一种基于深度梯度学习的高效伪装目标检测方法

季葛鹏[1], 范登平✉[2], 周昱程[1], 代登信[3], Alexander Liniger[2] and Luc Van Gool[2]

[1]计算机学院, 武汉大学, 武汉, 中国.
[2]计算机视觉实验室, 苏黎世联邦理工学院, 苏黎世, 瑞士.
[3]视觉自动化系统组, 马克思普朗克信息研究所, 萨尔布吕肯, 德国.

Abstract

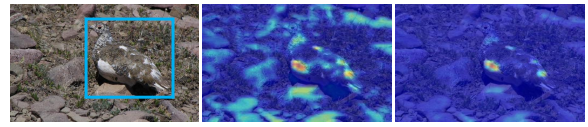
本文提出了**DGNet**模型，一种利用对象梯度监督实现伪装对象检测（COD）的新型深度框架。它将任务解偶成两个相关联的分支，即语义编码器和纹理编码器。分支之间的核心关联是梯度诱导转换，用来表征语义特征和纹理特征之间的软分组。得益于简单而高效的结构设计，DGNet模型大幅地超越了现有的前沿伪装目标检测模型。特别地，本文的高效版本DGNet-S模型取得了实时推理速度（80 fps）且仅有最好模型JCSOD-CVPR$_{21}$ 6.82%的参数量。应用结果也显示DGNet模型在息肉分割、缺陷检测和透明目标分割任务上取得了优异的性能。本文的源码可在https://github.com/GewelsJI/DGNet中获取。

Keywords: 伪装目标检测、目标梯度、软分组、高效模型和图像分割

1 引言

伪装目标检测 [1, 2]（Camouflaged Object Detection, COD）旨在分割具有人工或者自然模式的目标，这些目标能够“完美地”融入背景之中，以避免被发现 [2]。一些成功应用已展现出伪装目标检测任务的科学和工业价值，例如：医疗图像分析（即息肉 [3–5]和肺部感染分割 [6–8]）、视频理解（例如：运动分割 [9]、视频监控 [10]和自动驾驶 [11]）和休闲艺术 [12, 13]。

近期的一些研究工作 [1, 2, 15–17]在基于完整的目标级别真值掩膜监督之下展现出了卓越的性能。随后，各类前沿的技术被开发用于增强伪装目标检测的底层表征，例如：基于边界 [18, 19]和基于不确定性引导的 [20, 21]。然而，从边界监督或基于不确定性的模型中学习到的特征，通常会对伪装目标的稀疏边缘做出响应，从而引入噪声特征，特别是对于复杂场景而言（见图 1-a）。此外，伪装目标的边缘通常“难以定义”或“不明确”，因而不会从快速视觉扫描的过程中被弹出。本文注意到，尽管目标具有伪装性，但仍然留下



伪装图像　(a) 目标边缘　(b) 目标梯度

图1 纹理特征可视化。本文观察到DGNet-S模型在目标边缘的监督(a)下特征图背景中含有扩散的噪声。相比之下，基于目标梯度的监督(b)使网络更关注强度剧烈变化的区域。

✉通讯作者(dengpfan@gmail.com)。论文的主体部分是季葛鹏实习期间在范登平的指导下完成的。本文为MIR2022期刊论文 [14]的中文翻译稿。

一些线索，如图 1第一列中的白色斑纹。本文所感兴趣的不是仅提取边界或不确定区域，而是网络如何挖掘物体内部的“鉴别性模式”。

从这个角度出发，本文提出一种深度梯度网络（Deep Gradient Network, **DGNet**），它采用目标级别梯度图进行显式监督。其中，潜在假设是伪装目标内部具有一些像素强度变化。为了简化学习任务，本文将DGNet模型解耦为两个相关联的分支，即：语义编码器和纹理编码器。前者可视为上下文语义学习器，后者则是结构纹理提取器。通过这种方法，可以克服从单个分支中所提取高级特征和低级特征之间的特征歧义。为充分聚合两个分支生成的两类鉴别性特征，本文进一步设计了梯度诱导转换（Gradient-Induced Transition, GIT）模块，以协同的方式集成了不同分组尺度下的多源特征空间（即软分组策略）。在图 1-b中，DGNet模型采用一种聚焦于伪装对象内部区域的像素强度敏感策略，可在抑制背景噪声的同时检测纹理模式。

在三个具有挑战性的伪装目标检测基准上所进行的充分实验表明，本文的DGNet模型在不引入任何复杂结构的情况下，实现了最前沿的性能。此外，本文实现了一个仅有8.3M参数量的高效模型DGNet-S，它在伪装目标检测相关的基线模型对比中取得了最快的推理速度（80 fps）。值得注意的是，DGNet-S仅有最佳模型JCSOD-CVPR21 [21] 参数量的6.82%，且实现了相当的性能表现。上述结果表明，本文的模型大大缩小了科学研究与实际应用之间的差距。DGNet模型的三个下游应用（请参见第5节）也支撑了这个结论。本文的主要贡献可以归纳为：

- 引入了一个新颖的深度梯度学习框架用于解决伪装目标检测任务，名为**DGNet**模型。
- 提出了**梯度诱导转换**，它根据软分组策略对来自语义分支和纹理分支的特征进行自动分组。
- 展示了**三个下游应用**并取得了良好性能，包括息肉分割、缺陷检测和透明目标分割。

2 相关工作

传统的方法通过提取各种伪装目标和背景之间的手工特征，例如计算三维凸度 [22]、共生矩阵 [23]、最大期望估计 [24]、光流 [25]和高斯混合模型 [26]。这类方法在简单背景下表现良好，但在复杂背景下性能则会大幅缩减。

基于卷积神经网络的模型大概可以被分为三个类别：*a*）基于注意力的策略： Sun等人 [27, 28]提出了一个注意力诱导的跨级融合模块来整合多尺度特征，以及一个双分支的全局语义模块来挖掘多尺度语义信息。为了模仿猎食者的狩猎过程，Mei等人 [15]提出了PFNet模型，它包含定位模块以及聚焦模块来进行伪装识别。文献[19, 29]设计了一个巧妙的框架，它包含特征的协方差矩阵以及多变量矫正模块来提升模型的鲁棒性。Kajiura等人 [30]通过探索伪边界和伪掩膜的不确定性改进了检测的准确率。Zhuge等人 [31]提出了一个类似立方体的架构，它结合了注意力融合和X形连接以充分融合多层特征。*b*）双阶段策略： 搜索与识别策略 [1]是伪装目标检测的早期工作。在文献[2]中，近邻连接解码器和分组反向注意力被引入SINet模型 [1]中，用以进一步提升性能。*c*）联合学习策略： ANet模型 [32]是一个早期的尝试，它利用分类和分割的融合方案来实现伪装目标检测。近来，LSR模型 [17]和JCSOD模型 [21]重构了联合学习框架，引入伪装排序或者从显著对象到伪装对象的学习。ZoomNet模型 [33]是一个引入了混合尺度的三元组网络，它使用缩放策略来学习鉴别性的伪装语义。

基于Transformer和基于图的模型为近期的两种技术趋势。近期，Mao等人 [16]引入基于Transformer的困难等级感知学习概念，用于伪装目标和显著目标检测。UGTR模

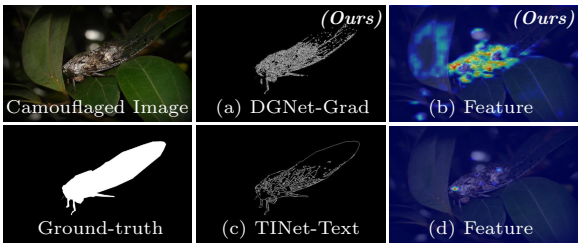


图2 与TINet模型 [34]中提出的纹理标签相比，本文的目标梯度标签(a)在伪装目标中保留了更多的几何线索。由于稀疏像素的数据分布不平衡(例如：细小目标边缘)，DGNet模型在纹理标签(c)的监督下无法推断注意力区域(d)。值得注意的是，这种改进使DGNet模型具有更可靠的辅助约束，如(b)中的特征所示。

型 [20]在Transformer框架下显式地利用概率表征模型来学习伪装对象中的不确定性。此外，Cheng等人 [35]收集了第一个用于伪装视频目标检测的数据集，并运用了基于Transformer的框架来利用短期动态和长期时序一致性来检测动态的伪装目标。随后，Zhai等人 [18]设计了互图学习模型，该模型将输入解耦到不同的特征中，以粗略定位目标并准确捕获其边界。

批注： 本文通过学习目标级别的梯度来挖掘纹理信息，而不是仅仅使用边缘信息或不确定性来建模。这其中的生物学灵感来源于伪装物体内丰富的梯度线索值得被探索，而稀疏的边缘线索却不足以实现这一点。如图 2所示，本文还注意到近期的工作 [34]试图利用纹理线索，然而该工作丢弃了由于Canny边缘检测器中不同阈值设定所带来丰富的目标梯度线索。简而言之，本文旨在设计一个优雅的框架，以更简洁的思想（即目标梯度学习）实现高效的伪装目标检测。更多的实验验证请参见第4.3节。

3 深度梯度网络

如文献[36]所讨论的，低级特征和高级特征在场景理解中占据着同等地位。文献[37]建议，不鼓励对这两种特征进行同时编码。如图 3所示，本文提出使用两个独立的编码器即：语义编码器和纹理编码器对伪装表征进行建模。

3.1 语义编码器

给定一张伪装图像$\mathcal{I} \in \mathbb{R}^{3\times H\times W}$，本文采用广泛使用的EfficientNet [38]作为语义编码器来获取金字塔特征$\{\mathbf{X}_i\}_{i=1}^{5}$。

维度缩减： 受文献[3]启发，本文采用如下两个步骤来确保解码阶段不同层级之间高效的逐元素操作: *a*）仅使用高三层语义特征（即当$i=3,4,5$时），它保留了视觉场景中丰富的语义。*b*）进一步利用两个具有$C_i\times 3\times 3$卷积核大小的ConvBR层[1]，用于将每个候选特征维度降低到C_i，有助于降低后续操作的计算负担。最后生成了三个语义特征$\{\mathbf{X}_i^R\}_{i=3}^{5} \in \mathbb{R}^{C_i\times H_i\times W_i}$，其中$C_i$、$H_i=\frac{H}{2^i}$和$W_i=\frac{W}{2^i}$分别表示特征的通道数目、高度和宽度。

3.2 纹理编码器

本文还引入了一个由目标级别梯度监督的定制纹理分支，弥补了高三层语义特征对几何纹理表征能力弱所导致的模式退化问题。

目标梯度的生成： 图像梯度用于描述图像强度的方向变化或者相邻像素位置之间的颜色变化，它被广泛应用于边缘检测 [41]和超分辨率 [42]任务之中。图 3的右侧展示了四种广泛使用的标签。通过计算真值$\mathbf{Z}^C$（b）和原始图像（a）的梯度，分别生成目标边界（c）和图像梯度（e）。然而，用含有不相关背景噪声的原始图像梯度图（e）作为纹理学习的监督信号，可能会误导优化过程。为了缓解这一问题，本文引入了一种新的伪装学习范式，将目标级别梯度$\mathbf{Z}^G$（d）作为监督，它包含了目标边界和内部区域的梯度线索。

