

文章编号:1671-6833(2025)05-0009-09

用于伪装目标检测的边缘-纹理引导增强网络

魏明军^{1,2}, 陈晓茹¹, 刘铭¹, 刘亚志^{1,2}, 李辉¹

(1. 华北理工大学人工智能学院, 河北唐山 063210; 2. 河北省工业智能感知重点实验室, 河北唐山 063210)

摘要: 伪装目标检测(COD)因目标物体与背景极为相似而具有很大的挑战性。针对目前COD方法未能充分利用边缘和纹理信息辅助检测任务而造成的边缘预测模糊、检测结果不完整且存在干扰等问题,提出了一种边缘-纹理引导增强网络(ETGNet),通过显式且充分的边缘和纹理引导策略来进一步提升COD的性能。首先,ETGNet中包含了一个关键的特征引导增强模块(FGEM),该模块能够利用并行的特征细化分支处理并增强对象特征,引导分支通过引导注意力来获取对象特征与边缘-纹理线索之间的相关性,以加强网络对于对象细节信息的理解并抑制噪声干扰;而自增强分支则利用自注意力机制从全局角度对伪装对象特征进行细化。其次,提出了一个特征交互融合模块(FIFM)来渐进融合相邻特征,FIFM利用注意力交互机制和加权融合策略学习特征间的互补信息,以生成更完整的预测图。最后,在3个公共数据集CAMO、COD10K和NC4K上进行实验验证,结果表明:所提出的网络在结构度量 S 、自适应增强匹配度量 E 、加权 F 度量和平均绝对误差 M 指标上均优于相关领域的其他方法,尤其在最大的测试集NC4K上,加权 F 度量指标高于所对比12个COD方法中表现最佳的FSPNet 2.2百分点。

关键词: 伪装目标检测; 边缘信息; 纹理信息; 特征引导; 特征交互

中图分类号: TP391 **文献标志码:** A **doi:** 10.13705/j.issn.1671-6833.2025.02.009

伪装目标检测(camouflaged object detection, COD)任务旨在检测出图像中与背景高度相似且边缘辨识度低的对象。高效的伪装目标检测方法不仅可以应用于医学、农业及工业等各个领域,同时也能作为其他的检测任务提供解决方法。近年来,随着大规模数据集如CAMO^[1]、COD10K^[2]和NC4K^[3]等的提出,许多基于深度学习的COD方法不断涌现。例如,2020年,Fan等^[2]设计了一个伪装目标搜索识别网络SINet,将COD任务分为搜索和识别两个阶段来定位和细化伪装目标;2021年,Lv等^[3]提出了一个基于排序的LSR网络,该网络在统一的框架中构建三重任务来同时定位、分割和排序伪装对象;2022年,Pang等^[4]设计了一个混合尺度三元组网络ZoomNet,通过对伪装对象的尺度变化行为进行显式建模来实现检测任务;2023年,Song等^[5]设计了一种双骨干网络FDNet,使用基于特征嫁接和干扰感知的策略聚合双骨干中的信息并消除潜在的干扰;同年,Huang等^[6]提出了一个特征收缩金字塔网

络FSPNet,利用特征探索与特征收缩策略来挖掘和聚合伪装对象线索。

虽然上述方法能够定位和分割出图像中包含伪装对象的区域,但它们仍未能精准地分割伪装对象的边界,且分割结果中普遍存在检测不完整的现象。针对上述问题,研究者们提出了多种基于边缘信息引导或基于纹理信息引导的方法来增强分割特征,以获取更精细的预测结果。例如,2022年,Sun等^[7]提出了一种边缘引导网络BGNet,将探索到的边缘线索与对象特征进行聚合,引导网络学习对象的结构语义特征表示;同年,Zhu等^[8]设计了一种边缘引导分离注意力模型BSA-Net;2023年,Sun等^[9]提出了一种边缘感知镜像网络EAMNet,该网络采用包含边缘分支和分割分支的双分支架构来协同增强边缘特征和分割特征;同年, Ji等^[10]提出一个利用双编码器提取对象与纹理特征的网络DGNet,该网络融合这两种特征进行最终预测。然而,以上方法中仍存在两个主要问题:①没有考虑对边缘细节

收稿日期:2025-01-09;修订日期:2025-02-23

基金项目:国家自然科学基金资助项目(52074126);河北省高等学校科学技术研究项目(ZD2022102)

作者简介:魏明军(1969—),男,河北唐山人,华北理工大学教授,主要从事计算机视觉、入侵检测、机器学习、数据挖掘等方面的研究,E-mail:weimj@ncst.edu.cn。

引用本文:魏明军,陈晓茹,刘铭,等.用于伪装目标检测的边缘-纹理引导增强网络[J].郑州大学学报(工学版),2025,46(5):9-17.(WEI M J, CHEN X R, LIU M, et al. Edge-texture guided enhancement network for camouflaged object detection [J]. Journal of Zhengzhou University (Engineering Science), 2025, 46(5): 9-17.)

