

基于交互图学习的伪装目标检测*

翟强^{1,†} 李鑫^{2,‡} 杨帆^{2,†} 陈程立³ 程洪¹ 范登平⁴

¹ 电子科技大学 ² Group 42 (G42) ³ 青岛大学 ⁴ 起源人工智能研究院 (IIAI)

摘要

目前的模型很难自动检测或分割融入周围环境的对象。一个主要的挑战在于，这些前景对象与背景环境之间的内在相似性，使得深度模型提取的特征无法区分。为了克服这一难点，一个理想的模型应该能够从给定的场景中寻找有价值的、额外的线索，并将其纳入到用于表征增强的联合学习框架中。以此为启发，本文设计了一个新颖的交互图学习 (MGL) 模型，该模型将传统的交互学习思想从规则网格空间推广至图域。具体来说，MGL 将图像接耦为两个特定任务的特征图——一个用于大致定位目标，另一个用于准确捕捉目标的边界细节——并通过图对它们的高阶关系进行反复推理来充分利用交互的好处。重要的是，与大多数使用共享函数来建模所有任务间交互的交互学习方法相比，MGL 配备了处理不同互补关系的类型化函数用以最大化信息交互。在多个具有挑战性的数据集（包括 CHAMELEON, CAMO 和 COD10K）上的实验证明本文 MGL 的性能显著优于现有最先进的方法。代码可参考：<https://github.com/fanyang587/MGL>

1. 引言

伪装在自然界中是一项重要的技能，因为它可以帮助某些动物通过融入周围环境来躲避捕食者。伪装能力与人类的感知能力密切相关，在过去几十年里吸引了越来越多的关注。生物和心理学研究表

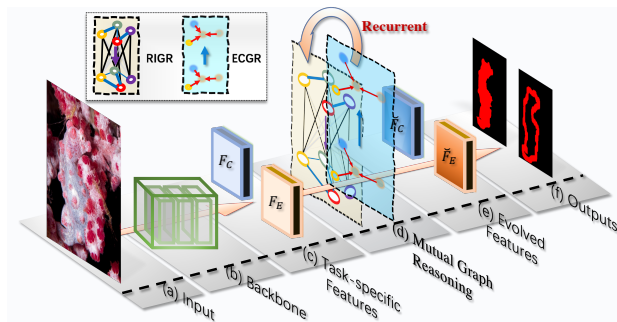


图 1: MGL 示意图。给定图像 (a), MGL 使用 ResNet-FCN 作为骨干网络 (b) 分别提取与任务特定的特征用于伪装目标检测 (COD) 和伪装目标感知边缘提取 (COEE)(c), 然后, 利用区域诱导图推理 (RIGR) 模块和边缘约束图推理 (ECGR) 模块的协作, 以循环的方式推理它们之间的相互关系 (d), 最后, 将衍化的特征 (e) 映射为最终的预测结果 (f)。

明, 人类很难快速发现伪装的动物或物体 [4,48]。一个可能的原因是人类的视觉系统的原始功能可能被设计用于识别拓扑属性 [2], 因此很难识别伪装的动物或物体, 这些动物或物体破坏了它们‘真实’身体的视觉边缘信息。尽管有了这些生物学上的发现, 遗憾的是, 在计算机视觉领域如何通过机器来弥补人类感知的这个“缺陷”仍然是有待探索的课题。

从背景中识别伪装对象, 也被称为伪装对象检测 (COD) [7], 是一个有价值的, 但具有挑战性的任务 [9]。”捕捉伪装”在很多现实生活任务中有很好的前景, 包括图像检索 [29]、物种发现 [42]、交通风险管理、医学图像分析 [10,12,58] 等等。然而, 现有的深度模型仍然不能完全解决前景物体与背景环境内在的视觉相似性。为了克服这一困难, 目前的方法从共享上下文中提取辅助特征来获取额外的知

*本文为 CVPR'21 论文 [66] 的中文翻译版。

†通讯作者: 杨帆 (fanyang_uestc@hotmail.com); ‡ 共一。

识, 例如, 用于识别 [9] 或分类 [20] 的特征显著地增强了伪装对象检测的底层表示。尽管他们成功证明了利用额外的知识对伪装对象检测的益处, 但仍然有三个主要的问题有待解决。首先, COD 任务与其辅助任务之间的交互影响被忽视或缺乏研究。更具体地说, 由于现有的工作 [9, 20, 63] 只利用额外的信息来指导主任务 (即 COD), 而忽略了它们之间重要的协作关系, 进而可能导致这类模型陷入局部最小 [49]。其次, 由于交叉任务依赖关系仅在原始坐标空间中建模, 更全局更高阶的知道信息可能被丢掉。经验表明, 当前的 COD 模型在严重遮挡以及不确定边界下将变得无效, 因为它们未能将高阶信息纳入到表征学习过程。最后, 根据最近的生物学发现 [17, 53, 54], 隐藏或伪装的关键因素是边缘破坏。遗憾的是, 现有的方法 [9, 20] 并没有研究如何提高真实边缘可见性来促进 COD 的表征学习, 这会削弱, 至少不能充分利用 COD 模型的学习能力。

针对这些不足, 本文提出了一种新颖的交互图学习模型 (MGL), 以充分和全面地利用伪装对象检测 (COD) 及其辅助任务之间的交互利益。考虑到边缘破坏是伪装 [17, 53, 54] 的关键因素之一, 本文将伪装对象的边缘提取 (COEE) 作为一个辅助任务, 并将其纳入到 MGL 中进行交互学习。如图 1 所示, MGL 拥有一个精心设计的交织架构, 加强了任务之间的交互与协作。重要的是, 与现有研究中“天真地”融合从两个任务学习的特征不同, MGL 通过明确推理 COD 和 COEE 两种类型功能之间的互补关系, 精确地利用对应任务中的有用信息来协同增强表征。为了从 COD 中挖掘语义指导信息并辅助 COEE, 本文开发了一种新的区域诱导图推理模块 (RIGR) 来推理高级的依赖关系, 并从 COD 中传递语义信息来增强 COEE 的底层表示; 为了提高真实边缘的可见性, 本文使用了一个新的边缘约束图推理模块 (ECGR) 来显式地合并 COEE 的边缘信息, 反过来更好地指导 COD 的表征学习。重要的是, 本文的 RIGR 和 ECGR 可以以循环的方式递归地挖掘交互的益处, 并互相吸收有价值的信息。

本文在多个数据集上与强基准以及当前最先进的方法进行对比实验, 证明了 MGL 的有效性。实

验结果清晰地表明了 MGL 在伪装目标检测中挖掘交互指导信息的优越性。本文的贡献归纳如下:

- **一种新颖的基于图的、互学习的伪装目标检测方法。** 据我们所知, 这是第一次尝试利用两个密切相关的任务 (即 COD 和 COEE) 之间的相互引导知识, 并利用基于图的技术来实现伪装目标检测。该方法能够捕获语义引导知识和空间支持信息, 从而相互促进两个任务的检测性能。
- **精心设计的基于图形的交互功能, 以充分挖掘类型化的指导信息。** 和传统的交互学习方法不同, 本文的 MGL 集成了两个不同的基于图的交互模块来推理类型关系: 用于从 COD 挖掘语义指导信息以协助 COEE 的 RIGR, 以及用于合并真实边缘先验以增强 COD 的底层表征的 ECGR。
- **在广泛使用的基准上的最好结果。** MGL 在多个数据集上创下了新纪录, 如 CHAMELEON [47], CAMO [20] 以及 COD10K [9], 并大幅超过了现有 COD 模型的性能。