[1]本文中，ConvBR表示标准的卷积层，其后接有BN层 [39]和ReLU层 [40]。

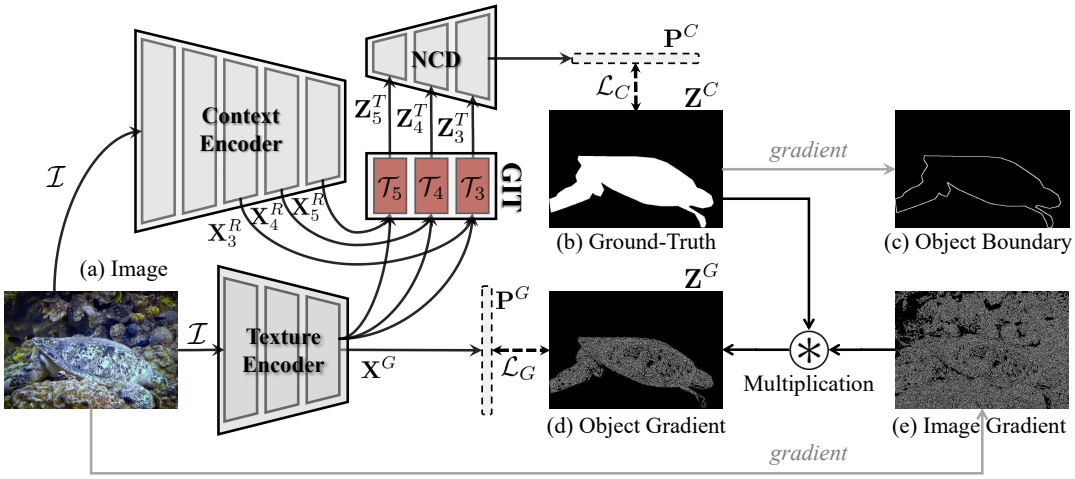


图3 DGNet模型的框架流程图。它由两个相关联的学习分支组成，即语义编码器（请参见第3.1节）和纹理编码器（请参见第3.2节）。然后引入梯度诱导转换（GIT）（请参见第3.3节）以协作的方式聚合来自上述两个编码器的特征。最后，采用近邻连接解码器（NCD）[2]生成预测$\mathbf{P}^C$（请参见第3.4节）。

表1 定制纹理编码器的细节。k代表卷积核的大小，c代表输出通道数目，s代表层级的步长大小，p代表零填充数目。本文将通道默认设置为$C_g = 32$。

层	输入大小	输出大小	Component	k	c	s	p
#01	$3 \times H \times W$	$64 \times \frac{H}{2} \times \frac{W}{2}$	ConvBR	7	64	2	3
#02	$64 \times \frac{H}{2} \times \frac{W}{2}$	$64 \times \frac{H}{4} \times \frac{W}{4}$	ConvBR	3	64	2	1
#03	$64 \times \frac{H}{4} \times \frac{W}{4}$	$C_g \times \frac{H}{8} \times \frac{W}{8}$	ConvBR	3	C_g	2	1
#04	$C_g \times \frac{H}{8} \times \frac{W}{8}$	$1 \times \frac{H}{8} \times \frac{W}{8}$	ConvBR	1	1	1	0

该过程可以表示为：

$$\mathbf{Z}^G = \mathcal{F}_E(\mathcal{I}(x,y)) \circledast \mathbf{Z}^C, \tag{1}$$

其中$\mathcal{F}_E$ 表示标准的Canny边缘检测器 [43]，输入为带有离散像素坐标(x,y) 的原始图像$\mathcal{I}$ 。$\circledast$ 表示逐像素乘法运算。

纹理编码器： 由于具有高分辨率的低层特征会引入不小的计算负担，本文设计了定制的轻量级编码器，并非使用开箱即用的骨架网络。本文从#03 层（请参见表 1）中获取纹理特征$\mathbf{X}^G \in \mathbb{R}^{C_g \times H_g \times W_g}$。同时，使用目标梯度$\mathbf{Z}^G$ 对随后的#04层进行监督。本文保留了较大分辨率的纹理特征（即：$H_g = \frac{H}{8}$和$W_g = \frac{W}{8}$），因为较小分辨率的特征丢弃了大部分细节。

3.3 梯度诱导转换

语义特征和纹理特征之间的潜在关联性为自适应聚合提供了巨大的潜力，而不是采用普通的聚合策略（例如：拼接和相加操作)。在此，本文设计了一个灵活的即插即用的梯度诱导转换（*GIT*）模块（请参见图 4），它以分组的角度使用纹理特征来辅助多源聚合过程。具体而言，它包括如下三个步骤。

梯度诱导的特征分组： 受文献[2]的启发，首先使用梯度诱导分组学习策略将三个语义特征$\{\mathbf{X}_i^R\}_{i=3}^5$ 和一个纹理特征$\mathbf{X}^G$ 沿通道切割为固定的分组。对于每个$\mathbf{X}_i^R$ 和$\mathbf{X}^G$特征对，该策略可以被表述为：

$$\begin{aligned} \{\mathbf{X}_{i,m}^R\}_{m=1}^M \in \mathbb{R}^{K_i \times H_i \times W_i} &\leftarrow \mathbf{X}_i^R \in \mathbb{R}^{C_i \times H_i \times W_i}, \\ \{\mathbf{X}_m^G\}_{m=1}^M \in \mathbb{R}^{K_g \times H_g \times W_g} &\leftarrow \mathbf{X}^G \in \mathbb{R}^{C_g \times H_g \times W_g}, \end{aligned} \tag{2}$$

其中$\leftarrow$ 是特征分组操作。$K_i = C_i/M$ 和$K_g = C_g/M$ 分别表示每个特征组的通道数，M表示其对应组的数目。然后周期性地将语义特征$\mathbf{X}_{i,m}^R$ 和

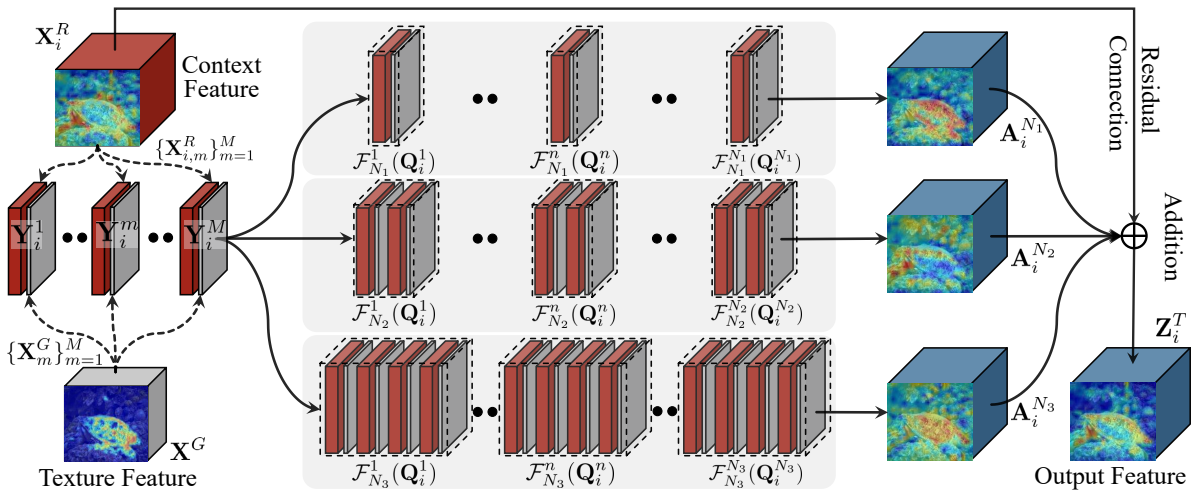


图4 梯度诱导转换（GIT）的图示。

纹理特征$\mathbf{X}_m^G$ 进行重排列，并通过：

$$\mathbf{Q}_i \in \mathbb{R}^{(C_i+C_g)\times H_i\times W_i} = \langle \mathbf{Y}_i^1;\ \ldots;\ \mathbf{Y}_i^m;\ \ldots;\ \mathbf{Y}_i^M \rangle, \tag{3}$$

得到重组后的特征$\mathbf{Q}_i$，其中$\langle\cdot;\ \cdot\rangle$ 表示通道拼接操作。在此，第m个子部分$\mathbf{Y}_i^m$来源于：

$$\mathbf{Y}_i^m \in \mathbb{R}^{(K_i+K_g)\times H_i\times W_i} = \langle \mathcal{F}_{\downarrow}(\mathbf{X}_m^G);\ \mathbf{X}_{i,m}^R \rangle, \tag{4}$$

其中,下采样函数$\mathcal{F}_{\downarrow}(\cdot)$ 用于确保$\mathbf{X}_m^G$的空间分辨率大小与$\mathbf{X}_{i,m}^R$相匹配。

软分组策略： 由于缺乏进一步的多源交互，朴素的特征聚合策略可能会忽略语义表征和纹理表征之间的相关性或者差异性。受到在 [44]中用于并行地捕捉多尺度目标的设计启发，本文提出使用软分组策略在多个不同细粒度的子空间中，提供并行的非线性映射关系，这使得网络能够联合探索多源表征。具体而言，本文实验中设置了三个并行子分支$\{N_1, N_2, N_3\}$（即图 4中的灰色区域）用于软分组。为了简化说明，本文省略了符号下标并以其中的第N个子分支为例，表示为：

$$\mathbf{A}_i^N = \langle \mathcal{F}_N^1(\mathbf{Q}_i^1);\ \ldots;\ \mathcal{F}_N^n(\mathbf{Q}_i^n)\ldots\mathcal{F}_N^N(\mathbf{Q}_i^N) \rangle, \tag{5}$$

其中$\mathcal{F}_N^n(\mathbf{Q}_i^n) \in \mathbb{R}^{(C_i/N)\times H_i\times W_i} = f_n(\mathbf{Q}_i^n, \omega_n)$用于在每个多源子空间中引入软非线性。映射函数$f_n$ 由C_i个具有$\frac{(C_i+C_g)}{N} \times 1 \times 1$卷积核大小标准卷积层所实现，它由$\omega_n$进行参数化。$\mathbf{Q}_i^n$代表重组特征$\mathbf{Q}_i$中的第n个子集，它被切分成$N$ 组。

并行残差学习： 本文进而以并行方式在不同的分组尺度上引入残差学习 [45]。因此将GIT函数$\mathcal{T}_i(\cdot,\cdot)$ （参见图 3中的红色块）定义为：

$$\mathbf{Z}_i^T = \mathcal{T}_i(\mathbf{X}_i^R, \mathbf{X}^G) = \mathbf{X}_i^R \oplus \sum_N \mathbf{A}_i^N, \tag{6}$$

其中$N \in \{N_1, N_2, N_3\}$ 表示不同的组缩放因子，这个参数设定将在第4.3节中讨论。$\oplus$ 表示逐元素相加操作，$\sum$ 表示多项和操作。最后的输出为$\{\mathbf{Z}_i^T\}_{i=3}^5 \in \mathbb{R}^{C_i\times H_i\times W_i}$。