和纹理线索的共同关注。②在使用边缘或纹理信息来辅助检测任务时仅使用较直接的特征引导方式,如BGNet仅使用逐元素相乘的策略来结合边缘与对象特征;BSA-Net仅通过较简单的加权相乘并相加的方式将边缘信息嵌入到伪装对象的前景与背景特征中;DGNet则设计了软分组策略,通过分组学习的方式较直接地使用梯度纹理信息来增强对象特征。总之,上述方法没有考虑多源信息间的相关性,可能会引入额外的噪声干扰,且导致网络无法充分学习多源信息中的对象线索,从而影响检测结果。

为了有效地减少引入边缘和纹理信息所带来的干扰,并充分利用这些信息来辅助检测任务以获得更精细的检测结果,本文提出了一种高效的边缘-纹理引导增强网络(edge-texture guided enhancement network, ETGENet)。首先,利用边缘与纹理感知解码器(edge & texture perception decoder, ETPD)来获取伪装对象精细的边缘和纹理先验信息;其次,设计了一种特征引导增强模块(feature guided enhancement module, FGEM),该模块通过计算边缘和纹理特征与对象特征间的引导注意力权重,使网络更加关注伪装对象的边缘和纹理信息,从而增强对象特征的代表能力;最后,提出了一种特征交互融合模块(feature interaction fusion module, FIFM),利用交互融合策略,FIFM能够有效地增强和聚合相邻层级特征,以获得整体语义更加完善的预测结果。

1 边缘-纹理引导增强网络

1.1 整体结构

本文提出的ETGENet的整体结构如图1所示。具体来说,对于给定的一张RGB图像,首先,采用PVT v2(pyramid vision transformer v2)^[11]作为编码器来逐层级提取图像特征,记为 $\{f_i, i \in 1, 2, 3, 4\}$ 。随后,利用感受野模块(receptive field block, RFB)^[12]处理各层级信息,以获取更丰富的多尺度特征,即 $\{f_{R_i}, i \in 1, 2, 3, 4\}$ 。经过RFB后,使用边缘与纹理感知解码器(ETPD)从低级特征 f_{R_1} 和高级特征 $(f_{R_3} \ f_{R_4})$ 中提取出与伪装目标相关的边缘信息 x_e 和纹理信息 x_t 。同时,利用后3个层级的特征通过拼接和卷积操作生成具有更丰富语义的特征 F_5 。随后,通过多个特征引导增强模块FGEM,自上而下地将来自ETPD的 x_e 、 x_t 与多级特征 f_{R_2} 、 f_{R_3} 、 f_{R_4} 和 F_5 集成在一起,以引导网络关注边缘和纹理信息,从而增强伪装目标及其边界特征的代表能力。最后,本文采用多个特征交互融合模块FIFM,以渐进融合的方式聚合多级特征,逐步细化对象预测。在推理过程中,本文选择最后一个FIFM的预测作为最终结果。

1.2 基于Transformer的特征提取

本文采用PVT v2作为特征提取的骨干网络。具体来说,PVT v2结合了CNN中的金字塔结构,使用渐进式收缩金字塔来生成多尺度特征,随着网络深度增加,Transformer的序列长度逐渐减小,从而显

