wu, Xiaoyu Xiang, Fanyi Xiao, Chenchen Zhu, Xiaoliang Dai, Dilin Wang, Fei Sun, Forrest Iandola, et al. Efficientsam: Leveraged masked image pretraining for efficient segment anything. *arXiv preprint arXiv:2312.00863*, 2023. 3
- [65] Qiong Yan, Li Xu, Jianping Shi, and Jiaya Jia. Hierarchical saliency detection. In *CVPR*, 2013. 4
- [66] Zhiling Yan, Weixiang Sun, Rong Zhou, Zhengqing Yuan, Kai Zhang, Yiwei Li, Tianming Liu, Quanzheng Li, Xiang Li, Lifang He, et al. Biomedical sam 2: Segment anything in biomedical images and videos. *arXiv preprint arXiv:2408.03286*, 2024. 3
- [67] Chuan Yang, Lihe Zhang, Huchuan Lu, Xiang Ruan, and Ming-Hsuan Yang. Saliency detection via graph-based manifold ranking. In *CVPR*, 2013. 4
- [68] Fan Yang, Qiang Zhai, Xin Li, Rui Huang, Ao Luo, Hong Cheng, and Deng-Ping Fan. Uncertainty-guided transformer reasoning for camouflaged object detection. In *ICCV*, pages 4146–4155, 2021. 3
- [69] Bowen Yin, Xuying Zhang, Deng-Ping Fan, Shaohui Jiao, Ming-Ming Cheng, Luc Van Gool, and Qibin Hou. Camoformer: Masked separable attention for camouflaged object detection. *IEEE TPAMI*, 2024. 6, 7
- [70] Qiang Zhai, Xin Li, Fan Yang, Chenglizhao Chen, Hong Cheng, and Deng-Ping Fan. Mutual graph learning for camouflaged object detection. In *CVPR*, pages 12997–13007, 2021. 3
- [71] Wei Zhai, Yang Cao, Jing Zhang, and Zheng-Jun Zha. Exploring figure-ground assignment mechanism in perceptual organization. *NeurIPS*, 35:17030–17042, 2022. 3
- [72] Miao Zhang, Tingwei Liu, Yongri Piao, Shunyu Yao, and

- Huchuan Lu. Auto-msfnet: Search multi-scale fusion network for salient object detection. In *ACM MM*, pages 667–676, 2021. [3](#), [6](#), [7](#)
- [73] Miao Zhang, Shuang Xu, Yongri Piao, Dongxiang Shi, Shusen Lin, and Huchuan Lu. Preynet: Preying on camouflaged objects. In *ACM MM*, pages 5323–5332, 2022. [3](#)
- [74] Xiaoning Zhang, Tiantian Wang, Jinqing Qi, Huchuan Lu, and Gang Wang. Progressive attention guided recurrent network for salient object detection. In *CVPR*, pages 714–722, 2018. [3](#)
- [75] Xiaoqi Zhao, Youwei Pang, Wei Ji, Baicheng Sheng, Jiaming Zuo, Lihe Zhang, and Huchuan Lu. Spider: A unified framework for context-dependent concept segmentation. In *ICML*, 2024. [1](#), [3](#), [6](#), [7](#)
- [76] Xiaoqi Zhao, Youwei Pang, Lihe Zhang, Huchuan Lu, and Lei Zhang. Suppress and balance: A simple gated network for salient object detection. In *ECCV*, pages 35–51, 2020. [6](#), [7](#)
- [77] Mingchen Zhuge, Deng-Ping Fan, Nian Liu, Dingwen Zhang, Dong Xu, and Ling Shao. Salient object detection via integrity learning. *IEEE TPAMI*, 45(3):3738–3752, 2022. [1](#), [2](#), [3](#), [6](#), [7](#)