3.4 学习细节

解码器： 给定语义特证$\{\mathbf{X}_i^R\}_{i=3}^5$，首先使用GIT函数$\mathcal{T}_i(\cdot,\cdot)$（请参见公式(6)）来获得输出特征$\{\mathbf{Z}_i^T\}_{i=3}^5$。为了更有效地利用上述的梯度诱导特征$\mathbf{Z}_i^T$，本文使用近邻连接解码器(NCD) [2]生成最终预测，它支持从高层到低层的特征传播。因此最终的伪装预测$\mathbf{P}^C$由$\mathbf{P}^C \in \mathbb{R}^{1\times H\times W} = \mathrm{NCD}(\mathbf{Z}_3^T, \mathbf{Z}_4^T, \mathbf{Z}_5^T)$ 生成。

表2 DGNet-S 和 DGNet模型的超参数设定。

模型	骨架网络	C_i	C_g	M	$\{N_1, N_2, N_3\}$
DGNet-S	EfficientNet-B1	32	32	8	$\{8, 16, 32\}$
DGNet	EfficientNet-B4	64	32	8	$\{4, 8, 16\}$

损失函数： 总体优化目标可定义为：

$$\mathcal{L} = \mathcal{L}_C(\mathbf{P}^C, \mathbf{Z}^C) + \mathcal{L}_G(\mathbf{P}^G, \mathbf{Z}^G), \quad (7)$$

其中，$\mathcal{L}_C$和$\mathcal{L}_G$ 分别表示分割损失函数和目标梯度损失函数。前者定义为$\mathcal{L}_C = \mathcal{L}^w_{IoU} + \mathcal{L}^w_{BCE}$，其中，$\mathcal{L}^w_{IoU}$ 和$\mathcal{L}^w_{BCE}$ 分别表示加权交并比损失和加权二值交叉熵损失。它会根据像素难易程度为每个像素点自适应地分配权值，用以关注全局结构并更多地关注困难像素点。这些损失的定义与 [1, 2, 46]中相同，其有效性在二值分割领域中已得到验证。对于后者，本文使用标准均方误差损失函数。

训练配置： DGNet模型在PyTorch [47]和Jittor [48]框架中实现，并在单块NVIDIA RTX TITAN GPU上进行训练与推理。本文使用文献[49]的策略进行模型参数初始化，并使用在ImageNet [50]上预训练模型进行骨架网络的初始化，以防止过拟合。本文丢弃EfficientNet [38]骨架模型中Conv1×1层、池化层和全连接层，并提取其高三层的侧输出特征，包括`stage-4`（$\mathbf{X}_3$）、`stage-6`（$\mathbf{X}_4$）和`stage-8`（$\mathbf{X}_5$）。考虑到性能与效率的均衡，本文实例化了两个模型实例，以适应不同计算开销下的特定需求（请参见表 2）。

本文使用Adam优化器以端到端的方式训练模型。使用SGDR策略 [51]的余弦退火部分来调整学习率，其中最小/最大学习率和最大调整迭代分别设定为$10^{-5}/10^{-4}$和20。批大小设定为12，最大迭代周期设置为100。在训练期间，本文将每张输入图像大小调整为352×352，并使用了四种数据增强方法：色彩增强、随机翻转、随机裁剪和随机旋转。最后，DGNet模型和DGNet-S 分别需要8.8小时和7.9小时达到模型训练收敛。

测试配置： 当网络完成训练，本文将输入图像的大小调整为352×352，并在三个未知的测试数据集上测试DGNet-S 和DGNet。我们将最终输出的$\mathbf{P}^C$ 作为预测图，不使用任何如DenseCRF [52]的后处理技术。

4 实验

4.1 测评基准

数据集： 伪装目标检测领域中有三个常见数据集：*a*）CAMO [32] 数据集有1,250张伪装图像，并被切分成CAMO-`Tr`（1,000张样本）和CAMO-`Te`（250张样本）。*b*）COD10K [2] 是目前最大的伪装目标检测数据集，它包含了COD10K-`Tr`（3,040张样本）和COD10K-`Te`（2,026张样本）。其图像下载自多个免费摄影网站，涵盖了5个父类以及69个子类。*c*）NC4K-`Te` [17] 作为最大的测试数据集，包含了4,121张图像，用来评测现有模型的泛化能力。遵循文献[2]中的协议，本文在4,040张图像的混合训练集（即：COD10K-`Train` + CAMO-`Train`）上进行模型训练，并在三个基准测评上评测本文方法（请参见表 3）。

指标： 遵循文献[2]，本文使用五个常用的指标来评测：结构指标（$\mathcal{S}_\alpha$） [53]、增强匹配指标（E_ϕ） [54, 55]、F指标（F_β） [56, 57]、加权F指标（F^w_β） [58]和平均绝对误差（$\mathcal{M}$）。此外，通过改变阈值范围$[0, 255]$，本文进一步得到准确率-召回率（PR）曲线 [56]、F指标曲线和E指标曲线。再者，本文采用三个指标来衡量模型的复杂性[2]和效率：模型参数量（Para）用百万单

[2]模型的参数和MACs由此代码库来评测:https://github.com/sovrasov/flops-counter.pytorch.

位M来衡量，乘积累加运算次数（MACs）用吉咖单位G来衡量，推理速度用每秒帧数（fps）。

对比模型： 本文和其他20个前沿模型（见表 3）进行对比，含8个显著目标检测模型和12个伪装目标检测模型。为公平比较，结果均来源于网站或使用原始设定在相同训练集上重新训练得来。

4.2 结果和分析

定量结果： 如表 3所示，DGNet模型在所有指标对比中取得了优越的性能。特别地，基于梯度的学习策略提升了预测图的完整性，并且在CAMO-Te数据集的F_β^w指标上以2.6%的差距超越了排名第一的SINetV2模型[2]。

定量曲线： 如图 5所示，本文通过不同的阈值来绘制所有与伪装目标检测相关的对比模型的精确率-召回率（第1行）、F指标（第2行)和E指标（第3行）曲线。在三个数据集上的对比表明，本文的紫红色实线/虚线的曲线明显优于其他方法。

定性结果： 四个顶尖的伪装目标检测基线模型和DGNet模型的视觉对比请参见图 6。有趣的是，这些对比模型无法在目标与图像边界重合时提供完整的检测结果。相反，由于本文的方法使用了梯度学习策略，可以准确地定位目标区域并提供精确的预测结果。

效率分析： 为了更好地揭示本文方法的效率，相较于现有的对比模型而言，本文的两个模型实例始终取得了最佳的权衡（请参见图 7）。DGNet模型大幅（F_β^w指标: +2.6%）超越了排名第一的SINetV2模型 [2]。值得注意的是，高效的模型实例DGNet-S 以低于排名第二模型JCSOD [21]113.33M参数量的情况下，超越其性能表现。此外，本文在表 4中提供所有伪装目标检测相关模型的推理速度，其在NVIDIA RTX TITAN显卡上进行测试。这清楚地展现了DGNet-S模型和DGNet模型能达到超越实时的推理速度（即：80 fps和58 fps）。

4.3 消融实验

本节进一步对核心模块进行消融实验来验证每个模块和配置的有效性。由于资源原因，本节仅对DGNet-S模型进行消融实验。

基础网络的作用： 如表 5-(a)所示，本节移除了DGNet-S 的纹理编码器和GIT，并称为基础网络（#01）。相较于#01模型，本文的DGNet-S模型 (#**S**) 仅略微增加了0.06M的模型参数量，但在性能上有大幅提升。

维度缩减的配置： 改变通道数C_i为16（#02）、32（#**S**）、64（#03）和128 (#04)，发现模型参数量越大可能会导致性能饱和。为实现计算资源和速度上的权衡，本文选择$C_i = 32$ 作为默认配置。

网络解耦策略的作用： 本节探索了解耦策略的必要性。受文献 [19]的启发，本节将从纹理编码器中提取的特征替换成从语义编码器中提取的低层特征$\mathbf{X}_2$，用以构建单流网络（#05）。值得注意的是，本节只改变纹理特征的提取方式，并保留了两个变体的梯度监督（即：#05 和#**S**），以确保消融实验的公平性。如表 5-(c)所示，将模型解耦成双流网络能提升性能（在CAMO-Te数据集的F_β^w指标上提升了5.3%），这得益于独立分支建模方式能避免不同层级特征中的二义性。

目标梯度监督的作用： 本节将梯度图$\mathbf{Z}^G$（#**S**）替换成边缘图$\mathbf{Z}^B$（#06）以监督语义学习过程。基于梯度图的监督所带来的性能提升（在CAMO-Te数据集的F_β^w指标上提升了1.7%）展示了其有效性。图 8第一行代表在不同监督信号下从纹理学习分支中所提取到的低层特征，它说明了本文方法能使得模型去捕捉伪装目标内部的梯度敏感信息，这类像素点更倾向于引发观察者的关注。

本文进一步使用基于纹理标签的监督 [34]（参见图 2）进行消融实验，表 6中的结果表明本文的梯度监督方式（即：w/ DGNet-Grad）比纹理监督方式（即：w/ TINet-Text）更

表3 针对前沿模型的全指标量化评测结果，包含8个显著目标检测模型和12个伪装目标检测模型。@$\mathcal{R}$表示当前指标的排名，Mean@$\mathcal{R}$代表所有指标的平均排名。注意，E_ϕ^{mx}/F_β^{mx}、E_ϕ^{mn}/F_β^{mn}和E_ϕ^{ad}/F_β^{ad}分别代表增强匹配指标和F指标的最大值、平均值和自适应值。↑/↓代表分数越高/低，效果越好。所有的评测结果图可以在OneDrive-3.54GB中获取到。

