伪装物体检测

范登平 [1,2] 季葛鹏 [3] 孙国磊 [4] 程明明 [2] 沈建冰 [1] * 邵岭 [1]
[1] 阿联酋起源人工智能研究院（IIAI） [2] 南开大学计算机学院
[3] 武汉大学计算机学院 [4] 瑞士苏黎世联邦理工学院

http://dpfan.net/Camouflage/

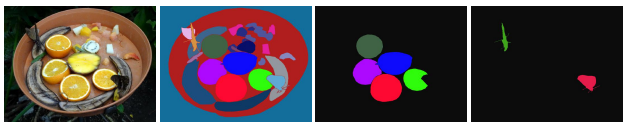
图 1: *COD10K* 数据集示例图。你能找到隐藏在图片中的伪装物体吗？**以彩色电子版阅览时视觉效果最佳**。答案见补充材料。

摘要

伪装物体检测（***C**amouflaged **O**bject **D**etection*, ***COD***），顾名思义，旨在识别“无缝”嵌入其周围环境的物体，本文对这项新任务展开了全面的研究。与传统的物体检测相比，通常伪装物体与其背景之间具有高度相似性，因此伪装物体检测更具挑战。为解决这一问题，本文精心构建了 ***COD10K*** 数据集，它包含了 *10,000* 张图像，且涵盖了各种自然场景，具有超过 *78* 个类别的伪装物体。所有的图像都进行了稠密的标注，包括类别、包围盒、对象级/实例级，以及抠图级的标签。*COD10K* 数据集可以助力许多视觉任务，例如目标定位、图像分割和抠图技术等。同时，本文也为伪装物体检测任务提供了一个简单且有效的框架，称为搜索识别网络（***S**earch **I**dentification **N**etwork*, ***SINet***）。没有借助过多技巧，*SINet* 在所有数据集上的表现均优于其它先进的物体检测基准模型。因此，*SINet* 是一个鲁棒的、通用的架构，这有助于促进伪装物体检测的发展。最后，通过对 *13* 种最先进模型进行系统评估，本文给出了许多有趣的发现并且展示了一些伪装物体检测的潜在应用。希望本文的研究能为这一新领域的学者提供更多探索机会。详见：*https://github.com/DengPingFan/SINet/*。

1. 引言

你能在图. 1中找到伪装的物体吗？生物学家将这类伪装方式称为背景匹配 [9]，即动物为避免被识别 [49]，会尝试改变其自身颜色以“完美”地融入周围环境。感官生态学家的研究表明，这种伪装策略是通过欺骗观察者的视觉感知系统而产生的 [58]。因此，解决伪装物体检测（**C**amouflaged **O**bject **D**etection **COD**）任务需要大量的视觉感知知识 [61]。如图. 2所示，物体物与背景之间高度的相似性使 COD 远比传统的显著物体检测（SOD）[1, 5, 18, 26, 63–67, 69] 或通用物体检测（GOD）[4, 80] 更具挑战性。



（a）原图 （b）通用物体 （c）显著物体 （d）伪装物体

图 2: 给定一张输入图像(a)，(b)为全景分割 [31] 的真值（全景分割检测中的**通用**物体 [40, 45] 包括 stuff 和 things），(c)是 **显著**实例/物体检测 [17, 34, 62, 77]（检测最吸引人注意力的目标），(d) 是本文提出的**伪装**物体检测任务，即检测出与周围环境具有相似模式（例如边缘，纹理或颜色）的物体。如图所示，两只蝴蝶的边缘与香蕉融合在一起，难以识别。

除了其学术价值外，伪装物体检测还有助于推动下列领域的实际应用：计算觉领域（可用于搜索和救援工作，或寻找稀有物种）、医学图像分析领域（如息肉分割 [15] 和肺炎分割 [19, 68]）、农业领域（如蝗虫入侵监控）和艺术领域（用于真实感图像融合 [22] 或艺术消遣 [6]）。

目前，由于缺乏规模足够大的数据集，伪装物体检测的研究还不够深入。为了对 COD 课题进行

* 本文为 CVPR2020 论文 [14] 的中文翻译版。

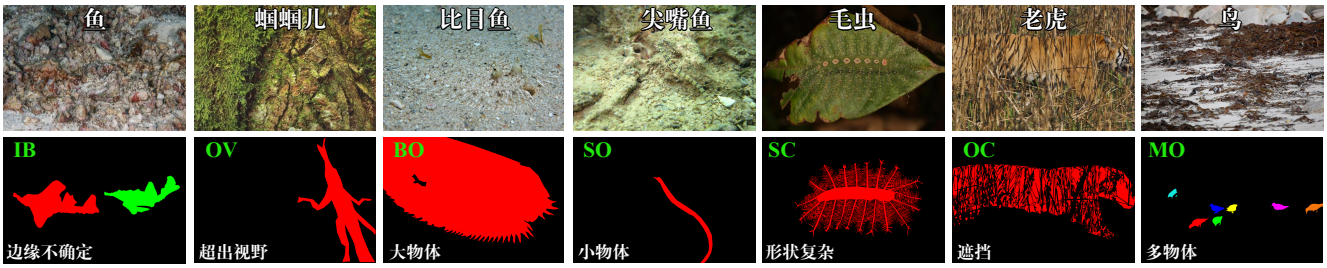


图 3: *COD10K* 数据集中各式各样的具有挑战性的属性示例图。属性表详见表. 2。以彩色电子版阅览视觉效果最佳。

全面的研究，本文做出了两项贡献。首先，本文专门为 COD 任务精心构建了 ***COD10K*** 数据集。它与现有的数据集有以下方面的区别：

- ***COD10K*** 数据集包含了 1 万张图像，涵盖了 78 种伪装物体类别，属于水生、飞行、两栖和陆地等。
- 所有的伪装图像都赋予了不同的层级标签，如类别、包围盒、对象级和实例级。这些标签会使得许多视觉任务受益，如目标定位，似物性检测，语义边缘检测 [43]，任务迁移学习 [70] 等。
- 每张伪装图像都加上了真实环境中遇到的具有挑战的属性以及抠图级 [74]（标注一张图像耗时约 60 分钟）的标签。这些高质量的标注有助于对算法性能进行更深入的分析。

其次，使用本文构建的 *COD10K* 数据集和两个现有数据集 [33, 57]，来共同构成最大的伪装物体检测训练数据集，对 13 种最先进 (SOTA) 的基准模型 [3, 24, 28, 33, 36, 39, 41, 52, 69, 76, 78, 79, 83] 进行严格的评估。本文的评估成为了目前最大规模的 COD 研究。此外，本文还提供了一个简单而有效的框架，名为 ***SINet*** (**S**earch **I**dentification **Net**)。值得注意的是，*SINet* 的训练总时长仅为 1 小时左右，并且在目前所有 COD 数据集上都达到了 SOTA 的性能。这表明 *SINet* 可能是解决 COD 问题的潜在方案。在深度学习时代下，本文是第一个完整的 COD 任务评测，同时以伪装的角度去重新理解物体检测任务。

2. 相关工作

根据 [80] 的研究，物体大致可以分为三类：通用物体、显著物体和伪装物体。接下来本文将逐一介绍这些检测策略。

2.1. 通用物体和显著物体检测

通用物体检测（**G**eneric **O**bject **D**etection, **COD**）. 通用物体检测是计算机视觉中最热门的方向之一 [11, 31, 38, 56]。值得注意的是，通用物体既可以是显著的也可以是伪装的；伪装物体可视为通用物体中较难的特例（如图. 9中第二行和第三行）。典型的 GOD 任务包括语义分割和全景分割（见图. 2 b）。