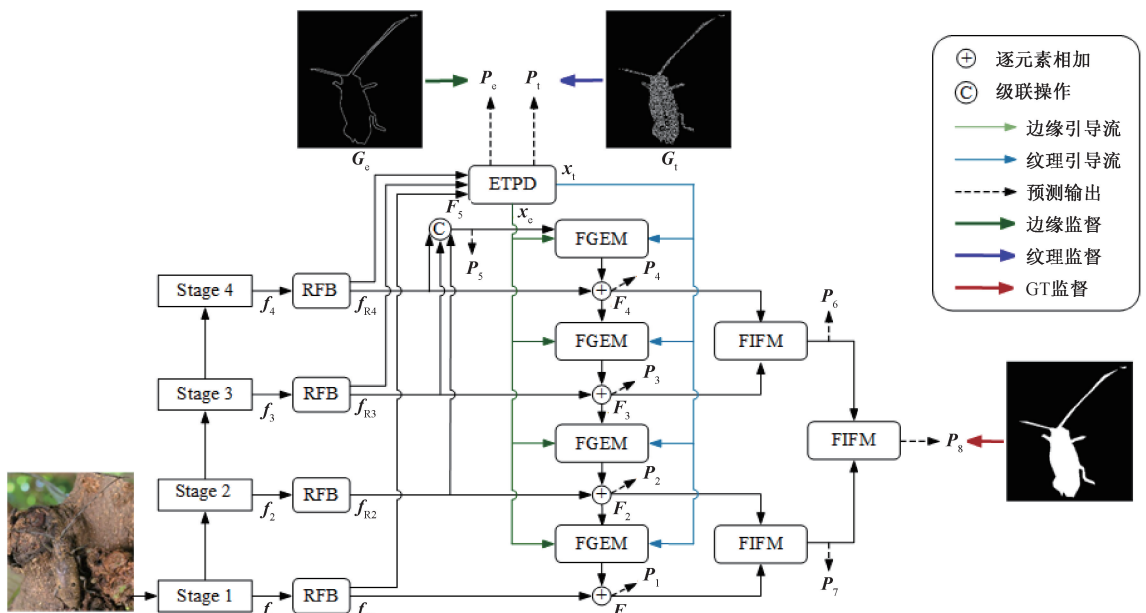


图1 ETGENet整体结构

Figure 1 Overall structure of ETGENet

著减少了计算量。同时,利用线性空间缩减注意力的设计,PVT v2 进一步缩减了自注意力计算时的空间维度,从而有效降低了学习高分率特征时的资源消耗。此外,PVT v2 使用重叠补丁嵌入的序列化方式在相邻图像的 Patch 之间建立重叠区域,使得模型能够在全局建模的同时更好地捕捉图像中的局部细节信息。

1.3 边缘-纹理感知解码器

在 COD 任务中,良好的边缘先验信息常被作为有效的辅助线索广泛应用^[7]。同时,伪装物体内部丰富的梯度信息也是一个有价值的探索方向^[10]。故本文同时结合两种线索来辅助检测任务。具体来说,本文提出的边缘-纹理感知解码器(ETPD)包含两个结构相同的解码器,即边缘感知解码器(edge perception decoder, EPD)和纹理感知解码器(texture perception decoder, TPD),分别用于生成具有针对性的伪装对象的边缘特征 x_e 与纹理特征 x_t 。以 EPD 为例,使用经 RFB 增强后的骨干特征 f_{R1} 、 f_{R3} 和 f_{R4} 进行解码。其中,低层特征包含了丰富的局部细节线索^[13],而高层特征则具有能够减少干扰的全局位置信息。图 2 为边缘感知解码器。首先,EPD 对 f_{R1} 与 f_{R3} 、 f_{R3} 与 f_{R4} 进行逐元素乘法,并通过跳跃连接和 3×3 卷积得到跨层级细化特征;其次,对 3 个分支的细化特征进行拼接,并利用 3×3 卷积生成 x_e ;最后,使用 1×1 卷积及 Sigmoid 操作得到伪装对象的边缘预测 P_e 。同理,通过 TPD 也可以获得 x_t 及其纹理预测 P_t 。对于得到的边缘与纹理预测图,本文使用对象边缘图 G_e 和对象纹理图 G_t 对二者进行监督。其中,对象边缘图通过 Canny 算法生成,对象纹理图则直接使用文献[10]中所提供的梯度级纹理标签。为了降低 ETPD 所占用的计算复杂度,这里未使用 f_{R2} 。上述过程可以表示为

$$\begin{cases} P_e = \sigma \text{Conv1}(x_e); \\ P_t = \sigma \text{Conv1}(x_t). \end{cases} \quad (1)$$

$$\begin{aligned} x_e \text{ 或 } x_t &= \text{Conv3}([\mathbf{f}_{R4} \uparrow, \\ &(\text{Conv3}(\text{Cat}(\mathbf{f}_{R4} \uparrow, \mathbf{f}_{R4} \uparrow \otimes \mathbf{f}_{R3}))) \uparrow, \\ &\text{Conv3}(\text{Cat}(\mathbf{f}_{R3} \uparrow, \mathbf{f}_{R3} \uparrow \otimes \mathbf{f}_{R1}))]. \end{aligned} \quad (2)$$

式中: $\sigma(\cdot)$ 为 Sigmoid 操作; $\text{Conv1}(\cdot)$ 为 1×1 卷积; \uparrow 为上采样操作; $\text{Conv3}(\cdot)$ 为 3×3 卷积; $\text{Cat}(\cdot)$ 为级联操作; \otimes 为逐元素相乘。