		SOD-related Models								COD-related Models													
	Metric	EGNet [59]	SCRN [60]	CPD [61]	CSNet-R [62]	F3Net [46]	UCNet [63]	ITSD [64]	MINet [65]	SINet [1]	PraNet [3]	BAS [66]	C2FNet [27]	TINet [34]	UGTR [20]	PFNet [15]	S-MGL [18]	R-MGL [18]	LSR [17]	JCSOD [21]	SINetV2 [2]	**DGNet-S** *Ours*	**DGNet** *Ours*
NC4K-Te [17]	$\mathcal{S}_\alpha\uparrow$	.777	.830	.788	.750	.780	.811	.811	.805	.808	.822	.817	.838	.829	.839	.829	.829	.833	.840	.842	.847	.845	.857
	@$\mathcal{R}$	21	9	19	22	20	16	15	18	17	13	14	7	12	6	11	10	8	5	4	2	3	1
	$E_\phi^{mx}\uparrow$	.864	.897	.865	.793	.848	.886	.883	.881	.883	.888	.872	.904	.890	.899	.898	.893	.893	.907	.907	.914	.913	.922
	@$\mathcal{R}$	20	9	19	22	21	14	15	17	16	13	18	6	12	7	8	11	10	5	4	2	3	1
	$E_\phi^{mn}\uparrow$	.841	.854	.804	.773	.824	.871	.845	.846	.871	.876	.859	.897	.879	.874	.888	.863	.867	.895	.898	.903	.902	.911
	@$\mathcal{R}$	19	16	21	22	20	11	18	17	12	9	15	5	8	10	7	14	13	6	4	2	3	1
	$E_\phi^{ad}\uparrow$	.826	.864	.842	.812	.853	.883	.855	.876	.882	.871	.868	.898	.880	.886	.892	.884	.889	.902	.903	.898	.899	.907
	@$\mathcal{R}$	21	17	20	22	19	11	18	14	12	15	16	6	13	9	7	10	8	3	2	5	4	1
	$F_\beta^w\uparrow$	.639	.698	.632	.603	.656	.729	.680	.705	.723	.724	.732	.762	.734	.747	.745	.731	.740	.766	.771	.770	.764	.784
	@$\mathcal{R}$	20	17	21	22	19	13	18	16	15	14	11	6	10	7	8	12	9	4	2	3	5	1
	$F_\beta^{mx}\uparrow$	.731	.793	.738	.669	.719	.782	.762	.768	.775	.786	.782	.810	.793	.807	.799	.797	.800	.815	.816	.823	.819	.833
	@$\mathcal{R}$	20	11	19	22	21	15	18	17	16	13	14	6	12	7	9	10	8	5	4	2	3	1
	$F_\beta^{mn}\uparrow$	.696	.757	.695	.655	.705	.775	.729	.753	.769	.762	.772	.795	.773	.787	.784	.777	.782	.804	.806	.805	.799	.814
	@$\mathcal{R}$	20	16	21	22	19	11	18	17	14	15	13	6	12	7	8	10	9	4	2	3	5	1
	$F_\beta^{ad}\uparrow$	.671	.744	.709	.672	.710	.776	.717	.763	.768	.753	.767	.788	.766	.779	.779	.771	.778	.802	.803	.792	.789	.803
	@$\mathcal{R}$	22	17	20	21	19	10	18	15	12	16	13	6	14	8	7	11	9	3	2	4	5	1
	$\mathcal{M}\downarrow$	.075	.059	.074	.088	.070	.055	.064	.060	.058	.059	.058	.049	.055	.052	.053	.055	.052	.048	.047	.048	.047	.042
	@$\mathcal{R}$	21	15	20	22	19	10	18	17	14	16	13	6	12	8	9	11	7	5	2	4	3	1
	Mean@$\mathcal{R}$	21	15	20	22	19	12	18	17	16	13	14	6	11	7	8	10	9	5	2	3	4	1
CAMO-Te [32]	$\mathcal{S}_\alpha\uparrow$	.732	.779	.726	.771	.711	.739	.750	.737	.745	.769	.749	.796	.781	.785	.782	.772	.775	.787	.800	.820	.826	.839
	@$\mathcal{R}$	20	10	21	13	22	18	15	19	17	14	16	5	9	7	8	12	11	6	4	3	2	1
	$E_\phi^{mx}\uparrow$	.820	.850	.801	.849	.780	.820	.830	.818	.829	.837	.808	.864	.848	.854	.855	.842	.842	.854	.873	.895	.907	.915
	@$\mathcal{R}$	17	9	21	10	22	18	15	19	16	14	20	5	11	8	6	13	12	7	4	3	2	1
	$E_\phi^{mn}\uparrow$	.800	.797	.723	.795	.741	.787	.780	.767	.804	.824	.796	.854	.836	.823	.842	.807	.812	.838	.859	.882	.893	.901
	@$\mathcal{R}$	14	15	22	17	21	18	19	20	13	9	16	5	8	10	6	12	11	7	4	3	2	1
	$E_\phi^{ad}\uparrow$	.811	.848	.810	.847	.802	.811	.830	.826	.825	.833	.806	.864	.845	.859	.852	.850	.847	.855	.865	.875	.892	.901
	@$\mathcal{R}$	18	10	20	12	22	19	15	16	17	14	21	5	13	6	8	9	11	7	4	3	2	1
	$F_\beta^w\uparrow$	.604	.643	.553	.642	.564	.640	.610	.613	.644	.663	.646	.719	.678	.686	.695	.664	.673	.696	.728	.743	.754	.769
	@$\mathcal{R}$	20	15	22	16	21	17	19	18	14	12	13	5	9	8	7	11	10	6	4	3	2	1
	$F_\beta^{mx}\uparrow$	.688	.738	.667	.740	.630	.708	.694	.683	.708	.728	.703	.771	.745	.754	.758	.739	.740	.753	.779	.801	.810	.822
	@$\mathcal{R}$	19	13	21	10	22	15	18	20	16	14	17	5	9	7	6	12	11	8	4	3	2	1
	$F_\beta^{mn}\uparrow$	.670	.705	.614	.705	.616	.700	.663	.667	.702	.710	.692	.762	.728	.738	.746	.721	.726	.744	.772	.782	.792	.806
	@$\mathcal{R}$	18	13	22	14	21	16	20	19	15	12	17	5	9	8	6	11	10	7	4	3	2	1
	$F_\beta^{ad}\uparrow$	.667	.733	.678	.730	.661	.716	.692	.704	.712	.715	.696	.764	.729	.749	.751	.733	.738	.756	.779	.779	.786	.804
	@$\mathcal{R}$	21	10	20	12	22	14	19	17	16	15	18	5	13	8	7	11	9	6	4	3	2	1
	$\mathcal{M}\downarrow$	.109	.090	.114	.092	.109	.094	.102	.096	.092	.094	.096	.080	.087	.086	.085	.089	.088	.080	.073	.070	.063	.057
	@$\mathcal{R}$	21	12	22	13	20	15	19	18	14	16	17	6	9	8	7	11	10	5	4	3	2	1
	Mean@$\mathcal{R}$	20	12	21	13	22	16	18	19	15	14	17	5	9	8	7	11	10	6	4	3	2	1
COD10K-Te [2]	$\mathcal{S}_\alpha\uparrow$	.736	.789	.748	.778	.739	.776	.767	.769	.776	.789	.802	.813	.793	.818	.800	.811	.814	.804	.809	.815	.810	.822
	@$\mathcal{R}$	22	13	20	15	21	17	19	18	16	14	10	5	12	2	11	6	4	9	8	3	7	1
	$E_\phi^{mx}\uparrow$	.855	.880	.842	.871	.819	.867	.861	.864	.874	.879	.870	.900	.878	.891	.890	.890	.890	.892	.891	.906	.905	.911
	@$\mathcal{R}$	20	11	21	15	22	17	19	18	14	12	16	4	13	7	10	9	8	5	6	2	3	1
	$E_\phi^{mn}\uparrow$	.810	.817	.766	.810	.795	.857	.808	.823	.864	.861	.855	.890	.861	.853	.877	.845	.852	.880	.884	.887	.888	.896
	@$\mathcal{R}$	19	17	22	18	21	11	20	16	8	10	12	2	9	13	7	15	14	6	5	4	3	1
	$E_\phi^{ad}\uparrow$	.753	.789	.768	.791	.818	.867	.787	.837	.867	.839	.869	.886	.848	.850	.868	.851	.865	.882	.882	.863	.868	.877
	@$\mathcal{R}$	22	19	21	18	17	8	20	16	9	15	5	1	14	13	7	12	10	3	2	11	6	4
	$F_\beta^w\uparrow$	.517	.575	.509	.569	.544	.633	.557	.601	.631	.629	.677	.686	.635	.667	.660	.655	.666	.673	.684	.680	.672	.693
	@$\mathcal{R}$	21	17	22	18	20	13	19	16	14	15	5	2	12	8	10	11	9	6	3	4	7	1
	$F_\beta^{mx}\uparrow$	.633	.699	.634	.679	.609	.691	.658	.672	.691	.704	.729	.743	.712	.742	.725	.733	.738	.732	.738	.752	.743	.759
	@$\mathcal{R}$	21	14	20	17	22	16	19	18	15	13	10	3	12	5	11	8	7	9	6	2	4	1
	$F_\beta^{mn}\uparrow$	.582	.651	.582	.635	.593	.681	.615	.654	.679	.671	.715	.723	.679	.712	.701	.702	.711	.715	.721	.718	.710	.728
	@$\mathcal{R}$	22	17	21	18	20	12	19	16	13	15	6	2	14	7	11	10	8	5	3	4	9	1
	$F_\beta^{ad}\uparrow$	.526	.593	.555	.589	.588	.673	.573	.639	.667	.640	.707	.703	.652	.671	.676	.667	.681	.699	.705	.682	.680	.698
	@$\mathcal{R}$	22	17	21	18	19	10	20	16	12	15	1	3	14	11	9	13	7	4	2	6	8	5
	$\mathcal{M}\downarrow$	.061	.047	.058	.047	.051	.042	.051	.043	.043	.045	.038	.036	.042	.035	.040	.037	.035	.037	.035	.037	.036	.033
	@$\mathcal{R}$	22	18	21	17	20	13	19	15	14	16	10	5	12	4	11	9	3	8	2	7	6	1
	Mean@$\mathcal{R}$	22	16	21	18	20	14	19	17	13	15	9	2	12	8	10	11	7	6	3	4	5	1