数据集	年份	数量	类	Att.	BBox.	Ml.	Ins.	Cate.	Spi.	Obj.
CHAMELEON [57]	2018	76	-							✓
CAMO [33]	2019	2.5*k*	8	✓					✓	✓
COD10K **(Ours)**	2020	**10***k*	78	✓	✓	✓	✓	✓	✓	✓

表 1: 从 COD 数据集对比信息来看，*COD10K* 数据集提供了更丰富的注释标签。数量（Img.）：图片数。类（Cls.）：类别。Att.：属性。BBox.；包围盒。Ml.：抠图 [74] 级标注（图. 7）。Ins.：实例级标签。Cate.：类别标签。Spi.：显式拆分训练集与测试集。Obj.：物体。

显著物体检测 (**S**alient **O**bject **D**etection, **SOD**). SOD 旨在识别图像中最引人注意的物体，然后对其轮廓进行像素级分割 [29, 39, 73, 78]。虽然“显著”一词本质上与“伪装”相反（突出与浸入），但是显著物体也可以为伪装物体检测提供重要信息，即可以把包含显著对象的图像作为伪装物体检测的负样本。

2.2. 伪装物体检测

伪装物体检测在生物学、艺术、医学等领域有着悠久而丰富的历史，对提高人类的视觉感知能力影响巨大。Abbott Thayer [59] 和 Hugh Cott [8] 的关于伪装动物的两项杰出研究至今仍然影响广泛。感兴趣的读者可以仔细阅读 Stevens 等人 [58] 关于这段历史的描述。

数据集. CHAMELEON [57] 是一个未经同行评议的数据集，仅包含 76 张图像，手工标注了对象级的真值图（GTs）。这些图像是以“伪装的动物”为关键字，通过谷歌搜索引擎收集的。另一个同期数据集称为 CAMO [33]，它具有 2500 张图像 (其中 2000 张用于训练，500 张用于测试)，涵盖了八个类别。它的两个子集 CAMO 和 MS-COCO，分别包含 1250 张图像。

不同于现有数据集，*COD10K* 数据集旨在提供

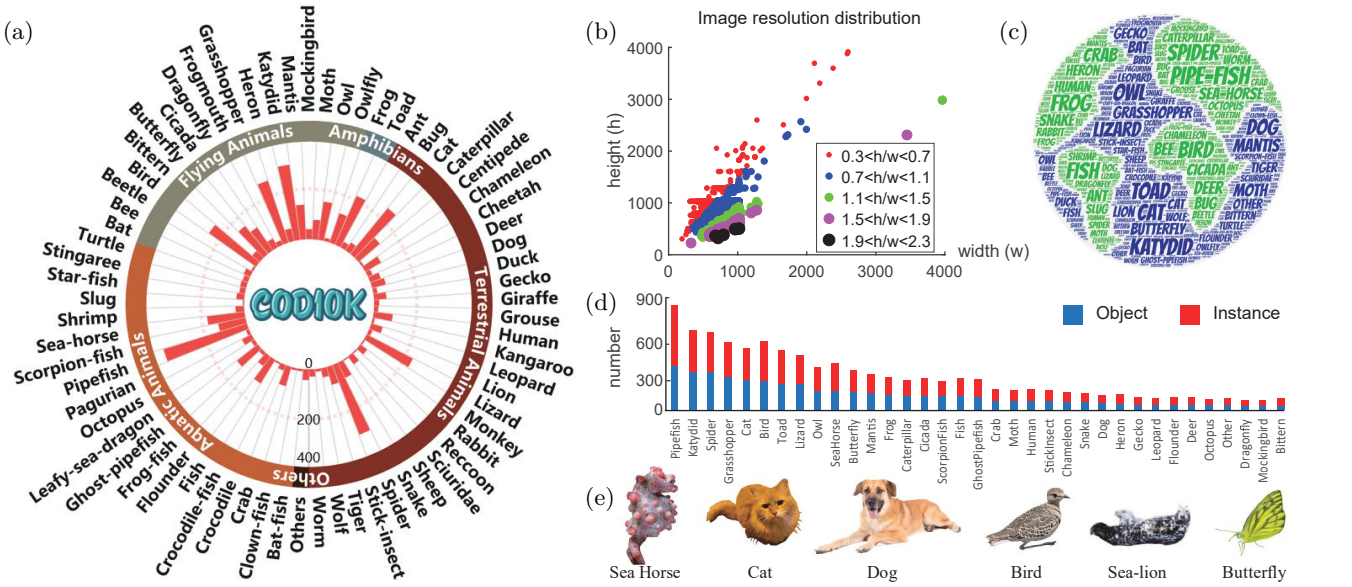


图 4: *COD10K* 数据集的统计信息和伪装类别示例。(a) 分类系统及其直方图分布。(b) 图像分辨率分布。(c) 词云分布。(d) 对象/实例的类别数量。(e) 子类示例。

一个更具挑战性，质量更高并且稠密标注的数据集。*COD10K* 是迄今为止最大的伪装物体检测数据集，包含有 1 万张图像（6000 张用于训练，4000 张用于测试）。详见表. 1。

伪装的类型. 伪装图像大致可分为两类，自然伪装和人工伪装的图像。动物（例如昆虫、头足类动物）的自然伪装是一项生存技能，用来防止被捕食者识别。相比之下，人工伪装通常被应用于视频监控环境；也会出现在产品的生产过程中（即产品缺陷检测）；或者在游戏或艺术中用来隐藏信息。

COD 定义. 和类别相关的语义分割任务不同，COD 任务与类别无关。因此，COD 任务简单且容易定义。输入一张图像，通过伪装物体检测方法，为每个像素 i 分配置信度 $p_i \in [0,1]$，其中 p_i 表示像素 i 的概率值。0 表示此像素不属于伪装物体，而 1 则表示完全属于伪装物体。本文主要研究对象级的 COD 任务，实例级 COD 作为未来的工作。

评估指标. 平均绝对误差（Mean absolute error，MAE）已被广泛用于 SOD 任务。和 Perazzi 等人 [50] 一样，本文也使用了平均绝对误差 M 指标（MAE(M)）来评估预测图像 C 和真值图像 G 之间的像素级精度。尽管 MAE 指标能有效地评估错误的出现和数量，但它不能评估错误出现的位置。最近，范等人提出了一种基于人类视觉感知机制的增强-匹配评价指标（E-measure，E_ϕ）[13]，它同时考虑了像素级信息的匹配和图像级信息的统计。该评价指标显然适用于评估伪装物体检测结果的整体和局部精度。由于伪装物体通常包含复杂的形状，COD 还需要一个可以判断结构相似性的评价指标，因此本文引入了 S 测评法（S-measure，S_α）[12]。最新研究 [12, 13] 表明，加权的 F 测评法（F-measure，F_β^w）[44] 比传统的 F_β 评测法更可靠。因此，本文也在 COD 领域中使用这一评估指标。

3. 本文数据集

新任务和数据集 [7, 11, 37, 48, 82] 的出现促进了计算机视觉各个领域的快速发展。例如，ImageNet [53] 数据集的出现使得深度模型在视觉识别任务上成为了可能。本文构建 COD 数据集的初衷与之相似，是为了：(1) 提供一项新的具有挑战性的任务，(2) 促进对新主题的研究，(3) 激发新想法。图. 1、图. 3以及图. 4 (e) 为 *COD10K* 的示例。接下来，本文从以下三个方面详细阐述 *COD10K* 数据集。数据集详情请点击此处。

3.1. 图像收集

根据 [18, 51] 的研究，数据集的标注质量和规模决定了基准数据集的时间长短。为了延长其寿命，*COD10K* 数据集包含 10,000 张图像（5,066 张伪装图像，3,000 张背景图像，1,934 张无伪装的图像），