伪装目标检测。 伪装目标检测 (COD) 任务 [21, 36, 38] 通过推动通用或显著目标检测 [13, 22–24, 27, 28, 32, 33, 41, 46, 55, 70, 72, 75] 的边界将目标隐藏至周围环境, 对研究工作提出了新的挑战。Fan 等人 [9] 提出了搜索和识别网络 (SINet) 来解决这一挑战, SINet 首先粗略的搜索伪装目标, 然后对其进行分割。Le 等人 [20] 引入了 Anabranched Network (ANet) [20], 该网络将分类信息整合到表征学习中。Yan 等人 [63] 提出了 MirrorNet, 使用实例分割和对抗攻击来解决 COD 任务。这类仿生模型背后的共同想法是, 探索并集成额外的线索到表征学习中, 以大幅优于传统的通用目标检测 (GOD) 方法和显著性目标检测 (SOD) 方法 [11, 13, 16, 27, 31, 44, 45, 69, 73]。与之前的研究不同, 本文使用了一个统一的、基于图的模型, 通过综合推理多层次关系来同时执行伪装目标检测 (COD) 和伪装目标边缘检测 (COEE)。

图卷积网络。 GCNs 是图数据分析的有力工具, 而且已经产生了许多应用 [35, 39, 56, 60, 61, 64, 68]。在 (通用或显著性) 目标检测方面, GCNs 用于检测或者分割图像、视频或点云中的 2D/3D 目标 [50, 52]。

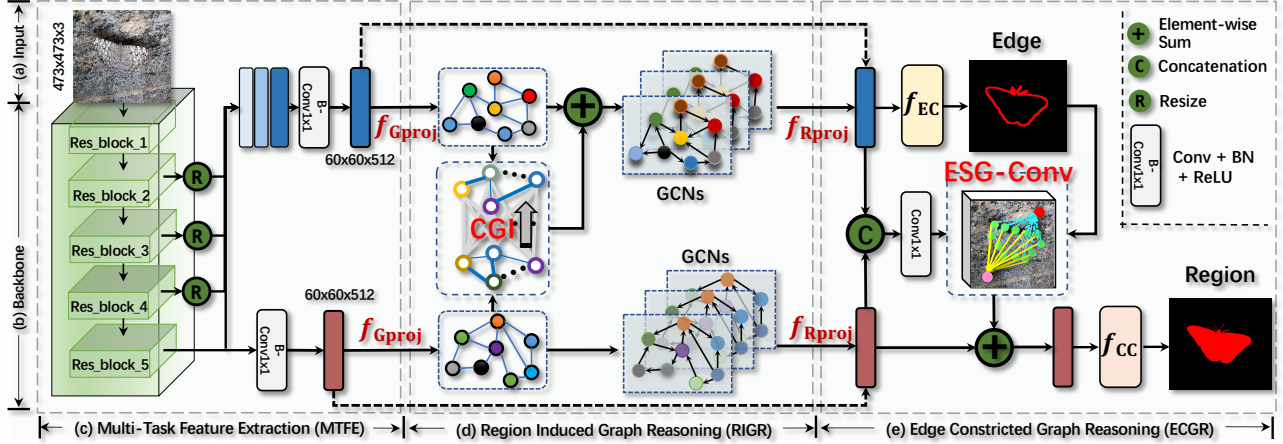


图 2: 交互图学习框架 (单个阶段, S-MGL) 示意图。流程图的主要组成部分如 (a) 至 (e) 所示。CGI 代表交叉图交互模块, ESG-Conv 代表边缘支撑图卷积, 是信息交互的关键操作。更多详情请参考²。

文献 [3, 25] 使用图卷积对长上下文进行建模来完成语义分割。Wu 等人 [57] 利用一个双向图来挖掘对象和背景之间的语义关系和共现。Luo 等人 [34] 引入了一个级联图模型, 利用多尺度、跨模态信息进行显著性目标检测。文献 [68] 引入了一种基于注意图聚类的自适应 GCN 模型进行联合显著性检测。针对伪装目标检测任务, 本文引入两个新颖的基于图的模块 (RIGR 和 ECGR) 以充分推理不同层次间的 COD 和 COEE 互补信息, 进而更好地从图像中学习表征, 克服多重挑战。

2. 本文的方法

2.1. 预备知识

动机。本文的方法受到生物学研究启发 [17, 53, 54]: 获取真实的身体或目标性状是捕捉伪装的关键。然后, 理想的伪装目标检测模型应该能够很好地检测到目标的真实边缘, 更重要的是, 将这些信息纳入至一个联合学习框架。直观地说, 通过在一个统一的、基于图的网络中传播信息, 所有相关的任务都可以相互受益。

问题形式化。设权重参数为 Θ 的函数 M_{Θ} 表示 COD 模型, 该函数接收图像 I 作为输入, 同时产生伪装对象区域图 $C \in [0, 1]$ 以及伪装对象边缘区域图 $E \in [0, 1]$, C 和 E 分别反映了每个像素属于伪装目标及其边缘的概率。给定打好标签的训练数据

集 $\{I_i, C_i, E_i\}_{i=1}^N$, 网络学习的目标是通过充分利用 COD 和 COEE 之间的交互益处来学习参数 Θ , 其中 I_i 是训练图像, C_i 是该图像对应的伪装目标区域图真值, E_i 代表伪装目标边缘图真值, E_i 可以自动从 C_i 生成得到。

2.2. 概述

MGL 包含三个主要的模块: 多任务特征提取模块 (MTFE), 区域诱导图推理 (RIGR) 模块以及边缘约束图推理 (ECGR) 模块。

- MTFE。给定输入图像 $I \in \mathbb{R}^{H \times W \times 3}$, 一个多任务骨干网络 f_{MTFE} 被解耦为两个特定任务的表征: 用于粗略检测对象的 $F_C \in \mathbb{R}^{h \times w \times c}$ 以及用于正确捕获其真实边缘的 $F_E \in \mathbb{R}^{h \times w \times c}$ 。
- RIGR。在这一阶段, 首先通过图投影操作 f_{Gproj} 将 F_C 和 F_E 转换为样本依赖的语义图 $\mathcal{G}_C = (\mathcal{V}_C, \mathcal{E}_C)$ 以及 $\mathcal{G}_E = (\mathcal{V}_E, \mathcal{E}_E)$, 其中具有相似特征的像素形成一个顶点, 边度量特征空间中顶点之间的关系。然后, 使用跨图交互模块 (CGI) f_{CGI} 捕获 \mathcal{G}_C 和 \mathcal{G}_E 之间的高级依赖关系, 并将语义信息从 \mathcal{V}_C 传递到 \mathcal{V}_E : $\mathcal{V}'_E = f_{\text{CGI}}(\mathcal{V}_C, \mathcal{V}_E)$ 。然后图推理操作 f_{GR} 使用图卷积 [18] 产生演化的图表征 \mathcal{V}_C 和 \mathcal{V}'_E 。最后将 \mathcal{V}_C 和 \mathcal{V}'_E 逆映射回原始坐标空间: $\hat{F}_C = f_{\text{Rproj}}(\mathcal{V}_C)$ 以及 $\check{F}_E = f_{\text{Rproj}}(\mathcal{V}'_E)$ 。

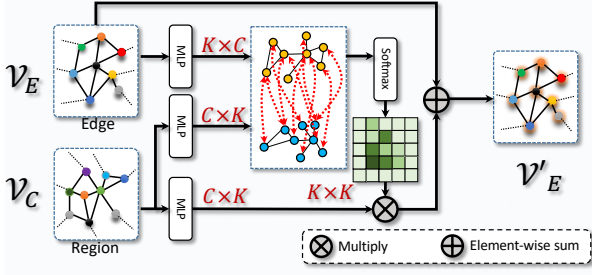


图 3: CGL.CGI 促进跨图 (任务) 交互, 并将 COD 的信息传递至用于学习 COEE 的演化图表征。

- ECGR。在空间关系分析之前, \check{F}_E 首先被输入到边缘分类器 f_{EC} 中以获得伪装对象的边缘映射 E 。此外, 模型还将 \check{F}_E 和 \hat{F}_C 融合 (例如通过 concatenate) 形成一个用于 COD 的新特性映射 F'_C , 然后在 E 的指导下, 使用一个新的边缘支撑图卷积 (ESG-Conv) 编码边缘信息并增强 F'_C 以便更好的定位对象: $\check{F}_C = \text{ESGConv}(F'_C; \mathcal{G}^e(E))$, $\mathcal{G}^e(E)$ 表示以 E 为条件的边缘支撑图。最后, 将 \check{F}_C 输入至分类器获得最终的预测结果 C 。