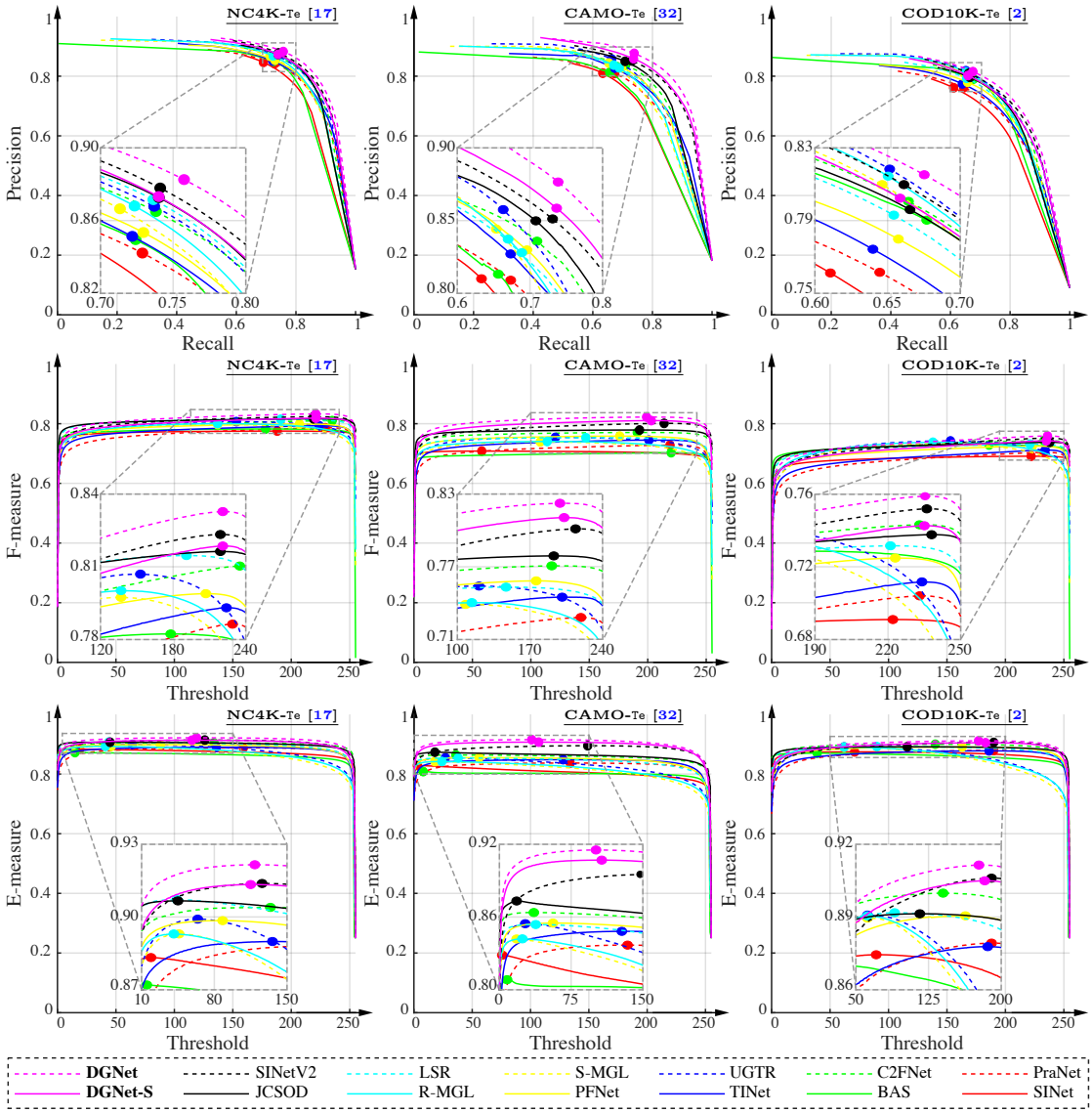


图5 与伪装目标检测相关的对比模型在三个常用数据集上的PR曲线（第1行）、F指标曲线（第2行)和增强匹配指标曲线（第3行）。PR曲线越靠近右上角，性能越好。F指标和E指标曲线越高，性能越好。彩版的阅读效果最佳。

表4 12个伪装目标检测模型与本文的两个模型实例（即：DGNet-S 和 DGNet）的推理速度对比。

Model	**DGNet-S**	**DGNet**	SINetV2 [2]	JCSOD [21]	LSR [17]	R-MGL [18]	S-MGL [18]
Input Size	352×352	352×352	352×352	352×352	352×352	473×473	473×473
Speed (fps)	**80**	58	68	43	31	9	13
Model	PFNet [15]	UGTR [20]	TINet [34]	C2FNet [27]	BAS [66]	PraNet [3]	SINet [1]
Input Size	416×416	473×473	352×352	352×352	288×288	352×352	352×352
Speed (fps)	78	15	50	68	31	73	63

好。此外，本文的方法也比TINet模型更简单且高效，例如：DGNet-S (8.0M) *vs.* TINet (28.6M)、DGNet-S (80 fps) *vs.* TINet (50 fps)。本文通过这样一个紧凑的设计在CAMO-Te上实现了最前沿的性能，例如：DGNet-S ($\mathcal{S}_\alpha = 0.826$)、DGNet模型 ($\mathcal{S}_\alpha = 0.839$) *vs.* TINet ($\mathcal{S}_\alpha = 0.781$)。

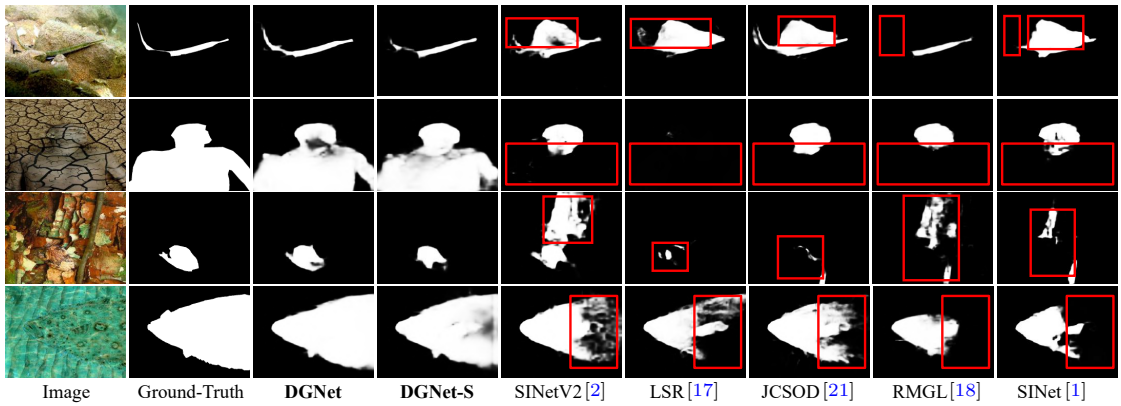


图6 常见的伪装目标检测基线模型和DGNet模型的可视化结果。红框表示假阳/阴性预测。更多可视化结果请参见https://github.com/GewelsJI/DGNet。

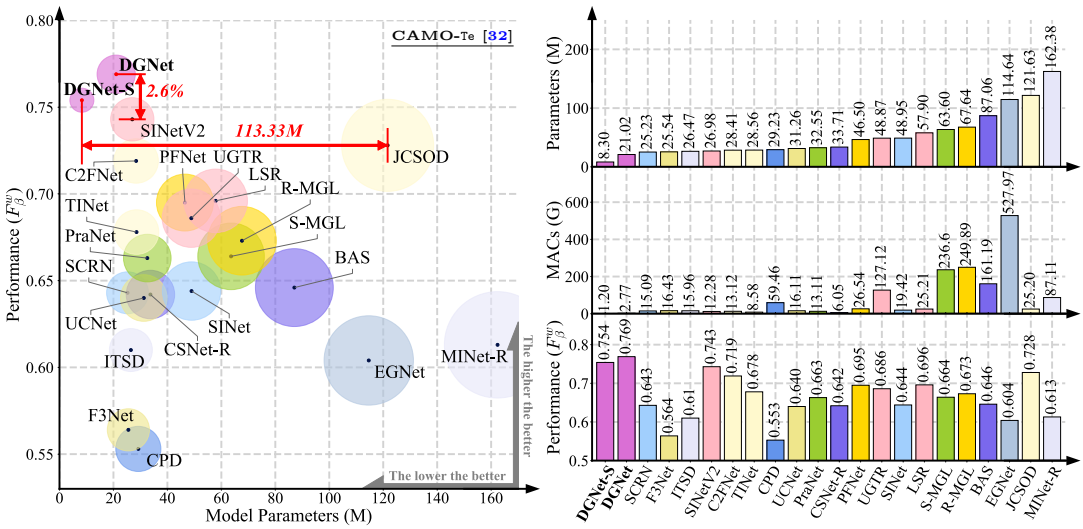


图7 （**左图**）本文呈现了在CAMO-Te [32]数据集上所有对比模型的性能（F_β^w）与模型参数量散点图。散点越大代表模型参数量越大。（**右图**）本文也提供了模型参数量、乘积累加运算和性能（F_β^w）的直方图对比。彩版的阅读效果最佳。

分组数目的配置： 在表5-(e)中展现不同分组数目M的变体模型，数目分别为1（#07）、4（#08）、8（#**S**）、16（#09）和32（#10）。注意到采用#07（$M=1$）未分组的候选特征时导致了性能的下降（在NC4K-Test数据集的E_ϕ^{mx}指标上降低了2.4%）。本文经验性地选择最佳性能M=8能作为默认配置。

放缩因子的配置： 本节在表5-(f)中也探讨缩放因子$N \in \{N_1, N_2, N_3\}$对模型性能的影响。相较于不同的配置（#11: $\{2,4,8\}$和#12: $\{4,8,16\}$），使用更细粒尺度的缩放因子(#**S**: $\{8,16,32\}$)可获得更好的预测性能。如图4所示，本文展示了三个并行的特征流（即：$\mathbf{A}_i^{N_1}$、$\mathbf{A}_i^{N_2}$和$\mathbf{A}_i^{N_3}$）的可视化结果，可见网络对目标内各个部分有着不同的关注度。这也验证了并行残差学习能以分组视角进一步改善输入的语义特征。

我们需要更多的子分支来进行软分组吗? 如表5-(g)所示，本节为不同的子分支设置了三个消融实验：3个子分支（#**S**: $N \in \{8,16,32\}$）、4个子分

表5 消融实验。#Para 和#MACs 代表模型的参数量和乘积累加运算次数。

No.	变体模型	Efficiency #Para	Efficiency #MACs	NC4K-Te [17] $\mathcal{S}_\alpha\uparrow$	NC4K-Te [17] $F_\beta^w\uparrow$	NC4K-Te [17] $\mathcal{M}\downarrow$	CAMO-Te [32] $\mathcal{S}_\alpha\uparrow$	CAMO-Te [32] $F_\beta^w\uparrow$	CAMO-Te [32] $\mathcal{M}\downarrow$	COD10K-Te [2] $\mathcal{S}_\alpha\uparrow$	COD10K-Te [2] $F_\beta^w\uparrow$	COD10K-Te [2] $\mathcal{M}\downarrow$
#S	**DGNet-S**	8.30M	1.20G	.845	.764	.047	.826	.754	.063	.810	.672	.036
(a) Base Network ↣ see 第3.1节												
#01	Base	8.24M	0.58G	.834	.676	.061	.814	.670	.072	.793	.550	.049
(b) Configuration of Dimensional Reduction ↣ see 第3.1节												
#02	$C_i = 16$	8.00M	0.81G	.842	.758	.048	.824	.749	.066	.806	.663	.037
#03	$C_i = 64$	9.36M	2.69G	.845	.764	.047	.827	.748	.065	.812	.673	.036
#04	$C_i = 128$	13.30M	8.55G	.847	.768	.046	.828	.751	.062	.810	.672	.036
(c) Network Decoupling Strategy ↣ see 第3.2节												
#05	*w/* $\mathbf{X}_2$	8.24M	0.59G	.840	.712	.055	.822	.701	.074	.805	.597	.043
(d) Should we use $\mathbf{Z}^G$ as supervision? ↣ see 公式(1)												
#06	*w/* $\mathbf{Z}^B$	8.30M	1.20G	.841	.753	.049	.821	.737	.067	.804	.654	.038
(e) Group Number M ↣ see 公式(2)												
#07	$M = 1$	8.30M	1.20G	.841	.756	.049	.822	.751	.064	.806	.662	.037
#08	$M = 4$	8.30M	1.20G	.842	.759	.048	.822	.742	.067	.809	.669	.036
#09	$M = 16$	8.30M	1.20G	.842	.752	.049	.829	.744	.065	.803	.651	.039
#10	$M = 32$	8.30M	1.20G	.845	.913	.047	.827	.745	.063	.809	.666	.036
(f) Scaling Factors $N \in \{N_1, N_2, N_3\}$ ↣ see 公式(5)												
#11	$\{2, 4, 8\}$	8.31M	1.20G	.842	.755	.048	.821	.741	.065	.808	.663	.036
#12	$\{4, 8, 16\}$	8.30M	1.20G	.844	.762	.047	.823	.744	.065	.806	.666	.037
(g) More sub-branches in Soft Grouping Strategy ↣ see 公式(5)												
#13	$N \in \{4, 8, 16, 32\}$	8.30M	1.20G	.844	.760	.048	.829	.748	.064	.811	.669	.037
#14	$N \in \{2, 4, 8, 16, 32\}$	8.31M	1.20G	.846	.765	.047	.825	.750	.063	.810	.670	.037
(h) Gradient-Induced Transition ↣ see 公式(6)												
#15	*w/o* $\mathcal{T}_i$	8.31M	1.20G	.839	.748	.050	.825	.741	.065	.802	.649	.039