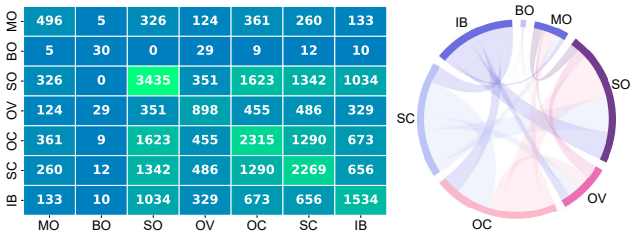


图 5: 左图：*COD10K* 数据集的共生属性分布。网格中的数字为图像总数。右图：属性间的相互依赖关系。弧长越长则两属性间的相关性就越高。

属性	描述
MO	多物体 图像至少包含两个物体
BO	大物体 物体面积和图像面积的比值大于等于 0.5。
SO	小物体 物体面积和图像面积的比值小于等于 0.1。
OV	超出视野 物体的部分区域超出了图像边界。
OC	遮挡 物体被部分遮挡。
SC	形状复杂 物体包含细小部分（如：动物脚）.
IB	边缘不确定 目标周围区域的前景和背景具有相似的颜色（与 RGB 直方图 χ^2 的距离 τ_{gc} 小于 0.9）

表 2: 属性描述 (示例见图. 3)。

包含 10 个超类和 78 个子类（69 个伪装类，9 个非伪装类），这些图像来源于多个摄影学网站。

大多数伪装图像来自 Flickr 网站，搜索关键字为：*camouflaged animal*, *unnoticeable animal*, *camouflaged fish*, *camouflaged butterfly*, *hidden wolf spider*, *walking stick*, *dead-leaf mantis*, *bird*, *sea horse*, *cat*, *pygmy seahorses* 等（见 图. 4 e），这些图像仅应用于学术研究。其它伪装图像（约 200 张）来自下列网站，如：Visualhunt，Pixabay，Unsplash，Free-images 等，这些图片不受版权和付费约束。为了避免选择偏见 [18]，我们还从 Flickr 收集了 3,000 张显著图像。为了进一步丰富负样本，又从互联网上下载了 1,934 张无伪装的图像，包括森林、雪地、草地、天空、海水和其他类别的场景。有关图像选择方案的更多信息，请见周等人的研究 [81]。

3.2. 专业标注

最新公布的数据集 [10, 16, 17] 表明，在创建大规模数据集时，建立分类系统至关重要。受 [46] 启发，本文进行了分层标注（采用众包方式标注）（类别 ↣ 包围盒 ↣ 属性 ↣ 对象/实例）。

- 类别. 如图. 4 (a) 所示，首先创建五个超类。然后从收集到的数据中归纳了 69 个最常见的子类。最后，标注每个图像的子类和超类。如果候选图像不属于任何已有类别，则将其归为“其他”类。
- 包围盒. 为了将 *COD10K* 数据集扩展到伪装

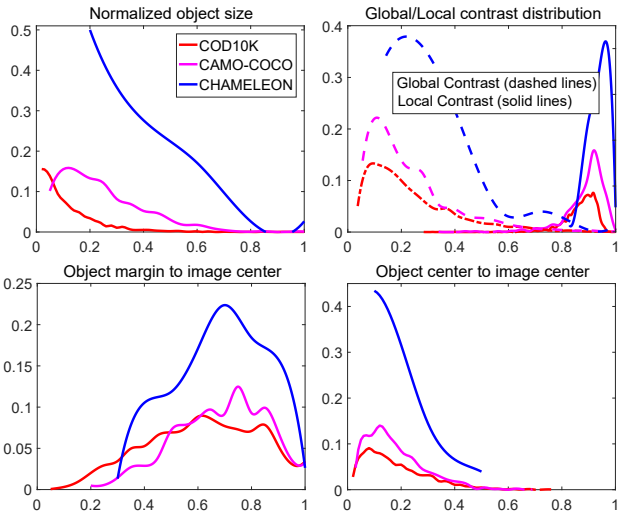


图 6: *COD10K* 数据集与现有 COD 数据集相比，它包含小物体（左上），包含更复杂的伪装（右上）并且受中心偏差影响更小（左下/右下）。

物采样（proposal）任务，本文还细致地为每张图像标注包围盒。

- 属性. 与文献 [18, 51] 保持一致，本文采用自然场景中常见的极具挑战的属性 (如遮挡、边缘不确定)，来标注每张图像。表. 2提供了属性描述，图. 5展示了共生属性分布情况。
- 对象/实例. 现有 COD 数据集仅提供了对象级标注（表. 1）。然而，对于计算机视觉研究人员而言，由于需要编辑和理解场景，将对象解析为实例更为重要。因此，与 COCO [37] 数据集类似，本文对数据集进行了实例级标注，共获得 5,069 张对象级的真值图像和 5,930 张实例级的真值图像。

3.3. 数据集特点与统计信息

- 物体大小. 根据 [18] 的研究，在图. 6（左上角）中绘制了归一化的物体尺寸，分布从 0.01%到 80.74%(平均：8.94%)。与 CAMO-COCO、CPD1K 和 CHAMELEON 相比，本文数据集具有更广的尺寸范围。
- 全局/局部对比度. 为了考察物体是否容易检测，本文使用全局与局部比值评估方法 [35]。图. 6(右上角) 表明了 *COD10K* 数据集比其他数据集更具挑战性。
- 中心偏见. 由于人类会自然地将注意力集中

图 7: 高质量的抠图级标注 [74]。

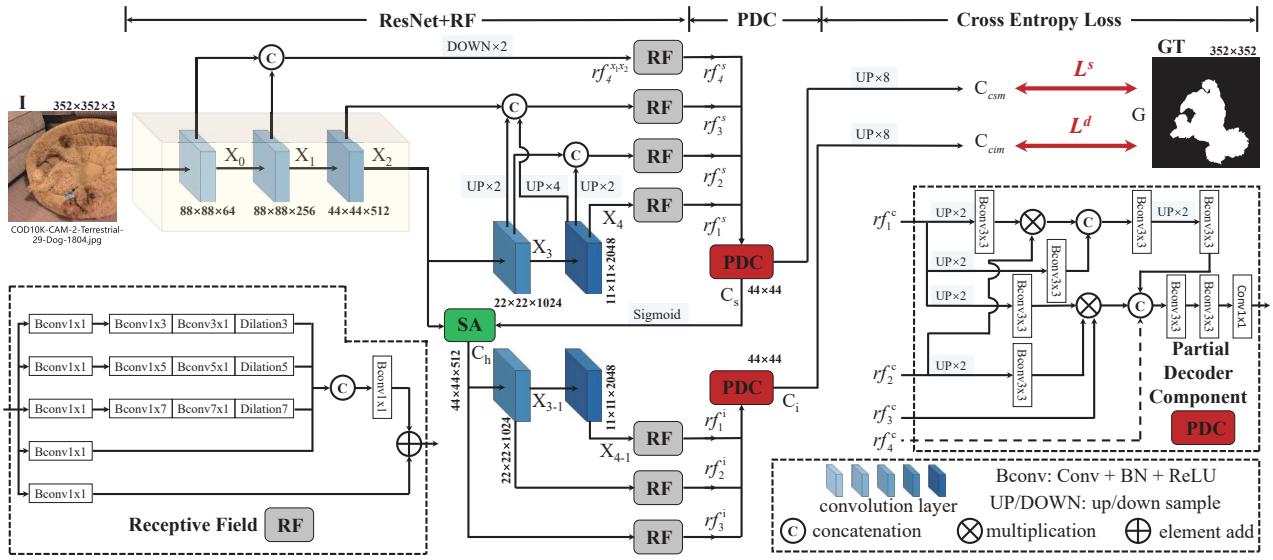


图 8: *SINet* 框架概要。本框架主要包括两个部分：感受野 (RF) 和部分解码组件 (PDC). RF 模块用来模拟人类视觉系统中的感受野结构。PDC 模块再现了动物捕食的搜索和识别阶段。SA 为 [69] 中描述的搜索注意函数。详见 § 4。