图 2 是本文方法的概况图。MGL 利用两个新的模块 (如 RIGR 和 ECGR) 在多层交互空间中推理了 COD 和 COEE 之间的关系。在表征学习过程中通过明确推理它们之间的关系, 可以直观地精确传播有价值的相互引导信息。值得一提的是, RIGR 和 ECGR 可以连续堆叠, 以实现循环性的相互学习。

2.3. 交互图学习

接下来本文详细介绍多任务特征提取 (MTFE), 区域诱导图推理 (RIGR) 以及边缘约束图推理 (ECGR) 模块。

多任务特征提取 (MTFE)。 f_{MTFE} 接受图像作为输入, 产生两个特定任务相关的特征图: 一个用于 COD, 另一个用于 COEE。具体来说, 给定一张输入图像 $I \in \mathbb{R}^{H \times W \times 3}$, 使用一个多任务的骨干网络 (一个以 Θ_{MTFE} 为参数的多分支 ResNet 的全联接网络) 同时获得 COD (F_C) 和 COEE (F_E) 的表征:

$$F_C = f_{\text{MTFE}}(I; \Theta_{\text{MTFE}}), \quad F_E = f_{\text{MTFE}}(I; \Theta_{\text{MTFE}}), \quad (1)$$

其中 $F_C \in \mathbb{R}^{h \times w \times c}$ 和 $F_E \in \mathbb{R}^{h \times w \times c}$ 是分别用于

COD 和 COEE 的 $h \times w$ 空间分辨率, c 通道特征, 因此可以很好的保存空间信息和高级语义信息。

边缘诱导图推理 (RIGR)。 RIGR 旨在推理 COD 内部以及 COD 和 COEE 之间的区域诱导语义关系, 而不考虑局部细节。它包含 4 个操作: (1) 图映射 f_{Gproj} , (2) 跨图交互 f_{CGI} , (3) 图推理 f_{GR} 以及 (4) 图逆映射 f_{Rproj} 。

(1) **图映射 f_{Gproj} 。** 给定输入特征 $F_C \in \mathbb{R}^{h \times w \times c}$ 或者 $F_E \in \mathbb{R}^{h \times w \times c}$, 首先使用 1×1 卷积层将它们变换至低维特征, 用 $F'_C \in \mathbb{R}^{(h \times w) \times C}$ 或者 $F'_E \in \mathbb{R}^{(h \times w) \times C}$ 表示。其次, 使用 f_{Gproj} 将特征向量 F'_C 或 F'_E 变换至图节点表征, 即 $\mathcal{V}_C \in \mathbb{R}^{C \times K}$ 或 $\mathcal{V}_E \in \mathbb{R}^{C \times K}$ 。参考 [26, 67], 本文将 f_{Gproj} 参数化为 $W \in \mathbb{R}^{K \times C}$, $\Sigma \in \mathbb{R}^{K \times C}$, W 的每一列 w_k 代表第 k 个节点的可学习的聚类中心。具体来说, 每个节点的表示可以按如下方式计算:

$$v_k = \frac{v'_k}{\|v'_k\|_2}, \quad v'_k = \frac{1}{\sum_i q_k^i} \sum_i q_k^i (f_i - w_k) / \sigma_k, \quad (2)$$

其中 σ_k 是 Σ 的列向量, v'_k 是特征向量 f_i 和 w_k 之间残差的加权平均值。 v_k 是第 k 个节点的表征, 也是特征矩阵 \mathcal{V} 的第 k 列。 q_k^i 是特征向量 f_i 对 w_k 的软赋值, 可由下式计算:

$$q_k^i = \frac{\exp(-\|(f_i - w_k) / \sigma_k\|_2^2 / 2)}{\sum_j \exp(-\|(f_i - w_j) / \sigma_j\|_2^2 / 2)}, \quad (3)$$

其中 ‘/’ 为元素级除法。本文通过测量节点内部表征之间的关系来计算图邻接矩阵: $\mathcal{A}^{\text{intra}} = f_{\text{norm}}(\mathcal{V}^T \times \mathcal{V}) \in \mathbb{R}^{K \times K}$, f_{norm} 为归一化操作。

(2) **跨图交互 f_{CGI} 。** f_{CGI} 建模图间的交互, 并引导图间信息从 \mathcal{V}_C 传递至 \mathcal{V}_E 。这一目标使我们从 non-local 操作 [51] 中获得灵感, 利用注意力机制计算图间的依赖关系。首先, 如图 3 所示, 本文使用不同的多层感知机 (MLPs) [43] 将 \mathcal{V}_C 变换至 key \mathcal{V}_C^k 以及 value \mathcal{V}_C^v , 将 \mathcal{V}_E 变换至 query \mathcal{V}_E^q 。然后相似性矩阵 $\mathcal{A}_{C \rightarrow E}^{\text{inter}} \in \mathbb{R}^{K \times K}$ 用矩阵乘法计算得到:

$$\mathcal{A}_{C \rightarrow E}^{\text{inter}} = f_{\text{norm}}(\mathcal{V}_E^{qT} \times \mathcal{V}_C^k), \quad (4)$$

where $\mathcal{A}_{C \rightarrow E}^{\text{inter}} \in \mathbb{R}^{K \times K}$ 。在此之后, 可以用以下公式将语义信息从 \mathcal{V}_C 传递至 \mathcal{V}_E :

$$\mathcal{V}'_E = f_{\text{CGI}}(\mathcal{V}_C, \mathcal{V}_E) = \chi(\mathcal{A}_{C \rightarrow E}^{\text{inter}} \times \mathcal{V}_C^v) + \mathcal{V}_E, \quad (5)$$

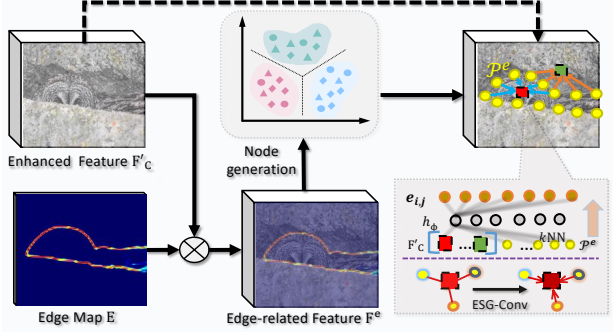


图 4: ECGR. ECGR 从 COEE 中挖掘有用信息用以指导 COD 的表征学习。支撑节点以一种‘soft’的方式产生。

其中 χ 作为权重参数来调整 CGI 关于 \mathcal{V}_E 的重要性。

(3) **图推理** f_{GR} 。在执行图间交互之后，本文使用 \mathcal{V}_C 和 \mathcal{V}'_E 作为输入来进行图内推理，以增强图表征。 f_{GR} 可以由图卷积实现 [18]:

$$\begin{cases} \mathcal{V}_C = f_{GR}(\mathcal{V}_C) = g(\mathcal{A}_C^{\text{intra}} \mathcal{V}_C W_C) \in \mathbb{R}^{C \times K}, \\ \mathcal{V}'_E = f_{GR}(\mathcal{V}'_E) = g(\mathcal{A}_E^{\text{intra}} \mathcal{V}'_E W_E) \in \mathbb{R}^{C \times K}, \end{cases} \quad (6)$$