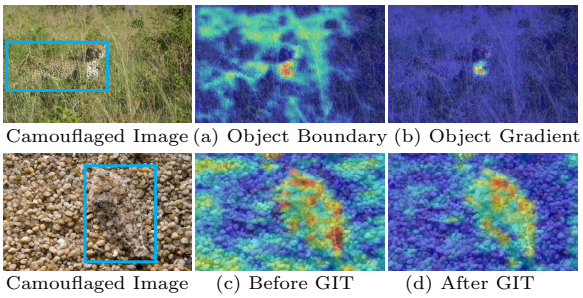


图8 两个核心设计的特征可视化，包含目标梯度监督（第1行）和GIT（第2行）。

表6 在不同监督方式下所训练的DGNet-S模型，包括：纹理标（TINet-Text [34]）和本文的目标梯度标签（DGNet-Grad）。

	NC4K-Te $\mathcal{S}_\alpha$	NC4K-Te F_β^w	NC4K-Te $\mathcal{M}$	CAMO-Te $\mathcal{S}_\alpha$	CAMO-Te F_β^w	CAMO-Te $\mathcal{M}$	COD10K-Te $\mathcal{S}_\alpha$	COD10K-Te F_β^w	COD10K-Te $\mathcal{M}$
w/ TINet-Text	.839	.747	.050	.820	.731	.068	.803	.652	.040
w/ DGNet-Grad	**.845**	**.764**	**.047**	**.826**	**.754**	**.063**	**.810**	**.672**	**.036**

支（#13: $N \in \{4, 8, 16, 32\}$）和5个子分支（#14: $N \in \{2, 4, 8, 16, 32\}$）。结果表明，分支越多，在数据集上的整体性能就越不稳定。

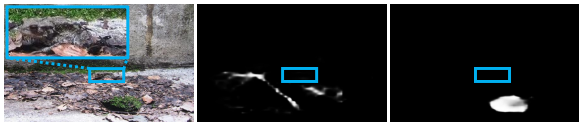
梯度诱导转换的贡献： 本节将模型中的GIT模块替换成朴素的通道拼接操作（表 5-(h)中的#15: *w/o* $\mathcal{T}_i$），用以验证其有效性。可见在CAMO-Te数据集上，配备GIT模块的模型（#S: *w/* $\mathcal{T}_i$）能在F_β^w指标上获得2.3%的提升。进一步地，如图 8中第二行所示，模型能获得更干净、更精细的表征$\mathbf{Z}_i^T$（即：After GIT）并抑制表征$\mathbf{X}_i^R$（即：Before GIT）中的背景噪声。这得益于GIT模块能够充分地聚合语义信息和纹理信息。

表7 本文模型在不同骨架网络下的性能，包括：EfficientNet [38] (即：EffNet-B1 & EffNet-B4) *vs.* MobileNet [67] (即：MobNet-S & MobNet-L)。

	#Para	MACs	NC4K-Te $\mathcal{S}_\alpha$	NC4K-Te F_β^w	CAMO-Te $\mathcal{S}_\alpha$	CAMO-Te F_β^w	COD10K-Te $\mathcal{S}_\alpha$	COD10K-Te F_β^w
MobNet-S	2.96M	1.27G	.779	.638	.735	.587	.729	.517
EffNet-B1	8.30M	1.20G	.845	.764	.826	.754	.810	.672
MobNet-L	6.96M	3.17G	.820	.723	.791	.686	.780	.620
EffNet-B4	21.02M	2.77G	.857	.784	.839	.769	.822	.693

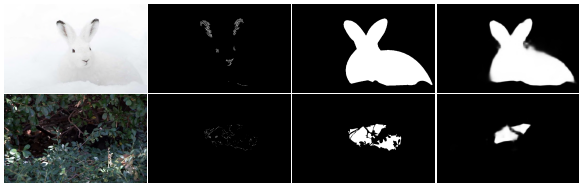
4.4 局限性

高效骨架网络vs. 轻量级骨架网络： 本文将高效的骨架网络（EfficientNet [38]）替换成轻量级的骨架网络（MobileNet [67]）来验证本文的方法在硬件受限条件下的潜在价值。在表 7中的结果表明本文的方法基于轻量级骨架网络取得了不太满意的性能，即MobNet-S（2.96M）和MobNet-L（6.96M），这也为未来探索留下了巨大空间。



Camouflaged Image SINetV2 [2] **DGNet (Ours)**

图9 具有小伪装对象的困难样本。



Camouflaged Image Object Gradient Ground-Truth Prediction

图10 不具有明显的目标梯度线索的视觉对比。

表8 DGNet模型和近期发布的ZoomNet模型[33]在三个测试集合上的性能对比。

	#Para	MACs	NC4K-Te		CAMO-Te		COD10K-Te	
			$\mathcal{S}_\alpha$	E_ϕ^{mx}	$\mathcal{S}_\alpha$	E_ϕ^{mx}	$\mathcal{S}_\alpha$	E_ϕ^{mx}
ZoomNet	32.38M	34.96G	.853	.912	.820	.892	**.838**	**.911**
DGNet	**21.02M**	**2.77G**	**.857**	**.922**	**.839**	**.915**	.822	**.911**

具有挑战性的情况： 尽管本文方法取得了令人满意的性能，它在如下具有挑战性的伪装场景中仍无法得到很好的预测：首先，本文认为所提出策略在如此有限的小目标区域中难以提供足够的纹理信息，这会导致假阳性的预测。如图 9所示，该案例也极易误导排名第一的SINetV2模型[2]，因此值得进一步的研究。

其次，我们观察到并不是所有的伪装对象内部都具有明显的梯度变化。如图 10的第一行所示，本文的方法可以很好地分割不具明显梯度变化的白兔。然而，本文的方法在如图 10第二行这种只包含相当稀少的梯度线索的极端条件下失败了。这激励我们在未来设计中去考虑整合更多启发式的和可学习的模式。此外，本文在提交之后注意到近期发布的伪装目标检测方法ZoomNet [33]。如表 8所示，本文的DGNet模型以微弱优势超越ZoomNet（即：NC4K-Te: +1.3% E_ϕ^{mx} 和CAMO-Te: +2.3% E_ϕ^{mx}），但未能在COD10K-Te上超越它。ZoomNet比本文的DGNet模型（21.02M 参数量）占用更多的计算资源（32.38M 参数量）。这激励我们在未来的扩展中考虑将缩放策略纳入本文的网络。

表9 两个常用的息肉分割测试数据集的定量结果。

	CVC-ColonDB [68]				ETIS-LPDB [69]			
Baseline	$\mathcal{S}_\alpha\uparrow$	$E_\phi^{mx}\uparrow$	$F_\beta^w\uparrow$	$D^{mx}\uparrow$	$\mathcal{S}_\alpha\uparrow$	$E_\phi^{mx}\uparrow$	$F_\beta^w\uparrow$	$D^{mx}\uparrow$
UNet [70]	.710	.781	.491	.560	.684	.740	.366	.444
UNet++ [71]	.692	.764	.467	.550	.683	.776	.390	.509
PraNet [3]	.820	.872	.699	.728	.794	.841	.600	.639
MSNet [72]	.838	.883	.736	.766	.845	.890	.677	.736
†**DGNet**	**.858**	**.898**	**.765**	**.789**	**.847**	**.904**	**.690**	**.741**

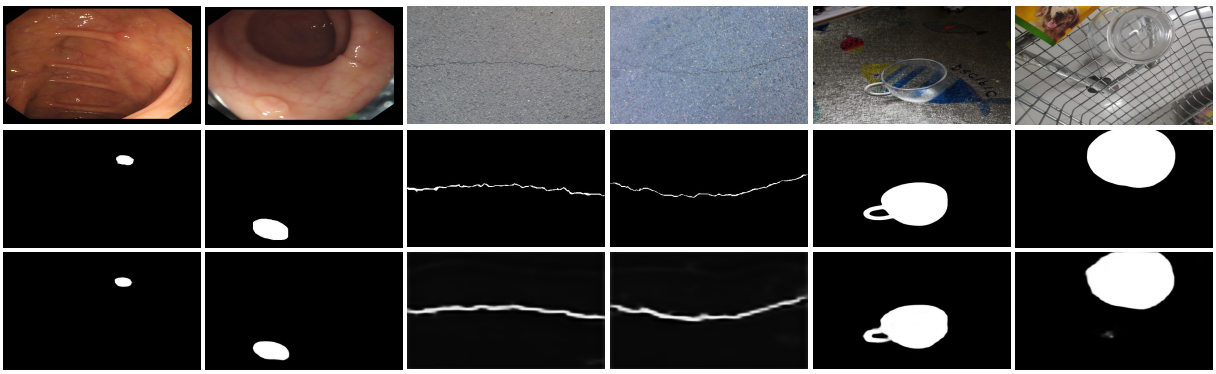
5 下游应用

本节在三个下游应用中评估了模型的泛化能力。

息肉分割： 在早期结肠镜筛查中，息肉与相似的周边环境具有较低的边缘对比度，它降低了直肠癌的检出能力。为展示本文方法在医疗领域的泛化能力，本节遵循同样的评测基准协议 [3]，在Kvasir-SEG [73]和CVC-ClinicDB [74]数据集上重新训练了DGNet 模型。采用两个未见过的测试数据集：CVC-ColonDB [68]和ETIS-LPDB [69]。表 9展示了†DGNet 在四个指标上一致地超越了四个前沿的息肉分割方法，包含$\mathcal{S}_\alpha$、E_ϕ^{mx}、F_β^w和Dice指标的最大值（D^{mx}）。值得注意的是，†DGNet 表示在特定任务的训练数据集上重新训练了DGNet。由†DGNet生成的可视化结果请参见图 11 (a)。

缺陷检测： 不合格的工业产品（例如：瓷砖、木板）会不可避免地造成无法挽回的经济损失。本节进一步在道路裂缝检测数据集上（即：CrackForest [75]）重新训练了四个经典的伪装目标检测模型和†DGNet模型，其中60%的数据用来训练，而40%的数据用来测试。如图 11所示，展示了一些可视化结果。

透明目标分割： 日常生活中，智能体（例如：机器人和无人机）需要学习去识别不明显的透明物体（例如：玻璃、水瓶和镜子）来避免事故。本



（a）息肉分割　　（b）道路裂缝检测　　（c）透明目标分割

图11 三种下游应用的可视化结果。从上到下：输入图像（第1行）、真值图（第2行）、预测图（第3行）。

节也在透明目标分割任务中验证了†DGNet 模型的有效性。为了简便，本节将Trans10K [76]数据集中的实例级别的标注重新组织成为目标级别的标注来训练模型。图 11中的可视化结果进一步展示了†DGNet模型的学习能力。

6 结语

本文提出了首个基于深度梯度学习的框架（*DGNet*）用以高效地分割伪装目标。为了挖掘伪装特征，本文提出将任务解耦成两个分支，即语义编码器和纹理编码器。本文设计了新颖的即插即用的梯度诱导转换（GIT），作为软分组模块来联合地学习来自两个分支的特征。这一简易且灵活的框架在三个具有挑战性的数据集上，相较于20个前沿对比模型展现了它强大的泛化能力。除此之外，本文的高效模型实例*DGNet-S*（8.3M & 80 fps）达到优异的性能与效率权衡。本文的方法在息肉分割、缺陷检测和透明目标分割的三个下游任务中，也得到了令人满意的结果，验证了其实际应用价值。

References

[1] D.-P. Fan, G.-P. Ji, G. Sun, M.-M. Cheng, J. Shen, and L. Shao, "Camouflaged object detection," in *Conference on computer vision and pattern recognition*. Seattle, WA, USA: IEEE, 2020, pp. 2777–2787, DOI: 10.1109/CVPR42600.2020.00285.