在场景中心，因此摄影时就会产生这样的偏见。本文采用 [18] 中的策略来分析这种偏见。图. 6（底部）表明了本文数据集受到的中心偏见的影响更小。

• *质量控制*. 为了确保高质量的标注，标记过程邀请了三位观察者参与，并进行 10 组的交叉验证。图. 7展示了通过和拒绝的标注示例。每张图像的实例级标注平均耗时 60 分钟。

• *超类和子类的分布*. *COD10K* 包括 5 个超类（陆地生物、空中生物、水生生物、两栖动物以及其他类）和 69 个子类（例如：蝙蝠鱼，狮子，蝙蝠，青蛙等），各类别的词云和对象/实例数量的示例分别显示在图. 4 c 和 d 中。

• *分辨率分布*. [71] 研究表明，高分辨率图像可以为模型训练提供更多物体边缘的细节从而在测试时获得更好的性能。图. 4 （b）展示了 *COD10K* 数据集图像分辨率的分布情况，它含有大量的 1080p 全高清图像。

• *数据集划分*. 为了给深度学习模型提供大量的训练数据，本文将 *COD10K* 数据集划分为：6,000 张图像的训练集和 4,000 张图像的测试集，图像分别从每个子类中随机选择。

4. 本文框架

动机. 生物学的研究 [23] 表明，捕食者在狩猎时，首先会判断是否存在潜在猎物，即*搜索猎物*；然后，目标动物被*识别*；最后，动物被*捕获*。

概述. 本文的 *SINet* 框架受到狩猎过程的前两阶段的启发。框架主要包括两个模块：搜索模块 (Search Module, SM) 和识别模块 (Identification Module, IM)。前者 (§ 4.1) 负责搜索被伪装的物体，而后者 (§ 4.2) 则用于精确检测物体。

4.1. 搜索模块（Search Module，SM）

神经科学实验已证实，在人类视觉系统中，一组多尺度的群感受野（population Receptive Fields, pRFs）有助于让靠近视网膜中央凹的区域更加显著，而该区域对小的空间位移非常敏感 [42]。于是，本文在搜索阶段（通常是在较小的、局部空间中）使用 RF [42, 69] 模块来整合更具鉴别性的特征表示。具体而言，对于输入图像 $\mathbf{I} \in \mathbb{R}^{W \times H \times 3}$，可利用 ResNet-50 [25] 模型提取出一组特征 $\{\mathcal{X}_k\}_{k=0}^{4}$。为了保留更多信息，本文将第二层特征层的步长参数设置为 1，使其和输入图像具有相同的分辨率。因此，每一层分辨率为 $\{[\frac{H}{k}, \frac{W}{k}], k = 4, 4, 8, 16, 32\}$ 。

最新研究 [79] 显示，更浅的卷积层中的低级特征保留了用于构建物体边缘的空间信息，而深层的深层卷积层的特征保留了用于定位目标的语义信息。由于神经网络本身的固有的特性，本文将提取的特征进行分层：低层 $\{\mathcal{X}_0, \mathcal{X}_1\}$，中层 $\mathcal{X}_2$ 和高层 $\{\mathcal{X}_3, \mathcal{X}_4\}$，并通过拼接、上采样和下采样等操作进行组合。与 [79] 不同，本文的 *SINet* 采用稠密连接策略 [27] 来保存来自不同特征层的更多信息，然后使用改进的 RF [42] 组件来扩大感受野。例如，先使用拼接操作来融合低级特征 $\{\mathcal{X}_0, \mathcal{X}_1\}$，然后将分辨率下采样为原始一半。再将融合后的新特征 rf_4^{x1x2} 进一步输入到 RF 组件生成 rf_4^s 特征。如图. 8所示，在

组合了三个层次的特征之后，得到了一组增强的特征 $\{rf_k^s, k=1,2,3,4\}$，用于鲁棒地学习伪装线索。

感受野 (Receptive Field, RF). RF 模块包括五个分支 $\{b_k, k=1,\ldots,5\}$。在每个分支中，第一个卷积层 (Bconv) 的尺寸为 1×1，用以将通道数降为 32。其后两层分别为：$(2k-1)\times(2k-1)$ Bconv 层和 3×3 Bconv 层。当 $k>2$ 时，空洞卷积率设置为 $(2k-1)$。前四个分支串联后，通过 1×1 Bconv 操作，其通道数降为 32。最后，加入第 5 个分支，并整体输入进 ReLU 函数以获得特征 rf_k。

4.2. 识别模块 (（Identification Module, IM)

通过之前的搜索模块获取到侯选特征后，在识别模块中，需要对伪装物体进行精确检测。本文采用密集连接方式对部分解码组件 (Partial Decoder Component, PDC) [69] 进行了扩展。具体来讲，PDC 整合了来自 SM 的四个特征层。可通过以下公式来计算粗糙的伪装图 C_s：

$$C_s = PD_s(rf_1^s, rf_2^s, rf_3^s, rf_4^s), \tag{1}$$

其中 $\{rf_k^s = rf_k, k=1,2,3,4\}$。现有文献 [41,69] 已表明，注意力机制可以有效地消除无关特征的干扰。因此引入*搜索注意力 (Search Attention, SA)* 模块来增强中间特征层 $\mathcal{X}_2$ 并获得增强的伪装图 C_h：

$$C_h = f_{max}(g(\mathcal{X}_2, \sigma, \lambda), C_s), \tag{2}$$

其中 $g(\cdot)$ 是 SA 函数，即为典型的归一化后的高斯滤波器，其标准差为：$\sigma=32$，核尺寸为：$\lambda=4$，$f_{max}(\cdot)$ 是一个最大化函数，用来突出伪装图 C_s 初始的伪装区域。

为了全面获取高层特征，本文进一步使用 PDC 来聚合另外三层的特征，并通过 RF 进行增强，以获得最终的伪装图 C_i：

$$C_i = PD_i(rf_1^i, rf_2^i, rf_3^i), \tag{3}$$

其中 $\{rf_k^i = rf_k, k=1,2,3\}$ 。PD_s 和 PD_i 之间的差别是输入特征的数量不同。

部分解码组件 (Partial Decoder Component, PDC). 形式上，给定一组来自搜索和识别阶段获取的特征 $\{rf_k^c, k\in[m,\ldots,M], c\in[s,i]\}$，本文使用上下文模块来生成新特征 $\{rf_k^{c1}\}$。并采用逐元素相乘方式来减少相邻特征之间的差距。具体来说，对于最浅层的特征，如 rf_4^s，当 $k=M$ 时，令 $rf_M^{c1}=rf_M^{c2}$。对于较深层的特征，如 rf_k^{c1}，$k<M$ 时，则将 rf_k^{c2} 更新为：

$$rf_k^{c2} = rf_k^{c1} \otimes \Pi_{j=k+1}^{M} Bconv(UP(f_j^{c1})), \tag{4}$$

其中 $k\in[m,\ldots,M-1]$，$Bconv(\cdot)$ 是结合了 3×3 卷积、批量归一化和 ReLU 函数的顺序操作。$UP(\cdot)$ 为具有 2^{j-k} 倍的上采样操作。最后，通过拼接方式将这些具有区分性的特征组合在一起。训练 *SINet* 模型损失函数为交叉熵 [78] 损失函数 L_{CE}。总损失函数 L 为:

$$L = L_{CE}^{s}(C_{csm}, G) + L_{CE}^{i}(C_{cim}, G), \tag{5}$$

其中 C_{csm} 和 C_{cim} 是对 C_s 和 C_i 上采样后获得的两个伪装物体映射图，其分辨率为：352×352。

4.3. 实施细节.