其中 $g(\cdot)$ 是非线性激活函数， W_C 和 W_E 是图卷积层的可学习参数， $\mathcal{A}_C^{\text{intra}}$ 和 $\mathcal{A}_E^{\text{intra}}$ 分别是 \mathcal{V}_C 以及 \mathcal{V}'_E 的邻接矩阵。

(4) **图逆映射** f_{Rproj} 。为了将增强的图表征映射回原始坐标空间，回顾图投影步骤中的赋值矩阵。设 COD 的赋值矩阵表示为 $\mathcal{Q}_C = [q_{Ck}]_{k=0}^{(K-1)}$ ，其中 $q_{Ck} = [q_{Ck}^i]_{i=0}^{(h \times w) - 1}$ ，COEE 的赋值矩阵表示为 $\mathcal{Q}_E = [q_{Ek}]_{k=0}^{(K-1)}$ ，其中 $q_{Ek} = [q_{Ek}^i]_{i=0}^{(h \times w) - 1}$ 。图逆映射 f_{Rproj} 可被公式化为：

$$\begin{cases} \hat{F}_C = \mathcal{Q}_C \mathcal{V}_C^T + F_C^l, & \mathcal{Q}_C \in \mathbb{R}^{(h \times w) \times K}, \\ \check{F}_E = \mathcal{Q}_E \mathcal{V}'_E{}^T + F_E^l, & \mathcal{Q}_E \in \mathbb{R}^{(h \times w) \times K}, \end{cases} \quad (7)$$

其中 $\hat{F}_C \in \mathbb{R}^{(h \times w) \times C}$ 和 $\check{F}_E \in \mathbb{R}^{(h \times w) \times C}$ 分别是增强后用于 COD 和 COEE 的特征图。

边缘约束图推理 (ECGR)。为了进一步从 COEE 中提取有效信息以指导 COD 的表征学习，ECGR 侧重于边缘压缩关系推理。ECGR 的思路如图 4 所示。

(1) **目标**。ECGR 的目标是使模型具有明确的边缘感知能力，从而准确地定位对象。我们期望通过显式地感知和编码边缘的信息来更新 \hat{F}_C 。为实现这个目标，本文首先直接融合（如 concatenate） \check{F}_E 和 \hat{F}_C 生

成增强后的用于 COD 的特征图，然后以 E 为条件，使用一个新颖的边缘支撑图卷积 (ESG-Conv) 来更新这个特征。接下来，文章将描述边支撑图 $\mathcal{G}^e(E)$ 以及与之对应的图卷积 ESG-Conv。

(2) **支撑顶点生成**。构建 $\mathcal{G}^e(E)$ 的第一步是产生基于边缘的节点特征。首先，一个全连接层被用来将 \hat{F}_E 映射至一个伪装对象边缘图 $E \in \mathbb{R}^{h \times w \times 1}$ 。其次，如图 4 所示，模型以一种 ‘soft’ 方式用注意力机制得到规则网格和边缘相关的特征： $F^e = E \otimes F'_C$ ，其中 \otimes 代表通道级乘法。最后，使用图映射操作 f_{Gproj} 将 F^e 变换至 z 个基于边的节点特征，表示为 $\mathcal{P}^e = \{p_1^e, \dots, p_z^e\}$ 。

(3) **边缘支撑图卷积 ESG-Conv**。将边缘支撑图 $\mathcal{G}^e(E) = (\mathcal{V}^e, \mathcal{E}^e)$ 构造为 k 最近邻 (k -NN) 图 [52] 以连接 F'_C 和 \mathcal{P}^e ， \mathcal{V}^e 和 \mathcal{E}^e 分别代表顶点和边。具体来说，本文把每个特征向量 $f_i^c \in F'_C$ 作为中心节点， $\{p_j^e : (i, j) \in \mathcal{E}^e\}$ 作为其边缘支撑节点。边特征 $e_{i,j}$ 定义如下：

$$e_{i,j} = h_\phi(f_i^c, p_j^e) = f_{Conv}(f_i^c - p_j^e), \quad (8)$$

其中 h_ϕ 是以可学习参数 ϕ 为参数的非线性函数。第 i -th 个特征或顶点的 ESG-Conv 输出为：

$$\check{f}_i = \max_{j:(i,j) \in \mathcal{E}^e} h_\Phi(f_i^c, e_{i,j}), \quad (9)$$

其中 h_Φ 表示用以学习节点特征的以 Φ 为可学习参数的函数， $\check{f}_i \in \check{F}_C$ 为演化表征。使用 ESG-Conv 边缘信息可以显式地编码至底层表征中，即 $\check{F}_C = \text{ESGConv}(F'_C; \mathcal{G}^e(E))$ 。

循环学习过程。为了充分利用 COD 和 COEE 之间的相互利益，MGL 还可以按以下方式循环学习：

$$\begin{cases} \check{F}_E^{(t+1)} = f_{RIGR}(\check{F}_C^{(t)}, \check{F}_E^{(t)}), \\ \check{F}_C^{(t+1)} = f_{ECGR}(\check{F}_C^{(t)}, \check{F}_E^{(t+1)}, E^{(t+1)}), \end{cases} \quad (10)$$

其中 f_{RIGR} 和 f_{ECGR} 分别代表 RIGR 和 ECGR。注意当 $t = 1$ 时， $\check{F}_C^{(1)} = f_{MTFE}(I; \Theta_{MTFE})$ ， $\check{F}_E^{(1)} = f_{MTFE}(I; \Theta_{MTFE})$ 。

2.4. 实现细节

本文提供了 MGL 的两个版本。一种是 S-MGL，它是只挖掘一次互信息的单阶段模型。另一种 R-

表 1: 不同数据库上的定量结果。‘†’ 为 GOD 和 SOD 的最先进方法。↑ (or ↓) 表示越高 (或越低) 越好。评测网页: [link](#).

Methods	CHAMELEON [47]				CAMO-Test [20]				COD10K-Test [9]			
	$S_\alpha \uparrow$	$E_\phi \uparrow$	$F_\beta^w \uparrow$	$M \downarrow$	$S_\alpha \uparrow$	$E_\phi \uparrow$	$F_\beta^w \uparrow$	$M \downarrow$	$S_\alpha \uparrow$	$E_\phi \uparrow$	$F_\beta^w \uparrow$	$M \downarrow$
2017 FPN † [27]	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 † [13]	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 † [69]	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++ † [74]	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 † [31]	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 † [16]	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 PoolNet † [30]	0.776	0.779	0.555	0.081	0.702	0.698	0.494	0.129	0.705	0.713	0.416	0.074
2019 BASNet † [45]	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 † [71]	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 † [59]	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 † [1]	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 † [70]	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 [20]	†	†	†	†	0.682	0.685	0.484	0.126	†	†	†	†
2020 MirrorNet [63]	†	†	†	†	0.741	0.804	0.652	0.100	†	†	†	†
2020 PraNet [10]	0.860	0.898	0.763	0.044	0.769	0.833	0.663	0.094	0.789	0.839	0.629	0.045
2020 SINet [9]	0.869	0.891	0.740	0.044	0.751	0.771	0.606	0.100	0.771	0.806	0.551	0.051
S-MGL (ours)	0.892	0.921	0.803	0.032	0.772	0.850	0.664	0.089	0.811	0.851	0.655	0.037
R-MGL (ours)	0.893	0.923	0.813	0.030	0.775	0.847	0.673	0.088	0.814	0.865	0.666	0.035

MGL 包括执行两个循环阶段的循环学习过程。具体实现如下:

多任务特征提取器。参考已有方法 [9], 本文使用在 ImageNet [19] 上预训练的 ResNet-50 [14] 作为骨干网络。本文使用扩展版本 [65] 以保证 COD 特征 (F_C) 的分辨率为 60×60 。为提取 COEE (F_E) 特征, 本文首先从 ResNet-50 收集一组 side-output 特征 $\{S_k\}_{k=2}^5$, 然后使用双线性上/下采样保证这些特征拥有相同的分辨率 60×60 , 最后用一个 concatenate 层以及一个 1×1 卷积层将它们融合。

区域诱导图推理模块。本文参考 [26] 设计并实现 $f_{G_{proj}}$, 并将 F_C 和 F_E 分别编码至 $K = 32$ 个语义节点 (如表 4 所示)。 f_{CGI} 中的传递函数用 MLPs (1×1 convolution) 实现。在 RIGR 中, 用公式 4 来建立图间的关系, 用公式 5 来获取从 \mathcal{V}_C 传递至 \mathcal{V}_E 的语义指导信息, 并产生用于 \mathcal{V}_E 的演化图表征 V'_E 。 f_{GR} 用 GCNs [18] 实现, $f_{R_{proj}}$ 利用公式 7 复用赋值矩阵实现图逆映射。

边缘约束图推理模块。对于边缘支撑节点的个数, 可以从实验中观察到当 $z = 32$ 时, 模型可以在速度和准确度之间取得较好的平衡 (参考表 4)。公式 8 中的 $h_\phi(\cdot)$ 可以使用元素级减法以及一个 1×1 的卷积简单实现。公式 9 中的 $h_\phi(\cdot)$ 连接边缘和节点, 即 f_i^c & $e_{i,j}$, 再使用一个 1×1 的卷积层融合以产生 $\check{f}_i \in \check{F}_C$ 。

分类器和损失函数。获得演化的表征 $\check{F}_E^{(t)}$ 和 $\check{F}_C^{(t)}$ 之后, 本文使用 1×1 卷积层分类器将它们分别映射到相应的输出 E 和 C。在训练过程中, 本文使用双线性插值将输出概率图上采样至原始图像尺寸以计算损失, 并且, 在两个任务中均使用交叉熵损失 [5]:

$$L = L_{CE}^c(C, G_C) + \gamma L_{CE}^c(E, G_E), \quad (11)$$

其中 G_C 和 G_E 为真值标签, γ 为融合权重。实验中设置 $\gamma = 1$ 。

3. 实验

3.1. 实验设置

数据集: 本文在以下公开数据集上完成实验:

- CHAMELEON [47] 包含 76 张人工逐像素标注好的高分辨率图像。CHAMELEON 数据集中的所有图像都从互联网收集得到。
- CAMO [20] 是一个包含 8 个类别共有 2500 张图像的集合。在这个数据集中, 自然伪装的对象和人工伪装的对象都用精细的标签标注。
- COD10K [9] 是最大的 COD 数据集。它包含 10 个大类以及 78 子类, 共计 10,000 张图像。所有图像均来自于摄影网站。

本文的训练集是由 [7] 提供的 CAMO 和 COD10K 训练集的组合。

评价指标: 参考 [9, 20], 本文使用平均绝对误差 (MAE) 作为评价指标。此外, 平均 E-measure (E_ϕ) [8], S-measure (S_α) [6] 以及加权 F-measure (F_β^w) [37] 也用做评价指标。为了评价辅助任务 COEE, 本文还采用了 ODS 和 OIS 指标 [62]。评测工具参考: <https://github.com/DengPingFan/CODToolbox>。

表 2: 本文方法在 CHAMELEON, CAMO 测试集和 COD10K 测试集上的消融实验。

ResNet-50	Candidate				CHAMELEON [47]				CAMO-Test [20]				COD10K-Test [9]			
	RIGR	ECGR	RL	$S_\alpha \uparrow$	$E_\phi \uparrow$	$F_\beta^w \uparrow$	$M \downarrow$	$S_\alpha \uparrow$	$E_\phi \uparrow$	$F_\beta^w \uparrow$	$M \downarrow$	$S_\alpha \uparrow$	$E_\phi \uparrow$	$F_\beta^w \uparrow$	$M \downarrow$	
✓				0.767	0.799	0.535	0.094	0.742	0.786	0.538	0.130	0.729	0.692	0.436	0.079	
✓	✓			0.844	0.863	0.686	0.055	0.766	0.828	0.611	0.104	0.785	0.758	0.557	0.052	
✓	✓	✓		0.892	0.921	0.803	0.032	0.772	0.850	0.664	0.089	0.811	0.851	0.655	0.037	
✓	✓	✓	✓	0.893	0.923	0.813	0.030	0.775	0.847	0.673	0.088	0.814	0.865	0.666	0.035	

表 3: 不同底层特征增强算法的定量结果。

Method	CAMO-Test [20]				COD10K-Test [9]			
	$S_\alpha \uparrow$	$E_\phi \uparrow$	$F_\beta^w \uparrow$	$M \downarrow$	$S_\alpha \uparrow$	$E_\phi \uparrow$	$F_\beta^w \uparrow$	$M \downarrow$
Baseline (ResNet-50 FCN)	0.742	0.786	0.538	0.130	0.729	0.692	0.436	0.079
Baseline + NL [51]	0.748	0.791	0.541	0.122	0.731	0.711	0.459	0.073
MTFE + MUL [40]	0.751	0.799	0.551	0.118	0.736	0.721	0.498	0.070
S-MGL (ours)	0.772	0.850	0.664	0.089	0.811	0.851	0.655	0.037
R-MGL (ours)	0.775	0.847	0.673	0.088	0.814	0.865	0.666	0.035

训练设置: 在训练过程中, 使用在 ImageNet [19] 预训练的 ResNet-50 [14] 初始化 MTFE, 其余的层或模块随机初始化。本文对所有数据使用数据增强技术, 如随机裁剪、左右翻转以及在 [0.75, 1.25] 范围内缩放。本文使用带‘poly’学习率调整策略的随机梯度下降 (SGD) 算法: $lr = base_lr \times (1 - \frac{iter}{max_iter})^{power}$ 。初始的学习率 $base_lr=10^{-7}$, $power = 0.9$

复现: 本文的 S-MGL and R-MGL 基于 PyTorch 实现。为使用更大的 batch, 模型在 NVIDIA Tesla V100 GPU 上完成训练。在测试阶段, 所有的模型都在 12G 显存的 NVIDIA GTX Titan X GPU 上完成测试。

3.2. 与最先进方法的对比

Baselines/SOTAs: 参考 [9], 首先在相关或相近的领域选择强有力的取得了最先进性能的基准模型, 如 GOD 和 SOD。此外, 近期公开的所有用于 COD 的方法都用于对比。总之, 本文将方法 (S-MGL 和 R-MGL) 与 16 个 SOTAs 进行了比较, 这些方法是在它们推荐的参数设置下, 使用与本文相同的训练集训练得到。

在 CHAMELEON 数据集上的性能: 表 1 报告了在 CHAMELEON 数据集上与 14 个最先进方法的对比。为公平对比, 所有的模型都使用相同的训练集用于训练。可以看出, 和其他所有工作相比, 本文的 S-MGL 在所有指标上都取得了更好的性能。和当前最先进的方法 SINet [9] 相比, S-MGL 显著的降低了 27.3%MAE 并提升了 8.5% F_β^w 。循环学习模型 R-MGL 进一步提升了性能, 并刷新了最新的记