[2] D.-P. Fan, G.-P. Ji, M.-M. Cheng, and L. Shao, "Concealed object detection," *Transactions on pattern analysis and machine intelligence*, pp. 1–1, 2021, DOI: 10.1109/TPAMI.2021.3085766.

[3] D.-P. Fan, G.-P. Ji, T. Zhou, G. Chen, H. Fu, J. Shen, and L. Shao, "Pranet: Parallel reverse attention network for polyp segmentation," in *International Conference on Medical Image Computing and Computer Assisted Intervention*. Lima, Peru: Springer, 2020, pp. 263–273, DOI: https://doi.org/10.1007/978-3-030-59725-2_26.

[4] G.-P. Ji, Y.-C. Chou, D.-P. Fan, G. Chen, H. Fu, D. Jha, and L. Shao, "Progressively normalized self-attention network for video polyp segmentation," in *International Conference on Medical Image Computing and Computer Assisted Intervention*. Strasbourg, France: Springer, 2021, pp. 142–152, DOI: 10.1007/978-3-030-87193-2_14.

[5] G.-P. Ji, G. Xiao, Y.-C. Chou, D.-P. Fan, K. Zhao, G. Chen, H. Fu, and L. Van Gool, "Video polyp segmentation: A deep learning perspective," [Online], 2022, Available: https://arxiv.org/abs/2203.14291.

[6] D.-P. Fan, T. Zhou, G.-P. Ji, Y. Zhou, G. Chen, H. Fu, J. Shen, and L. Shao, "Inf-net: Automatic covid-19 lung infection segmentation from ct images," *Transactions on Medical Imaging*, vol. 39, no. 8, pp. 2626–2637, 2020, DOI: 10.1109/TMI.2020.2996645.

[7] Y.-H. Wu, S.-H. Gao, J. Mei, J. Xu, D.-P. Fan, R.-G. Zhang, and M.-M. Cheng, "Jcs: An explainable covid-19 diagnosis system by joint classification and segmentation," *Transactions on image processing*, vol. 30, pp. 3113–3126, 2021, DOI: 10.1109/TIP.2021.3058783.

[8] J. Liu, B. Dong, S. Wang, H. Cui, D.-P. Fan, J. Ma, and G. Chen, "Covid-19 lung infection segmentation with a novel two-stage cross-domain transfer learning framework," *Medical image analysis*, vol. 74, p. 102205, 2021, DOI: 10.1016/j.media.2021.102205.

[9] G.-P. Ji, K. Fu, Z. Wu, D.-P. Fan, J. Shen, and L. Shao, "Full-duplex strategy for video object segmentation," in *International conference on computer vision*. Montreal, Canada: IEEE, 2021, pp. 4922–4933, DOI: 10.1109/ICCV48922.2021.00488.

[10] W.-C. Chen, X.-Y. Yu, and L.-L. Ou, "Pedestrian attribute recognition in video surveillance scenarios based on view-attribute attention localization," *Machine Intelligence Research*, vol. 19, no. 2, pp. 153–168, 2022, DOI: 10.1007/s11633-022-1321-8.

[11] J.-R. Xue, J.-W. Fang, and P. Zhang, "A survey of scene understanding by event reasoning in autonomous driving," *International Journal of Automation and Computing*, vol. 15, no. 3, pp. 249–266, 2018, DOI: 10.1007/s11633-018-1126-y.

[12] R. Feng and B. Prabhakaran, "Facilitating fashion camouflage art," in *International conference on Multimedia*. Barcelona, Spain: ACM, 2013, pp. 793–802, DOI: 10.1145/2502081.2502121.

[13] M. Dean, R. Harwood, and C. Kasari, "The art of camouflage: Gender differences in the social behaviors of girls and boys with autism spectrum disorder," *Autism*, vol. 21, no. 6, pp. 678–689, 2017, DOI: 10.1177/1362361316671845.

[14] G.-P. Ji, D.-P. Fan, Y.-C. Chou, D. Dai, A. Liniger, and L. Van Gool, "Deep gradient learning for efficient camouflaged object detection," *Machine Intelligence Research*, 2022.

[15] H. Mei, G.-P. Ji, Z. Wei, X. Yang, X. Wei, and D.-P. Fan, "Camouflaged object segmentation with distraction mining," in *Conference on computer vision and pattern recognition*. Nashville, TN, USA: IEEE, 2021, pp. 8772–8781, DOI: 10.1109/CVPR46437.2021.00866.

[16] Y. Mao, J. Zhang, Z. Wan, Y. Dai, A. Li, Y. Lv, X. Tian, D.-P. Fan, and N. Barnes,

"Transformer transforms salient object detection and camouflaged object detection," [Online], 2021, Available: https://arxiv.org/abs/2104.10127.

[17] Y. Lyu, J. Zhang, Y. Dai, A. Li, B. Liu, N. Barnes, and D.-P. Fan, "Simultaneously localize, segment and rank the camouflaged objects," in *Conference on computer vision and pattern recognition*. Nashville, TN, USA: Springer, 2021, pp. 11 591–11 601, DOI: 10.1109/CVPR46437.2021.01142.

[18] Q. Zhai, X. Li, F. Yang, C. Chen, H. Cheng, and D.-P. Fan, "Mutual graph learning for camouflaged object detection," in *Conference on computer vision and pattern recognition*. Nashville, TN, USA: IEEE, 2021, pp. 12 997–13 007, DOI: 10.1109/CVPR46437.2021.01280.

[19] G.-P. Ji, L. Zhu, M. Zhuge, and K. Fu, "Fast camouflaged object detection via edge-based reversible re-calibration network," *Pattern Recognition*, vol. 123, p. 108414, 2022, DOI: 10.1016/j.patcog.2021.108414.

[20] F. Yang, Q. Zhai, X. Li, R. Huang, A. Luo, H. Cheng, and D.-P. Fan, "Uncertainty-guided transformer reasoning for camouflaged object detection," in *ICCV*. Santiago, Chile: IEEE, 2021, pp. 4146–4155, DOI: 10.1109/ICCV48922.2021.00411.

[21] A. Li, J. Zhang, Y. Lv, B. Liu, T. Zhang, and Y. Dai, "Uncertainty-aware joint salient object and camouflaged object detection," in *Conference on computer vision and pattern recognition*. Nashville, TN, USA: IEEE, 2021, pp. 10 071–10 081, DOI: 10.1109/CVPR46437.2021.00994.

[22] Y. Pan, Y. Chen, Q. Fu, P. Zhang, and X. Xu, "Study on the camouflaged target detection method based on 3d convexity," *Modern Applied Science*, vol. 5, no. 4, p. 152, 2011, DOI: 10.5539/mas.v5n4p152.

[23] P. Sengottuvelan, A. Wahi, and A. Shanmugam, "Performance of decamouflaging through exploratory image analysis," in *2008 First International Conference on Emerging Trends in Engineering and Technology*. Nagpur, India: IEEE, 2008, pp. 6–10, DOI: 10.1109/ICETET.2008.232.

[24] Z. Liu, K. Huang, and T. Tan, "Foreground object detection using top-down information based on em framework," *Transactions on image processing*, vol. 21, no. 9, pp. 4204–4217, 2012, DOI: 10.1109/TIP.2012.2200492.

[25] J. Y. Y. H. W. Hou and J. Li, "Detection of the mobile object with camouflage color under dynamic background based on optical flow," *Procedia Engineering*, vol. 15, pp. 2201–2205, 2011, DOI: 10.1016/j.proeng.2011.08.412.

[26] J. Gallego and P. Bertolino, "Foreground object segmentation for moving camera sequences based on foreground-background probabilistic models and prior probability maps," in *International Conference on Image Processing*. Paris, France: IEEE, 2014, pp.

3312–3316, DOI: 10.1109/ICIP.2014.7025670.

[27] Y. Sun, G. Chen, T. Zhou, Y. Zhang, and N. Liu, "Context-aware Cross-level Fusion Network for Camouflaged Object Detection," in *International Joint Conference on Artificial Intelligence*. Montreal-themed virtual reality: IJCAI, 2021, DOI: 10.24963/ijcai.2021/142.

[28] G. Chen, S. Liu, Y. Sun, G.-P. Ji, Y. Wu, and T. Zhou, "Camouflaged object detection via context-aware cross-level fusion," *Transactions on circuits and systems for video technology*, pp. 1–1, 2022, DOI: 10.1109/TCSVT.2022.3178173.

[29] J. Ren, X. Hu, L. Zhu, X. Xu, Y. Xu, W. Wang, Z. Deng, and P.-A. Heng, "Deep texture-aware features for camouflaged object detection," *Transactions on circuits and systems for video technology*, pp. 1–1, 2021, DOI: 10.1109/TCSVT.2021.3126591.

[30] N. Kajiura, H. Liu, and S. Satoh, "Improving camouflaged object detection with the uncertainty of pseudo-edge labels," in *Multimedia Asia*, Gold Coast, Australia, 2021, pp. 1–7, DOI: 10.1145/3469877.3490587.

[31] M. Zhuge, X. Lu, Y. Guo, Z. Cai, and S. Chen, "Cubenet: X-shape connection for camouflaged object detection," *Pattern Recognition*, vol. 127, p. 108644, 2022, DOI: 10.1016/j.patcog.2022.108644.

[32] T.-N. Le, T. V. Nguyen, Z. Nie, M.-T. Tran, and A. Sugimoto, "Anabranch network for camouflaged object segmentation," *Computer Vision and Image Understanding*, vol. 184, pp. 45–56, 2019, DOI: https://doi.org/10.1016/j.cviu.2019.04.006.

[33] Y. Pang, X. Zhao, T.-Z. Xiang, L. Zhang, and H. Lu, "Zoom in and out: A mixed-scale triplet network for camouflaged object detection," in *Conference on computer vision and pattern recognition*. New Orleans, LA, USA: IEEE, 2022, pp. 2160–2170.

[34] J. Zhu, X. Zhang, S. Zhang, and J. Liu, "Inferring camouflaged objects by texture-aware interactive guidance network," in *AAAI Conference on Artificial Intelligence*, vol. 35, no. 4. [Online]: AAAI Press, 2021, pp. 3599–3607.