SINet 是在 PyTorch 中实现的，并使用 Adam 优化器进行训练 [30]。在训练阶段，批处理大小为 36，学习率从 1e-4 开始。整个训练过程共有 30 个阶段 (采用提前停止策略)，仅耗时大约 70 分钟。运行时间是在以下平台上测试获得：Intel® i9-9820X CPU @3.30GHz × 20、TITAN RTX。分辨率为 352×352 的图像，其推理时间为 0.2s。

5. 评测实验

5.1. 实验设置

训练和测试细节. 为了验证 *SINet* 的通用性，使用以下三种训练集（伪装图像）：**(i)** CAMO [33]；**(ii)** COD10K；**(iii)** CAMO + COD10K + EXTRA。对于 CAMO，则使用默认的训练集。同样对于 COD10K，也使用默认的伪装图像训练集。本文在以下数据集上进行模型评估：整个 CHAMELEON [57] 数据集，CAMO 和 COD10K 两个数据集对应的测试集。

基线. 由于没有公开的基于深度网络的 COD 模型。因此，本文根据以下标准选择了 12 个深度学习的基准模型 [3,24,28,33,36,41,52,69,76,78,79,83] 并按照第 iv 个训练方案，使用原文推荐的参数来训练这些基准模型。(1) 经典框架，(2) 最新发布，(3) 在特定领域 (例如 GOD 或 SOD) 中达到 SOTA 性能。

5.2. 结果与数据分析

CHAMELEON 性能表现. 从表. 3中可以看出，相比 12 个 SOTA 的物体检测基准，本文提出的 *SINet* 在所有指标上均获胜。尤其是本文模型并未使用任何边缘/边界特征 (如 EGNet [78], PFANet [79])、预处理技术 [47] 或后处理策略 (如 CRF [32]，图割模型 [2])。

CAMO 性能表现. 本文还在最近提出的 CAMO [33] 数据集上进行了模型测试，它包括了各种伪装物体。根据表. 3中全面的分析，不难发现 CAMO 数据

基准模型	CHAMELEON [57]				CAMO-Test [33]				COD10K-Test (Ours)			
	$S_\alpha \uparrow$	$E_\phi \uparrow$	$F_\beta^w \uparrow$	$M \downarrow$	$S_\alpha \uparrow$	$E_\phi \uparrow$	$F_\beta^w \uparrow$	$M \downarrow$	$S_\alpha \uparrow$	$E_\phi \uparrow$	$F_\beta^w \uparrow$	$M \downarrow$
2017 FPN [36]	0.794	0.783	0.590	0.075	0.684	0.677	0.483	0.131	0.697	0.691	0.411	0.075
2017 MaskRCNN [24]	0.643	0.778	0.518	0.099	0.574	0.715	0.430	0.151	0.613	0.748	0.402	0.080
2017 PSPNet [76]	0.773	0.758	0.555	0.085	0.663	0.659	0.455	0.139	0.678	0.680	0.377	0.080
2018 UNet++ [83]	0.695	0.762	0.501	0.094	0.599	0.653	0.392	0.149	0.623	0.672	0.350	0.086
2018 PiCANet [41]	0.769	0.749	0.536	0.085	0.609	0.584	0.356	0.156	0.649	0.643	0.322	0.090
2019 MSRCNN [28]	0.637	0.686	0.443	0.091	0.617	0.669	0.454	0.133	0.641	0.706	0.419	0.073
2019 BASNet [52]	0.687	0.721	0.474	0.118	0.618	0.661	0.413	0.159	0.634	0.678	0.365	0.105
2019 PFANet [79]	0.679	0.648	0.378	0.144	0.659	0.622	0.391	0.172	0.636	0.618	0.286	0.128
2019 CPD [69]	0.853	0.866	0.706	0.052	0.726	0.729	0.550	0.115	0.747	0.770	0.508	0.059
2019 HTC [3]	0.517	0.489	0.204	0.129	0.476	0.442	0.174	0.172	0.548	0.520	0.221	0.088
2019 EGNet [78]	0.848	0.870	0.702	0.050	0.732	0.768	0.583	0.104	0.737	0.779	0.509	0.056
2019 ANet-SRM [33]	‡	‡	‡	‡	0.682	0.685	0.484	0.126	‡	‡	‡	‡
SINet'20 Training setting (i)	0.737	0.737	0.478	0.103	0.708	0.706	0.476	0.131	0.685	0.718	0.352	0.092
SINet'20 Training setting (ii)	0.846	0.871	0.691	0.050	0.665	0.662	0.470	0.128	0.758	0.796	0.517	0.054
SINet'20 Training setting (iii)	**0.869**	**0.891**	**0.740**	**0.044**	**0.751**	**0.771**	**0.606**	**0.100**	**0.771**	**0.806**	**0.551**	**0.051**

表 3: 各数据集上的定量评估结果。最高分以**粗体**突出。训练设置见 § 5.1：(i) CAMO, (ii) COD10K, (iii) CAMO + COD10K + EXTRA. 注意，ANet-SRM 模型码 (仅在 CAMO 上训练) 没有开源代，因此，因此其他数据集的结果无法得知 ('‡')。↑ 表示评分越高越好。E_ϕ 表示 E 测评法 [13] 的均值. 基准模型使用第 (iv) 集合进行训练。评测代码详见：https://github.com/DengPingFan/CODToolbox

集比之前的两个数据集 (CHAMELEON，CPD1K) 更具挑战性。*SINet* 再次获得最佳表现，因此进一步证明了它的鲁棒性。

COD10K 性能表现. 使用 COD10K 的测试集 (2,026 张图像) 进行测试，可以再次发现 *SINet* 一如既往地优于其他模型。这是因为其专门设计的搜索和识别模块可以自动学到丰富的高、中、低层的特征，而这些正是解决伪装物体边缘难以确定的这一具有挑战性问题的关键 (见 图. 9)。

GOD 与 SOD 的基准模型. 值得注意，在排名前三的模型中，GOD 模型 (FPN [36]) 比 SOD 的 CPD [69] 和 EG-Net [78] 表现差，这说明 SOD 框架可能更适合扩展到 COD 任务上。不论与 GOD [3, 24, 28, 36, 76, 83] 模型还是 SOD [39, 41, 52, 69, 78, 79] 模型相比，*SINet* 明显缩短了训练时间 (例如，*SINet*: 耗时 1 小时而 EGNet 需要耗费 48 小时)，并且在所有数据集上都达到了 SOTA 性能，这表明本模型有希望解决 COD 问题。