表 4: 不同参数的消融实验影响。‘K’代表语义节点个数; ‘z’代表边缘支撑节点个数; ‘t’是 MGL 模型中循环学习的次数。

Method	CAMO-Test [20]				COD10K-Test [9]			
	$S_\alpha \uparrow$	$E_\phi \uparrow$	$F_\beta^w \uparrow$	$M \downarrow$	$S_\alpha \uparrow$	$E_\phi \uparrow$	$F_\beta^w \uparrow$	$M \downarrow$
S-MGL (K=16, z=32)	0.771	0.832	0.661	0.092	0.805	0.832	0.638	0.042
S-MGL (K=32, z=32)	0.772	0.850	0.664	0.089	0.811	0.851	0.655	0.037
S-MGL (K=64, z=32)	0.774	0.849	0.661	0.089	0.809	0.854	0.648	0.037
S-MGL (K=32, z=16)	0.772	0.843	0.662	0.090	0.804	0.837	0.640	0.040
S-MGL (K=32, z=32)	0.772	0.850	0.664	0.089	0.811	0.851	0.655	0.037
S-MGL (K=32, z=64)	0.773	0.848	0.666	0.089	0.807	0.855	0.657	0.037
R-MGL (K=32, z=32, t=1)	0.772	0.850	0.664	0.089	0.811	0.851	0.655	0.037
R-MGL (K=32, z=32, t=2)	0.775	0.847	0.673	0.088	0.814	0.865	0.666	0.035
R-MGL (K=32, z=32, t=3)	0.773	0.848	0.672	0.088	0.815	0.862	0.661	0.036

录。显然, 本文的解决方案可以明显克服伪装场景中的模糊性, 并提供比现有方法更可靠的结果。

在 CAMO 数据集上的性能: 本文还在 CAMO 测试集上和当前最先进的方法进行了对比。从表 1 中可以看出, 本文的 S-MGL 和 R-MGL 明显优于其他解决方案。这是因为本文的模型可以充分利用交互的益处并确保模型的可靠性以克服复杂场景中的严重遮挡和难以定义边界的问题。

在 COD10K 数据集上的性能: 在最大的数据集 COD10K 测试集上, 本文的解决方案在所有评价指标下取得了最新的记录。特别地, S-MGL 大幅超过了当前最好的模型, 其中 $S_\alpha=81.1\%$, $E_\phi=85.1\%$, $F_\beta^w=65.5\%$, 并且刷新最优的 MAE 至 0.037。R-MGL 进一步提升了性能。强有力的基于图的交互模块使本文的模型能够很好的与辅助任务 COEE 一起工作, 以克服 COD 中的所有挑战。图 5 展示了部分可视化样例。

辅助任务 (COEE): 我们相信本文模型中的交互学习也可以显著地辅助伪装目标边缘检测 (COEE)。为了验证这一点, 本文将 MGL 和著名的 HED [62] 及其改进版本 DSS [15] 做了对比, 此外还比较了强多任务基准 MUL [40]。所有的模型都使用同样的训练集训练。如表 5 所示, 本文的 S-MGL 和 R-MGL 在该任务重取得了比现有模型更好的结果, 这表明

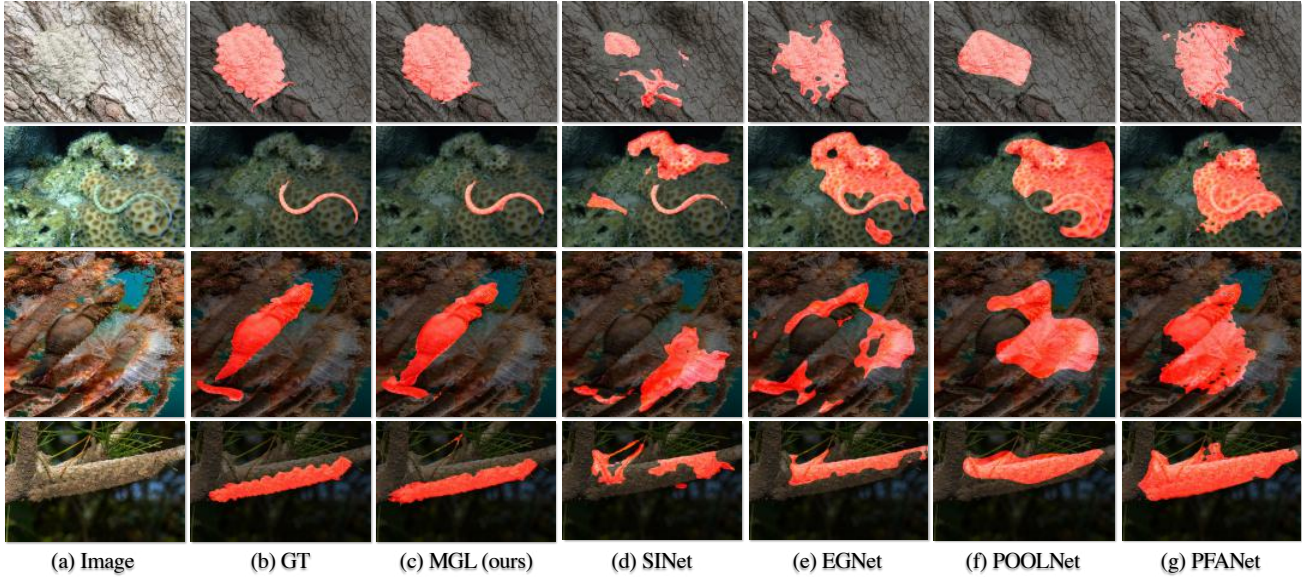


图 5: 不同模型之间的定性对比: (c) 本文的方法 (R-MGL), (d) SINet [9], (e) EGNet [70], (f) POOLNet [30], 和 (g) PFANet [71]。显然, 本文的方法可以更好地发现隐藏的目标以及更明确的边界。

表 5: 伪装目标边缘检测在 CAMO 测试集和 COD10K 测试集上的结果对比。

Method	CAMO-Test [20]		COD10K-Test [9]	
	ODS	OIS	ODS	OIS
HEd [62]	0.315	0.318	0.294	0.313
DSS [15]	0.316	0.336	0.347	0.372
Res50-FCN	0.509	0.511	0.505	0.524
MTEF + MUL [40]	0.521	0.539	0.516	0.534
S-MGL	0.536	0.545	0.535	0.557
R-MGL	0.543	0.551	0.540	0.558

本文的解决方案不仅提高了主任务 (COD) 的性能, 还提高了辅助任务 (COEE) 的性能。图 6 展示了部分可视化样例。

3.3. 消融实验

RIGR 和 ECGR 的有效性: 为了验证 RIGR 的作用, 本文使用 ResNet50-FCN 作为基准网络。首先, 如表 2 所示, 和基准网络在所有的数据集上相比, RIGR 都使模型获得了一定的性能提升, 这证明了 RIGR 的有效性。此外, 加入 ECGR 后, 可以进一步提升准确度。因此, 提高真实边缘可见性的重要性是显而易见的, 可以使模型具有更强的克服 COD 任务所面临困难的能力。此外, 本文还仔细研究了 RIGR 和 ECGR 中不同参数的影响, 不同参数设置的详细对比结果如表 4 所示。

循环学习的作用: MGL 可以很容易的被扩展为一

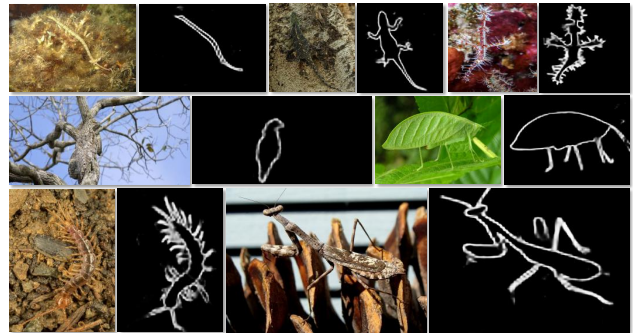


图 6: 本文方法 (R-MGL) 在 COD10K 测试集上对伪装目标边缘检测的可视化结果。

个更全面的循环推理过程。表 2 表明, 使用循环学习技术可以进一步提高模型的性能。这是因为循环过程可以用来细化初始结果/特征, 从而提高准确性。此外, 根据本文的实验 (表 4), 只使用两个循环步骤可以确保足够好的性能, 并且使 R-MGL 为所有数据集刷新了新的记录, 并大幅超过现有的方法。