[35] X. Cheng, H. Xiong, D.-P. Fan, Y. Zhong, M. Harandi, T. Drummond, and Z. Ge, "Implicit motion handling for video camouflaged object detection," in *Conference on computer vision and pattern recognition*. New Orleans, LA, USA: IEEE, 2022, pp. 13 864–13 873.

[36] T. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Conference on computer vision and pattern recognition*. Honolulu, HI, USA: IEEE, 2017, pp. 2117–2125, DOI: 10.1109/CVPR.2017.106.

[37] Z. Ke, J. Sun, K. Li, Q. Yan, and R. W. Lau, "Modnet: real-time trimap-free portrait matting via objective decomposition," in *AAAI Conference on Artificial Intelligence*, vol. 36,

no. 1. [Online]: AAAI Press, 2022, pp. 1140–1147, DOI: 10.1609/aaai.v36i1.19999.

[38] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International conference on machine learning*. New Orleans, LA, USA: PMLR, 2019, pp. 6105–6114.

[39] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International conference on machine learning*. Lille, France: PMLR, 2015, pp. 448–456.

[40] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proceedings of the fourteenth international conference on artificial intelligence and statistics*. Fort Lauderdale, FL, USA: JMLR Workshop and Conference Proceedings, 2011, pp. 315–323.

[41] Z. Su, W. Liu, Z. Yu, D. Hu, Q. Liao, Q. Tian, M. Pietikainen, and L. Liu, "Pixel difference networks for efficient edge detection," in *International conference on computer vision*. Montreal, Canada: IEEE, 2021, pp. 5117–5127, DOI: 10.1109/ICCV48922.2021.00507.

[42] C. Ma, Y. Rao, Y. Cheng, C. Chen, J. Lu, and J. Zhou, "Structure-preserving super resolution with gradient guidance," in *Conference on computer vision and pattern recognition*. Seattle, WA, USA: IEEE, 2020, pp. 7769–7778, DOI: 10.1109/CVPR42600.2020.00779.

[43] J. Canny, "A computational approach to edge detection," *Transactions on pattern analysis and machine intelligence*, no. 6, pp. 679–698, 1986, DOI: 10.1109/TPAMI.1986.4767851.

[44] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *Transactions on pattern analysis and machine intelligence*, vol. 40, no. 4, pp. 834–848, 2017, DOI: 10.1109/TPAMI.2017.2699184.

[45] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Conference on computer vision and pattern recognition*. Las Vegas, NV, USA: IEEE, 2016, pp. 770–778, DOI: 10.1109/CVPR.2016.90.

[46] J. Wei, S. Wang, and Q. Huang, "F^3net: fusion, feedback and focus for salient object detection," in *AAAI Conference on Artificial Intelligence*, vol. 34, no. 07. New York, New York, USA: AAAI Press, 2020, pp. 12 321–12 328, DOI: 10.1609/aaai.v34i07.6916.

[47] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga *et al.*, "Pytorch: An imperative style, high-performance deep learning library," in *Advances in neural information processing systems*, vol. 32. Vancouver, Canada: Curran Associates, Inc., 2019.

[48] S.-M. Hu, D. Liang, G.-Y. Yang, G.-W. Yang, and W.-Y. Zhou, "Jittor: a novel deep learning framework with meta-operators and unified graph execution," *Science China Information Sciences*, vol. 63, no. 12, pp. 1–21, 2020, DOI: 10.1007/s11432-020-3097-4.

[49] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *International conference on computer vision*. Santiago, Chile: IEEE, 2015, pp. 1026–1034, DOI: 10.1109/ICCV.2015.123.

[50] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, vol. 25. Stateline, NV, USA: Curran Associates, Inc., 2012.

[51] I. Loshchilov and F. Hutter, "Sgdr: Stochastic gradient descent with warm restarts," in *International Conference on Learning Representations*. Toulon, France: PMLR, 2017.

[52] P. Krähenbühl and V. Koltun, "Efficient inference in fully connected crfs with gaussian edge potentials," in *Advances in neural information processing systems*, vol. 24. Granada, Spain: Curran Associates, Inc., 2011, pp. 109–117.

[53] D.-P. Fan, M.-M. Cheng, Y. Liu, T. Li, and A. Borji, "Structure-measure: A new way to evaluate foreground maps," in *International conference on computer vision*. Venice, Italy: IEEE, 2017, pp. 4548–4557, DOI: 10.1109/ICCV.2017.487.

[54] D.-P. Fan, C. Gong, Y. Cao, B. Ren, M.-M. Cheng, and A. Borji, "Enhanced-alignment measure for binary foreground map evaluation," in *International Joint Conference on Artificial Intelligence*. Stockholm, Sweden: IJCAI, 2018, pp. 698–704, DOI: 10.24963/ijcai.2018/97.

[55] D.-P. Fan, G.-P. Ji, X. Qin, and M.-M. Cheng, "Cognitive vision inspired object segmentation metric and loss function," *SCIENTIA SINICA Informationis*, vol. 6, p. 6, 2021, DOI: 10.1155/2017/4037190.

[56] A. Borji, M.-M. Cheng, H. Jiang, and J. Li, "Salient object detection: A benchmark," *Transactions on image processing*, vol. 24, no. 12, pp. 5706–5722, 2015, DOI: 10.1109/TIP.2015.2487833.

[57] Q. Hou, M.-M. Cheng, X. Hu, A. Borji, Z. Tu, and P. H. Torr, "Deeply supervised salient object detection with short connections," in *Conference on computer vision and pattern recognition*. Honolulu, HI, USA: IEEE, 2017, pp. 3203–3212, DOI: 10.1109/TPAMI.2018.2815688.

[58] R. Margolin, L. Zelnik-Manor, and A. Tal, "How to evaluate foreground maps?" in *Conference on computer vision and pattern recognition*. Columbus, OH, USA: IEEE, 2014, pp. 248–255, DOI: 10.1109/CVPR.2014.39.

[59] J.-X. Zhao, J.-J. Liu, D.-P. Fan, Y. Cao, J. Yang, and M.-M. Cheng, "Egnet: Edge

guidance network for salient object detection," in *International conference on computer vision*. Seoul, Korea: IEEE, 2019, pp. 8779–8788, DOI: 10.1109/ICCV.2019.00887.

[60] Z. Wu, L. Su, and Q. Huang, "Stacked cross refinement network for edge-aware salient object detection," in *ICCV*. Seoul, Korea: IEEE, 2019, pp. 7264–7273, DOI: 10.1109/ICCV.2019.00736.

[61] Z. Wu, L. Su, and Q. Huang, "Cascaded partial decoder for fast and accurate salient object detection," in *Conference on computer vision and pattern recognition*. Long Beach, CA, USA: IEEE, 2019, pp. 3907–3916, DOI: 10.1109/CVPR.2019.00403.

[62] S.-H. Gao, Y.-Q. Tan, M.-M. Cheng, C. Lu, Y. Chen, and S. Yan, "Highly efficient salient object detection with 100k parameters," in *European conference on computer vision*. Glasgow, United Kingdom: Springer, 2020, pp. 702–721, DOI: 10.1007/978-3-030-58539-6_42.

[63] J. Zhang, D.-P. Fan, Y. Dai, S. Anwar, F. S. Saleh, T. Zhang, and N. Barnes, "Uc-net: Uncertainty inspired rgb-d saliency detection via conditional variational autoencoders," in *Conference on computer vision and pattern recognition*. Seattle, WA, USA: IEEE, 2020, pp. 8582–8591, DOI: 10.1109/CVPR42600.2020.00861.

[64] H. Zhou, X. Xie, J.-H. Lai, Z. Chen, and L. Yang, "Interactive two-stream decoder for accurate and fast saliency detection," in *Conference on computer vision and pattern recognition*. Seattle, WA, USA: IEEE, 2020, pp. 9141–9150, DOI: 10.1109/CVPR42600.2020.00916.

[65] Y. Pang, X. Zhao, L. Zhang, and H. Lu, "Multi-scale interactive network for salient object detection," in *Conference on computer vision and pattern recognition*. Seattle, WA, USA: IEEE, 2020, pp. 9413–9422, DOI: 10.1109/CVPR42600.2020.00943.

[66] X. Qin, D.-P. Fan, C. Huang, C. Diagne, Z. Zhang, A. C. Sant'Anna, A. Suarez, M. Jagersand, and L. Shao, "Boundary-aware segmentation network for mobile and web applications," [Online], 2021, Available: https://arxiv.org/abs/2101.04704.

[67] A. Howard, M. Sandler, G. Chu, L.-C. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, V. Vasudevan, Q. V. Le, and H. Adam, "Searching for mobilenetv3," in *International conference on computer vision*. Seoul, Korea: IEEE, 2019, pp. 1314–1324, DOI: 10.1109/ICCV.2019.00140.

[68] J. Bernal, J. Sánchez, and F. Vilarino, "Towards automatic polyp detection with a polyp appearance model," *Pattern Recognition*, vol. 45, no. 9, pp. 3166–3182, 2012, DOI: 10.1016/j.patcog.2012.03.002.

[69] J. Silva, A. Histace, O. Romain, X. Dray, and B. Granado, "Toward embedded detection of polyps in wce images for early diagnosis of colorectal cancer," *International journal*

of computer assisted radiology and surgery, vol. 9, no. 2, pp. 283–293, 2014, DOI: 10.1007/s11548-013-0926-3.

[70] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer Assisted Intervention*. Munich, Germany: Springer, 2015, pp. 234–241, DOI: 10.1007/978-3-319-24574-4_28.

[71] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: Redesigning skip connections to exploit multiscale features in image segmentation," *Transactions on Medical Imaging*, vol. 39, no. 6, pp. 1856–1867, 2019, DOI: 10.1109/TMI.2019.2959609.

[72] X. Zhao, L. Zhang, and H. Lu, "Automatic polyp segmentation via multi-scale subtraction network," in *International Conference on Medical Image Computing and Computer Assisted Intervention*. Strasbourg, France: Springer, 2021, pp. 120–130, DOI: 10.1007/978-3-030-87193-2_12.

[73] D. Jha, P. H. Smedsrud, M. A. Riegler, P. Halvorsen, T. de Lange, D. Johansen, and H. D. Johansen, "Kvasir-seg: A segmented polyp dataset," in *International Conference on Multimedia Modeling*. Daejeon, Korea: Springer, 2020, pp. 451–462, DOI: 10.1007/978-3-030-37734-2_37.

[74] J. Bernal, F. J. Sánchez, G. Fernández-Esparrach, D. Gil, C. Rodríguez, and F. Vilariño, "Wm-dova maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians," *Computerized Medical Imaging and Graphics*, vol. 43, pp. 99–111, 2015, DOI: 10.1016/j.compmedimag.2015.02.007.

[75] Y. Shi, L. Cui, Z. Qi, F. Meng, and Z. Chen, "Automatic road crack detection using random structured forests," *Transactions on Intelligent Transportation Systems*, vol. 17, no. 12, pp. 3434–3445, 2016, DOI: 10.1109/TITS.2016.2552248.

[76] E. Xie, W. Wang, W. Wang, M. Ding, C. Shen, and P. Luo, "Segmenting transparent objects in the wild," in *European conference on computer vision*. Glasgow, United Kingdom: Springer, 2020, pp. 696–711, DOI: 10.1007/978-3-030-58601-0_41.