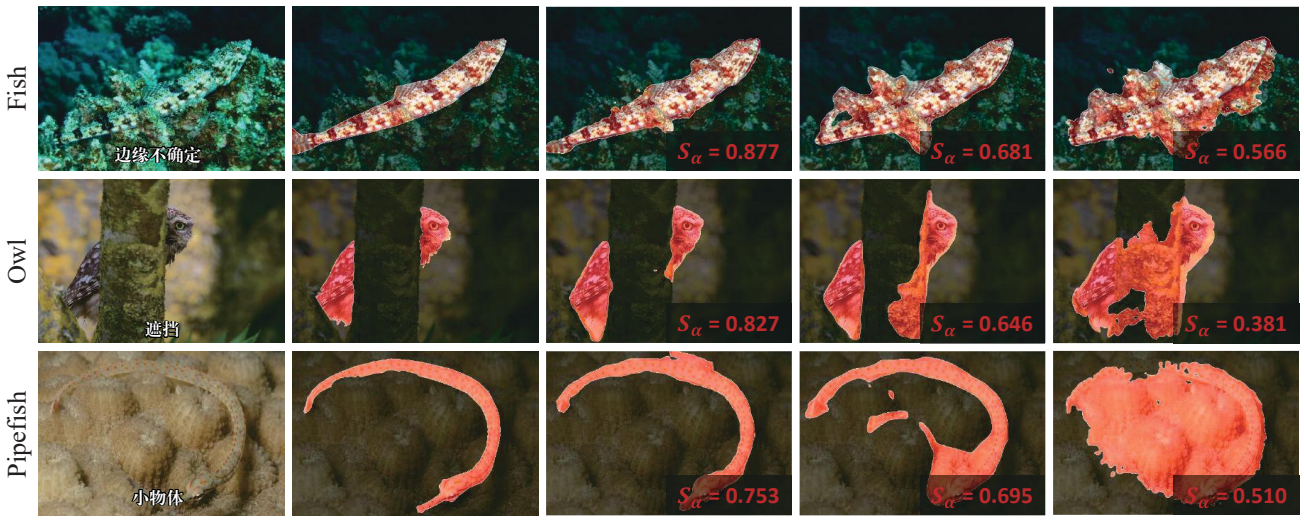
数据集间的泛化性. 数据集的泛化性和难度在训练和评估不同算法 [62] 中都起着至关重要的作用。因此，本文使用跨数据集分析法 [60]，对现有 COD 数据集进行研究，即在一个数据集上训练模型，然后在其他数据集上进行测试。本文选择了两个数据集，包括 CAMO [33] 和本文构建的 *COD10K* 数据集。参考 [62]，本文也对于每个数据集，随机选择 800 张图像作为训练集，选择 200 张图像作为测试集，因为 CPD1K 数据集仅包含 1,000 张图像。为公平比较，*SINet* 模型在每个数据集上都训练到损失稳定为止。

训练: \ 测试:	CAMO [33]	COD10K (Ours)	自身	其他均值	下降 ↓
CAMO [33]	**0.803**	0.702	0.803	0.678	15.6%
COD10K (Ours)	0.742	**0.700**	0.700	0.683	2.40%
其他均值	0.641	0.589			

表 4: *SINet* 模型在跨数据集下得到的 S 测评法 (S-measure↑ [12]) 结果。*SINet* 在某一行上的对应数据集上进行训练，然后分别在列上的所有数据集上进行测试。"自身 (Self)"：表示训练和测试都在同一个（对角线上）数据集上进行。"其他均值 (Mean others)" 表示在其他数据集上的平均分数。

表. 4 提供了在跨数据集上使用 S 测评法 (S-measure) 的结果。每行数据表示用这一行的数据集训练，再对各个列中的数据集进行测试。因此，从行的角度看，可以得知该行数据集的泛化性。每列数据表示在对应行数据集上进行训练后，再对列上数据集上进行测试，因此，从列的角度分析，可知该列数据集的难度。请注意，训练和测试的设置与表. 3中使用的设置不同。因此性能无法相互比较。正如预期，本文的 *COD10K* 是难度最大的数据集 (即，最后一行的其他平均为：0.589)。这是因为数据集包含了各种具有挑战性的伪装物体 (见 § 3)。

定性分析. 图. 9给出了本文 *SINet* 模型与两个基准模型之间的定性比较。观察发现，PFANet [79] 虽然定位到伪装物体，但是输出总是不准确。通过进一步使用边缘特征，EGNet [78] 获得了比 PFANet 相对更准确的定位。然而，它仍然忽略了物体的细微



(a) 原图 (b) 真值 (c) SINet (Ours) (d) EGNet [78] (e) PFANet [79]

图 9: *SINet* 和两个在 *COD10K* 数据集上表现最好的基准模型的定性分析。细节详见 补充材料。

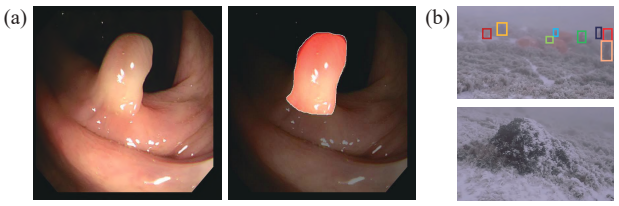


图 10: 更多应用示例。(a) 息肉检测/分割结果。(b) 用于灾区的搜救系统。

之处，尤其是第一行的鱼类。所有这些具有挑战的示例 (例如，*边缘不确定*，*遮挡*和*小物体*)，*SINet* 模型均可检测出伪装物体的精确细节从而预测出真正伪装物体，展示了本文框架的稳健性。

6. 潜在应用

伪装检测系统 (**C**amouflage **D**etection **S**ystems, **CDS**) 有许多潜在的应用。本文在此预想了两种潜在用途。

医学图像分割. 配备了 CDS 的医疗图像分割算法，用息肉等特殊对象训练后，则可以用于在真实场景中自动分割息肉 (图. 10 a)，或者在自然界中发现和保护稀有物种，甚至在灾难场景中用于搜索和救援。

搜索引擎. 图. 11是一个谷歌的搜索结果示例。从结果 (图. 11 a) 中，可以注意到搜索引擎无法检测到隐藏的蝴蝶，因此只能搜索到具有相似背景的图像。有趣的是，当搜索引擎配备了 CDS(在此，只是简单地修改了关键字)，引擎就可以识别伪装物体，并且

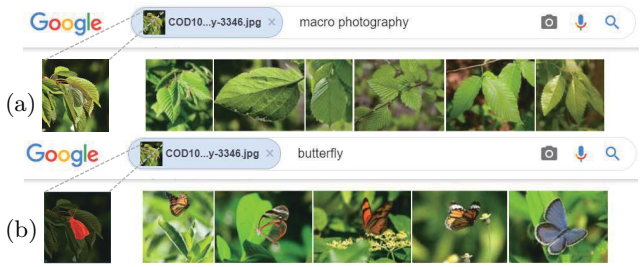


图 11: 配备 CDS 与否的搜索引擎示例，(a)：无 CDS，(b) 有 CDS。

反馈一些蝴蝶的图像 (图. 11 b)。

7. 结论

本文从伪装角度为物体检测构建了一个完整的评测。具体来说，本文构建了一个新的、有挑战性的、稠密标注的 *COD10K* 数据集，并进行了大规模的评估，开发了一个简单而有效的端到端 *SINet* 框架，并展示了一些潜在应用。与目前先进基准模型相比，*SINet* 框架极具竞争力竞，并且在视觉上能够得到更好的结果。以上贡献旨在为本领域提供一个为 COD 任务设计新模型的机会。在未来的工作中，我们计划扩展 *COD10K* 数据集，可提供各种形式的输入，例如 RGB-D 伪装物体检测 (类似于 RGB-D 显着物体检测 [20,72,75])，或者其他形式。探索一些新技术，例如弱监督学习 [54,55]、零样本学习 [84]、变分自动编码模型 [85] 和多尺度骨干网络 [21]。

参考文献

[1] Ali Borji, Ming-Ming Cheng, Qibin Hou, Huaizu Jiang, and Jia Li. Salient object detection: A survey. *Computational Visual Media*, 5(2):117–150, 2019.

[2] Yuri Boykov, Olga Veksler, and Ramin Zabih. Fast approximate energy minimization via graph cuts. In *IEEE CVPR*, pages 377–384, 1999.

[3] 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 *IEEE CVPR*, pages 4974–4983, 2019.

[4] Ming-Ming Cheng, Yun Liu, Wen-Yan Lin, Ziming Zhang, Paul L Rosin, and Philip HS Torr. Bing: Binarized normed gradients for objectness estimation at 300fps. *Computational Visual Media*, 5(1):3–20, 2019.

[5] 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.