交互图学习的优越性: 本文进行了全面的实验和比较, 以显示本文所提交互图学习方法的优越性。如表 3 所示, 和广泛使用的 non-local (NL) 操作相比, 显式互学习 (MUL) 可以保证更可靠的结果, 这表明挖掘有价值的辅助边缘信息可以帮助模型克服 COD

任务中严重遮挡和边界不确定等挑战。本文的思想是将 MUL 从规则网格空间拓展至图域。显然，本文的 S-MGL 和 R-MGL 优于常规 MUL，因为它具有更强的捕获高阶关系的能力。这些实验表明，深度挖掘 COD 和辅助 COEE 之间的高阶关系是有意义的，可以显著提高模型的可靠性，更好地克服具有挑战性的 COD 任务的内在模糊性。此外，通过图来推理高阶关系也会带来明显的性能改进。

4. 总结与展望

本文提出了一个基于图的联合学习框架交互图学习 (MGL) 模型来检测伪装目标以及真实边缘。该模型介绍了两个新颖的模块：区域诱导图推理 (RIGR) 模块和边缘约束图推理 (ECGR) 模块，它们可以共同挖掘有价值的互补信息，以提高 COD 的真实边缘可见性。本文还将 MGL 设计为一个循环推理过程，以充分利用所有有用的信息。大量实验表明，明确地挖掘真实边缘先验信息有助于克服 COD 任务重固有的困难，如遮挡和边界不可定义等。另外，MGL 也可以帮助其他相关的计算机视觉任务，例如需要多源信息以联合表征增强的全景分割任务。

References

- [1] Kai Chen, Jiangmiao Pang, Jiaqi Wang, Yu Xiong, Xiaoxiao Li, Shuyang Sun, Wansen Feng, Ziwei Liu, Jianping Shi, Wanli Ouyang, et al. Hybrid task cascade for instance segmentation. In CVPR, 2019. 6
- [2] Lin Chen. Topological structure in visual perception. *Science*, 1982. 1
- [3] Yunpeng Chen, Marcus Rohrbach, Zhicheng Yan, Yan Shuicheng, Jiashi Feng, and Yannis Kalantidis. Graph-based global reasoning networks. In CVPR, June 2019. 3
- [4] IC Cuthill. Camouflage. *Journal of Zoology*, 2019. 1
- [5] Pieter-Tjerk De Boer, Dirk P Kroese, Shie Mannor, and Reuven Y Rubinfeld. A tutorial on the cross-entropy method. *Annals OR*, 2005. 6
- [6] Deng-Ping Fan, Ming-Ming Cheng, Yun Liu, Tao Li, and Ali Borji. Structure-measure: A new way to evaluate foreground maps. In ICCV, 2017. 6
- [7] Deng-Ping Fan, Ge-Peng Ji, Ming-Ming Cheng, and Ling Shao. Concealed object detection. arXiv preprint arXiv:2102.10274, 2021. 1, 6
- [8] Deng-Ping Fan, Ge-Peng Ji, Xuebin Qin, and Ming-Ming Cheng. Cognitive vision inspired object segmentation metric and loss function. SSI, 2021. 6
- [9] Deng-Ping Fan, Ge-Peng Ji, Guolei Sun, Ming-Ming Cheng, Jianbing Shen, and Ling Shao. Camouflaged object detection. In CVPR, 2020. 1, 2, 6, 7, 8
- [10] Deng-Ping Fan, Ge-Peng Ji, Tao Zhou, Geng Chen, Huazhu Fu, Jianbing Shen, and Ling Shao. Pragnet: Parallel reverse attention network for polyp segmentation. In MICCAI, 2020. 1, 6
- [11] Deng-Ping Fan, Tengpeng Li, Zheng Lin, Ge-Peng Ji, Dingwen Zhang, Ming-Ming Cheng, Huazhu Fu, and Jianbing Shen. Re-thinking co-salient object detection. *IEEE TPAMI*, 2021. 2
- [12] Deng-Ping Fan, Tao Zhou, Ge-Peng Ji, Yi Zhou, Geng Chen, Huazhu Fu, Jianbing Shen, and Ling Shao. Inf-net: Automatic covid-19 lung infection segmentation from ct images. *IEEE TMI*, 39(8):2626 – 2637, 2020. 1
- [13] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In ICCV, 2017. 2, 6
- [14] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In CVPR, 2016. 6, 7
- [15] 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, 2017. 7, 8
- [16] Zhaojin Huang, Lichao Huang, Yongchao Gong, Chang Huang, and Xinggang Wang. Mask scoring r-cnn. In CVPR, 2019. 2, 6
- [17] Changku Kang, Martin Stevens, Jong-yeol Moon, Sang-Im Lee, and Piotr G Jablonski. Camouflage through behavior in moths: the role of background matching and disruptive coloration. *Behavioral Ecology*, 2015. 2, 3
- [18] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In ICLR, 2017. 3, 5, 6
- [19] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In NeurIPS, 2012. 6, 7

- [20] Trung-Nghia Le, Tam V. Nguyen, Zhongliang Nie, Minh-Triet Tran, and Akihiro Sugimoto. Anabranched network for camouflaged object segmentation. *CVIU*, 2019. 2, 6, 7, 8
- [21] Aixuan Li, Jing Zhang, Yunqiu Lyu, Bowen Liu, Tong Zhang, and Yuchao Dai. Uncertainty-aware joint salient object and camouflaged object detection. In *CVPR*, 2021. 2
- [22] Xin Li, Fan Yang, Leiting Chen, and Hongbin Cai. Saliency transfer: An example-based method for salient object detection. In *IJCAI*, 2016. 2
- [23] Xin Li, Fan Yang, Hong Cheng, Junyu Chen, Yuxiao Guo, and Leiting Chen. Multi-scale cascade network for salient object detection. In *ACM MM*, 2017. 2
- [24] Xin Li, Fan Yang, Hong Cheng, Wei Liu, and Ding-gang Shen. Contour knowledge transfer for salient object detection. In *ECCV*, 2018. 2
- [25] Xia Li, Yibo Yang, Qijie Zhao, Tiancheng Shen, Zhouchen Lin, and Hong Liu. Spatial pyramid based graph reasoning for semantic segmentation. In *CVPR*, 2020. 3
- [26] Yin Li and Abhinav Gupta. Beyond grids: Learning graph representations for visual recognition. In *NeurIPS*, pages 9225–9235, 2018. 4, 6
- [27] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection. In *CVPR*, 2017. 2, 6
- [28] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *ICCV*, 2017. 2
- [29] Guanghai Liu and Dengping Fan. A model of visual attention for natural image retrieval. In *ICISCCC*, 2013. 1
- [30] Jiang-Jiang Liu, Qibin Hou, Ming-Ming Cheng, Jiashi Feng, and Jianmin Jiang. A simple pooling-based design for real-time salient object detection. In *CVPR*, 2019. 6, 8
- [31] Nian Liu, Junwei Han, and Ming-Hsuan Yang. Picanet: Learning pixel-wise contextual attention for saliency detection. In *CVPR*, 2018. 2, 6
- [32] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg. Ssd: Single shot multibox detector. In *ECCV*, 2016. 2
- [33] Ao Luo, Xin Li, Fan Yang, Zhicheng Jiao, and Hong Cheng. Webly-supervised learning for salient object detection. *Pattern Recognition*, 2020. 2
- [34] Ao Luo, Xin Li, Fan Yang, Zhicheng Jiao, Hong Cheng, and Siwei Lyu. Cascade graph neural networks for rgb-d salient object detection. In *ECCV*, 2020. 3
- [35] Ao Luo, Fan Yang, Xin Li, Dong Nie, Zhicheng Jiao, Shangchen Zhou, and Hong Cheng. Hybrid graph neural networks for crowd counting. In *AAAI*, 2020. 2
- [36] Yunqiu Lyu, Jing Zhang, Yuchao Dai, Li Aixuan, Bowen Liu, Nick Barnes, and Deng-Ping Fan. Simultaneously localize, segment and rank the camouflaged objects. In *CVPR*, 2021. 2
- [37] Ran Margolin, Lihi Zelnik-Manor, and Ayellet Tal. How to evaluate foreground maps? In *CVPR*, 2014. 6
- [38] Haiyang Mei, Ge-Peng Ji, Ziqi Wei, Xin Yang, Xiaopeng Wei, and Deng-Ping Fan. Camouflaged object segmentation with distraction mining. In *CVPR*, 2021. 2
- [39] Abdullhah Mohamed, Kun Qian, Mohamed Elhoseiny, and Christian Claudel. Social-stgcnn: A social spatio-temporal graph convolutional neural network for human trajectory prediction. In *CVPR*, 2020. 2
- [40] Xuecheng Nie, Jiashi Feng, and Shuicheng Yan. Mutual learning to adapt for joint human parsing and pose estimation. In *ECCV*, 2018. 7, 8
- [41] Youwei Pang, Xiaoqi Zhao, Lihe Zhang, and Huchuan Lu. Multi-scale interactive network for salient object detection. In *CVPR*, 2020. 2
- [42] Ricardo Pérez-de la Fuente, Xavier Delclòs, Enrique Peñalver, Mariela Speranza, Jacek Wierzechos, Carmen Ascaso, and Michael S Engel. Early evolution and ecology of camouflage in insects. *PNAS*, 2012. 1
- [43] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *CVPR*, 2017. 4
- [44] Xuebin Qin, Deng-Ping Fan, Chenyang Huang, Cyril Diagne, Zichen Zhang, Adrià Cabeza Sant’Anna, Albert Suárez, Martin Jagersand, and Ling Shao. Boundary-aware segmentation network for mobile and web applications. *arXiv preprint arXiv:2101.04704*, 2021. 2
- [45] Xuebin Qin, Zichen Zhang, Chenyang Huang, Chao Gao, Masood Dehghan, and Martin Jagersand. Bas-

- net: Boundary-aware salient object detection. In CVPR, 2019. 2, 6
- [46] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In CVPR, 2016. 2
- [47] P Skurowski, H Abdulameer, J Błaszczyk, T Depta, A Kornacki, and P Koziel. Animal camouflage analysis: Chameleon database. Unpublished Manuscript, 2018. 2, 6, 7
- [48] Martin Stevens and Sami Merilaita. Animal camouflage: current issues and new perspectives. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 2009. 1
- [49] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In CVPR, 2016. 2
- [50] Lei Wang, Yuchun Huang, Yaolin Hou, Shenman Zhang, and Jie Shan. Graph attention convolution for point cloud semantic segmentation. In CVPR, 2019. 2
- [51] Xiaolong Wang, Ross Girshick, Abhinav Gupta, and Kaiming He. Non-local neural networks. In CVPR, 2018. 4, 7
- [52] Yue Wang, Yongbin Sun, Ziwei Liu, Sanjay E Sarma, Michael M Bronstein, and Justin M Solomon. Dynamic graph cnn for learning on point clouds. *TOG*, 2019. 2, 5
- [53] Richard J Webster. Does disruptive camouflage conceal edges and features? *Current Zoology*, 2015. 2, 3
- [54] Richard J Webster, Christopher Hassall, Chris M Herdman, Jean-Guy J Godin, and Thomas N Sherratt. Disruptive camouflage impairs object recognition. *Biology Letters*, 2013. 2, 3
- [55] Jun Wei, Shuhui Wang, Zhe Wu, Chi Su, Qingming Huang, and Qi Tian. Label decoupling framework for salient object detection. In CVPR, 2020. 2
- [56] Xin Wei, Ruixuan Yu, and Jian Sun. View-gcn: View-based graph convolutional network for 3d shape analysis. In CVPR, 2020. 2
- [57] Yangxin Wu, Gengwei Zhang, Yiming Gao, Xiajun Deng, Ke Gong, Xiaodan Liang, and Liang Lin. Bidirectional graph reasoning network for panoptic segmentation. In CVPR, 2020. 3
- [58] Yu-Huan Wu, Shang-Hua Gao, Jie Mei, Jun Xu, Deng-Ping Fan, Rong-Guo Zhang, and Ming-Ming Cheng. Jcs: An explainable covid-19 diagnosis system by joint classification and segmentation. *IEEE TIP*, 30:3113–3126, 2021. 1
- [59] Zhe Wu, Li Su, and Qingming Huang. Cascaded partial decoder for fast and accurate salient object detection. In CVPR, 2019. 6
- [60] Guo-Sen Xie, Jie Liu, Huan Xiong, and Ling Shao. Scale-aware graph neural network for few-shot semantic segmentation. In CVPR, 2021. 2
- [61] Guo-Sen Xie, Li Liu, Fan Zhu, Fang Zhao, Zheng Zhang, Yazhou Yao, Jie Qin, and Ling Shao. Region graph embedding network for zero-shot learning. In *ECCV*, 2020. 2
- [62] Saining Xie and Zhuowen Tu. Holistically-nested edge detection. In *ICCV*, 2015. 6, 7, 8
- [63] Jinnan Yan, Trung-Nghia Le, Khanh-Duy Nguyen, Minh-Triet Tran, Thanh-Toan Do, and Tam V Nguyen. Mirrornet: Bio-inspired adversarial attack for camouflaged object segmentation. *arXiv*, 2020. 2, 6
- [64] Han Yang, Xingjian Zhen, Ying Chi, Lei Zhang, and Xian-Sheng Hua. Cpr-gcn: Conditional partial-residual graph convolutional network in automated anatomical labeling of coronary arteries. In CVPR, 2020. 2
- [65] Fisher Yu and Vladlen Koltun. Multi-scale context aggregation by dilated convolutions. In *ICLR*, 2015. 6
- [66] Qiang Zhai, Xin Li, Fan Yang, Chenglizhao Chen, Hong Cheng, and Deng-Ping Fan. Mutual graph learning for camouflaged object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021. 1
- [67] Hang Zhang, Jia Xue, and Kristin Dana. Deep ten: Texture encoding network. In CVPR, 2017. 4
- [68] Yaobin Zhang, Weihong Deng, Mei Wang, Jiani Hu, Xian Li, Dongyue Zhao, and Dongchao Wen. Global-local gcn: Large-scale label noise cleansing for face recognition. In CVPR, 2020. 2, 3
- [69] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. Pyramid scene parsing network. In CVPR, 2017. 2, 6
- [70] Jia-Xing Zhao, Jiang-Jiang Liu, Deng-Ping Fan, Yang Cao, Jufeng Yang, and Ming-Ming Cheng. Egnnet:

- Edge guidance network for salient object detection. In ICCV, 2019. [2](#), [6](#), [8](#)
- [71] Ting Zhao and Xiangqian Wu. Pyramid feature attention network for saliency detection. In CVPR, 2019. [6](#), [8](#)
- [72] Xiaoqi Zhao, Youwei Pang, Lihe Zhang, Huchuan Lu, and Lei Zhang. Suppress and balance: A simple gated network for salient object detection. In ECCV, 2020. [2](#)
- [73] Huajun Zhou, Xiaohua Xie, Jian-Huang Lai, Zixuan Chen, and Lingxiao Yang. Interactive two-stream decoder for accurate and fast saliency detection. In CVPR, 2020. [2](#)
- [74] 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. [6](#)
- [75] Mingchen Zhuge, Deng-Ping Fan, Nian Liu, Dingwen Zhang, Dong Xu, and Ling Shao. Salient object detection via integrity learning. arXiv preprint arXiv:2101.07663, 2021. [2](#)