[6] Hung-Kuo Chu, Wei-Hsin Hsu, Niloy J Mitra, Daniel Cohen-Or, Tien-Tsin Wong, and Tong-Yee Lee. Camouflage images. *ACM Trans. Graph.*, 29(4):51–1, 2010.

[7] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In *IEEE CVPR*, pages 3213–3223, 2016.

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

[9] Innes C Cuthill, Martin Stevens, Jenna Sheppard, Tracey Maddocks, C Alejandro Párraga, and Tom S Troscianko. Disruptive coloration and background pattern matching. *Nature*, 434(7029):72, 2005.

[10] Dima Damen, Hazel Doughty, Giovanni Maria Farinella, Sanja Fidler, Antonino Furnari, Evangelos Kazakos, Davide Moltisanti, Jonathan Munro, Toby Perrett, Will Price, et al. Scaling egocentric vision: The epic-kitchens dataset. In *ECCV*, pages 720–736, 2018.

[11] Mark Everingham, SM Ali Eslami, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The PASCAL visual object classes challenge: A retrospective. *IJCV*, 111(1):98–136, 2015.

[12] Deng-Ping Fan, Ming-Ming Cheng, Yun Liu, Tao Li, and Ali Borji. Structure-measure: A New Way to Evaluate Foreground Maps. In *IEEE ICCV*, pages 4548–4557, 2017.

[13] Deng-Ping Fan, Cheng Gong, Yang Cao, Bo Ren, Ming-Ming Cheng, and Ali Borji. Enhanced-alignment Measure for Binary Foreground Map Evaluation. In *IJCAI*, pages 698–704, 2018.

[14] Deng-Ping Fan, Ge-Peng Ji, Guolei Sun, Ming-Ming Cheng, Jianbing Shen, and Ling Shao. Camouflaged object detection. In *IEEE CVPR*, pages 2777–2787, 2020.

[15] 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. *arXiv*, 2020.

[16] Deng-Ping Fan, Zheng Lin, Ge-Peng Ji, Dingwen Zhang, Huazhu Fu, and Ming-Ming Cheng. Taking a deeper look at the co-salient object detection. In *IEEE CVPR*, 2020.

[17] Deng-Ping Fan, Zheng Lin, Zhao Zhang, Menglong Zhu, and Ming-Ming Cheng. Rethinking RGB-D Salient Object Detection:Models, Datasets, and Large-Scale Benchmarks. *IEEE TNNLS*, 2020.

[18] Deng-Ping Fan, Jiang-Jiang Liu, Shang-Hua Gao, Qibin Hou, Ali Borji, and Ming-Ming Cheng. Salient objects in clutter: Bringing salient object detection to the foreground. In *ECCV*, pages 1597–1604. Springer, 2018.

[19] 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 Scans. *IEEE TMI*, 2020.

[20] Keren Fu, Deng-Ping Fan, Ge-Peng Ji, and Qijun Zhao. JL-DCF: Joint Learning and Densely-Cooperative Fusion Framework for RGB-D Salient Object Detection. In *IEEE CVPR*, 2020.

[21] Shanghua Gao, Ming-Ming Cheng, Kai Zhao, Xin-Yu Zhang, Ming-Hsuan Yang, and Philip HS Torr. Res2net: A new multi-scale backbone architecture. *IEEE TPMAI*, 2020.

[22] Shiming Ge, Xin Jin, Qiting Ye, Zhao Luo, and Qiang Li. Image editing by object-aware optimal boundary searching and mixed-domain composition. *CVM*, 4(1):71–82, 2018.

[23] Joanna R Hall, Innes C Cuthill, Roland Baddeley, Adam J Shohet, and Nicholas E Scott-Samuel. Camouflage, detection and identification of moving targets. *Proc. R. Soc. B: Biological Sciences*, 280(1758):20130064, 2013.

[24] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *IEEE ICCV*, pages 2961–2969, 2017.

[25] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *IEEE CVPR*, pages 770–778, 2016.

[26] Qibin Hou, Ming-Ming Cheng, Xiaowei Hu, Ali Borji, Zhuowen Tu, and Philip Torr. Deeply supervised salient object detection with short connections. *IEEE TPAMI*, 41(4):815–828, 2019.

[27] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *IEEE CVPR*, pages 4700–4708, 2017.

[28] Zhaojin Huang, Lichao Huang, Yongchao Gong, Chang Huang, and Xinggang Wang. Mask scoring r-cnn. In *IEEE CVPR*, pages 6409–6418, 2019.

[29] Laurent Itti, Christof Koch, and Ernst Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE TPAMI*, 20(11):1254–1259, 1998.

[30] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *ICLR*, 2015.

[31] Alexander Kirillov, Kaiming He, Ross Girshick, Carsten Rother, and Piotr Dollár. Panoptic segmentation. In *IEEE CVPR*, pages 9404–9413, 2019.

[32] Philipp Krahenbuhl and Vladlen Koltun. Efficient inference in fully connected crfs with gaussian edge potentials. In *NIPS*, pages 109–117, 2011.

[33] Trung-Nghia Le, Tam V Nguyen, Zhongliang Nie, Minh-Triet Tran, and Akihiro Sugimoto. Anabranch network for camouflaged object segmentation. *CVIU*, 184:45–56, 2019.

[34] Guanbin Li, Yuan Xie, Liang Lin, and Yizhou Yu. Instance-level salient object segmentation. In *IEEE CVPR*, pages 247–256, 2017.

[35] Yin Li, Xiaodi Hou, Christof Koch, James M Rehg, and Alan L Yuille. The secrets of salient object segmentation. In *IEEE CVPR*, pages 280–287, 2014.

[36] Tsungyi Lin, Piotr Dollar, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection. In *IEEE CVPR*, pages 936–944, 2017.

[37] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *ECCV*, pages 740–755. Springer, 2014.

[38] Ce Liu, Jenny Yuen, and Antonio Torralba. Sift flow: Dense correspondence across scenes and its applications. *IEEE TPAMI*, 33(5):978–994, 2010.

[39] Jiang-Jiang Liu, Qibin Hou, Ming-Ming Cheng, Jiashi Feng, and Jianmin Jiang. A simple pooling-based design for real-time salient object detection. *IEEE CVPR*, 2019.

[40] Li Liu, Wanli Ouyang, Xiaogang Wang, Paul Fieguth, Jie Chen, Xinwang Liu, and Matti Pietikäinen. Deep learning for generic object detection: A survey. *IJCV*, 2019.

[41] Nian Liu, Junwei Han, and Ming-Hsuan Yang. Picanet: Learning pixel-wise contextual attention for saliency detection. In *IEEE CVPR*, pages 3089–3098, 2018.

[42] Songtao Liu, Di Huang, et al. Receptive field block net for accurate and fast object detection. In *ECCV*, pages 385–400, 2018.

[43] Yun Liu, Ming-Ming Cheng, Deng-Ping Fan, Le Zhang, JiaWang Bian, and Dacheng Tao. Semantic edge detection with diverse deep supervision. *arXiv preprint arXiv:1804.02864*, 2018.

[44] Ran Margolin, Lihi Zelnik-Manor, and Ayellet Tal. How to evaluate foreground maps? In *IEEE CVPR*, pages 248–255, 2014.

[45] Gerard Medioni. Generic object recognition by inference of 3-d volumetric. *Object Categorization: Computer and Human Vision Perspectives*, 87, 2009.

[46] Kaichun Mo, Shilin Zhu, Angel X Chang, Li Yi, Subarna Tripathi, Leonidas J Guibas, and Hao Su. Partnet: A large-scale benchmark for fine-grained and hierarchical part-level 3d object understanding. In *IEEE CVPR*, pages 909–918, 2019.

[47] Greg Mori. Guiding model search using segmentation. In *IEEE ICCV*, pages 1417–1423, 2005.

[48] Gerhard Neuhold, Tobias Ollmann, Samuel Rota Bulo, and Peter Kontschieder. The mapillary vistas dataset for semantic understanding of street scenes. In *IEEE CVPR*, pages 4990–4999, 2017.

[49] Andrew Owens, Connelly Barnes, Alex Flint, Hanumant Singh, and William Freeman. Camouflaging an object from many viewpoints. In *IEEE CVPR*, pages 2782–2789, 2014.

[50] Federico Perazzi, Philipp Krähenbühl, Yael Pritch, and Alexander Hornung. Saliency filters: Contrast based filtering for salient region detection. In *IEEE CVPR*, pages 733–740, 2012.

[51] Federico Perazzi, Jordi Pont-Tuset, Brian McWilliams, Luc Van Gool, Markus Gross, and Alexander Sorkine-Hornung. A benchmark dataset and evaluation methodology for video object segmentation. In *IEEE CVPR*, pages 724–732, 2016.

[52] Xuebin Qin, Zichen Zhang, Chenyang Huang, Chao Gao, Masood Dehghan, and Martin Jagersand. Basnet: Boundary-aware salient object detection. In *IEEE CVPR*, pages 7479–7489, 2019.

[53] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, and et al. Imagenet large scale visual recognition challenge. *IJCV*, 115(3):211–252, 2015.

[54] Yunhang Shen, Rongrong Ji, Yan Wang, Yongjian Wu, and Liujuan Cao. Cyclic guidance for weakly supervised joint detection and segmentation. In *IEEE CVPR*, pages 697–707, 2019.

[55] Yunhan Shen, Rongrong Ji, Shengchuan Zhang, Wangmeng Zuo, and Yan Wang. Generative adversarial learning towards fast weakly supervised detection. In *IEEE CVPR*, pages 5764–5773, 2018.

[56] Jamie Shotton, John Winn, Carsten Rother, and Antonio Criminisi. Textonboost: Joint appearance, shape and context modeling for multi-class object recognition and segmentation. In *ECCV*, pages 1–15. Springer, 2006.

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

[58] Martin Stevens and Sami Merilaita. Animal camouflage: current issues and new perspectives. *Phil. Trans. R. Soc. B: Biological Sciences*, 364(1516):423–427, 2008.

[59] 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*. Macmillan Company, 1909.

[60] Antonio Torralba, Alexei A Efros, et al. Unbiased look at dataset bias. In *IEEE CVPR*, pages 1521–1528, 2011.

[61] Tom Troscianko, Christopher P Benton, P George Lovell, David J Tolhurst, and Zygmunt Pizlo. Camouflage and visual perception. *Phil. Trans. R. Soc. B: Biological Sciences*, 364(1516):449–461, 2008.

[62] Wenguan Wang, Qiuxia Lai, Huazhu Fu, Jianbing Shen, and Haibin Ling. Salient object detection in the deep learning era: An in-depth survey. *arXiv preprint arXiv:1904.09146*, 2019.

[63] Wenguan Wang and Jianbing Shen. Deep visual attention prediction. *IEEE TIP*, 27(5):2368–2378, 2017.

[64] 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 *IEEE CVPR*, pages 5968–5977, 2019.

[65] Wenguan Wang, Jianbing Shen, Xingping Dong, and Ali Borji. Salient object detection driven by fixation prediction. In *IEEE CVPR*, pages 1711–1720, 2018.

[66] Wenguan Wang, Jianbing Shen, Ling Shao, and Fatih Porikli. Correspondence driven saliency transfer. *IEEE TIP*, 25(11):5025–5034, 2016.

[67] Wenguan Wang, Shuyang Zhao, Jianbing Shen, Steven CH Hoi, and Ali Borji. Salient object detection with pyramid attention and salient edges. In *IEEE CVPR*, pages 1448–1457, 2019.

[68] Yu-Huan Wu, Shang-Hua Gao, Jie Mei, Jun Xu, Deng-Ping Fan, Chao-Wei Zhao, and Ming-Ming Cheng. JCS: An Explainable COVID-19 Diagnosis System by Joint Classification and Segmentation. *arXiv preprint arXiv:2004.07054*, 2020.

[69] Zhe Wu, Li Su, and Qingming Huang. Cascaded partial decoder for fast and accurate salient object detection. In *IEEE CVPR*, pages 3907–3916, 2019.

[70] Amir R Zamir, Alexander Sax, William Shen, Leonidas J Guibas, Jitendra Malik, and Silvio Savarese. Taskonomy: Disentangling task transfer learning. In *IEEE CVPR*, pages 3712–3722, 2018.

[71] Yi Zeng, Pingping Zhang, Jianming Zhang, Zhe Lin, and Huchuan Lu. Towards high-resolution salient object detection. In *IEEE ICCV*, 2019.

[72] Jing Zhang, Deng-Ping Fan, Yuchao Dai, Saeed Anwar, Fatemeh Sadat Saleh, Tong Zhang, and Nick Barnes. UC-Net: Uncertainty Inspired RGB-D Saliency Detection via Conditional Variational Autoencoders. In *IEEE CVPR*, 2020.

[73] Pingping Zhang, Dong Wang, Huchuan Lu, Hongyu Wang, and Xiang Ruan. Amulet: Aggregating multi-level convolutional features for salient object detection. In *IEEE CVPR*, pages 202–211, 2017.

[74] Yunke Zhang, Lixue Gong, Lubin Fan, Peiran Ren, Qixing Huang, Hujun Bao, and Weiwei Xu. A late fusion cnn for digital matting. In *IEEE CVPR*, pages 7469–7478, 2019.

[75] Zhao Zhang, Zheng Lin, Jun Xu, Wenda Jin, Shao-Ping Lu, and Deng-Ping Fan. Bilateral attention network for rgb-d salient object detection. *arXiv preprint arXiv:2004.14582*, 2020.

[76] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. Pyramid scene parsing network. In *IEEE CVPR*, pages 6230–6239, 2017.

[77] Jia-Xing Zhao, Yang Cao, Deng-Ping Fan, Ming-Ming Cheng, Xuan-Yi Li, and Le Zhang. Contrast prior and fluid pyramid integration for RGBD salient object detection. In *IEEE CVPR*, pages 3927–3936, 2019.

[78] 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 *IEEE ICCV*, 2019.

[79] Ting Zhao and Xiangqian Wu. Pyramid feature attention network for saliency detection. In *IEEE CVPR*, pages 3085–3094, 2019.

[80] Zhong-Qiu Zhao, Peng Zheng, Shou-tao Xu, and Xindong Wu. Object detection with deep learning: A review. *IEEE TNNLS*, 30(11):3212–3232, 2019.

[81] Bolei Zhou, Agata Lapedriza, Aditya Khosla, Aude Oliva, and Antonio Torralba. Places: A 10 million image database for scene recognition. *IEEE TPAMI*, 40(6):1452–1464, 2017.

[82] Bolei Zhou, Hang Zhao, Xavier Puig, Sanja Fidler, Adela Barriuso, and Antonio Torralba. Scene parsing through ade20k dataset. In *IEEE CVPR*, pages 633–641, 2017.

[83] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: A nested u-net architecture for medical image segmentation. In *DLMIA*, pages 3–11, 2018.

[84] Yizhe Zhu, Mohamed Elhoseiny, Bingchen Liu, Xi Peng, and Ahmed Elgammal. A generative adversarial approach for zero-shot learning from noisy texts. In *IEEE CVPR*, pages 1004–1013, 2018.

[85] Yizhe Zhu, Martin Renqiang Min, Asim Kadav, and Hans Peter Graf. S3VAE: Self-Supervised Sequential VAE for Representation Disentanglement and Data Generation. In *IEEE CVPR*, 2020.