1.4 特征引导增强模块

在形状大小各异且遮挡严重的复杂伪装场景下,增强网络对于对象边缘和纹理信息的表征能力至关重要。受到文献[14]中改进的自注意力模块

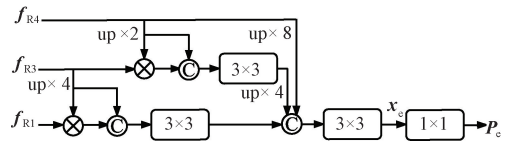


图 2 边缘感知解码器

Figure 2 Edge perception decoder

的启发,本文提出了一种特征引导增强模块 FGEM,以更充分地利用边缘-纹理线索对网络进行引导,并减少多源信息中不稳定因素所带来的干扰。图 3 为特征引导增强模块。以第 i 个层级中的 FGEM 为例,FGEM 接收对象特征 F_{i+1} 和由 ETPD 生成的边缘特征 x_e 、纹理特征 x_t ,将这 3 个特征作为输入,通过自增强分支、边缘引导分支以及纹理引导分支对对象特征进行引导和优化。

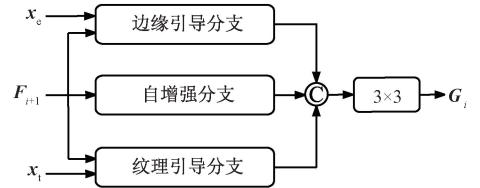


图 3 特征引导增强模块

Figure 3 Feature guide enhancement module

对于边缘引导分支,通过构建跨通道间的相关性矩阵来隐式地对不同特征之间的全局上下文关系进行建模,从而获得经过引导增强后的对象特征。图 4 为边缘引导分支。

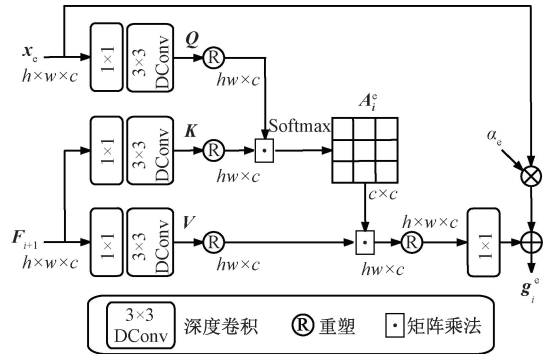


图 4 边缘引导分支

Figure 4 Edge guide branch

如图 4 所示,首先,对输入的边缘特征 x_e 和对象特征 F_{i+1} 进行卷积处理,即通过 1×1 卷积分别聚合两种特征不同通道上的信息,再利用 3×3 深度卷积在通道级别上对空间上下文特征进行编码,从而捕捉特征的局部结构。经过这个过程, x_e 生成查询 Q ,而 F_{i+1} 生成键 K 和值 V :

$$\begin{cases} Q = \text{DConv3}(\text{Conv1}(x_e)); \\ K = \text{DConv3}(\text{Conv1}(F_{i+1})); \\ V = \text{DConv3}(\text{Conv1}(F_{i+1})). \end{cases} \quad (3)$$

式中: $\text{DConv3}(\cdot)$ 表示 3×3 深度卷积。其次,对 \mathbf{Q} 和 \mathbf{K} 进行重塑,并对二者进行矩阵相乘和 Softmax 操作,以生成引导注意力权重矩阵 $\mathbf{A}_i^e, \mathbf{A}_i^s$ 中的每个像素分别代表了两个向量间对应位置的相似性度量,反映了边缘特征与对象特征之间的相关程度。其中,相关程度较高的有效特征被分配更大的权重,而相关程度较低的噪声干扰将得到抑制。再次,通过矩阵乘法运算,利用 \mathbf{A}_i^e 对重塑后的 \mathbf{V} 进行加权,从而引导模型关注与边缘特征高度相关的对象特征区域,最终生成具有针对性的对象特征。最后,将经过引导注意力细化后的对象特征进行重塑和 1×1 卷积处理,并与加权后的 \mathbf{x}_e 相加得到受边缘特征引导增强后的对象特征 \mathbf{g}_i^e 。上述过程可写为

$$\begin{cases} \mathbf{g}_i^e = \text{Conv1}(\mathbf{V} \cdot \mathbf{A}_i^e) + \alpha_e \otimes \mathbf{x}_e; \\ \mathbf{A}_i^e = \text{Softmax}(\mathbf{Q} \cdot \mathbf{K}^T / \alpha). \end{cases} \quad (4)$$

式中: α_e 为可学习的权重参数,用于自适应地对边缘特征进行补充; $\text{Softmax}(\cdot)$ 为 Softmax 操作; α 为可学习的缩放因子,用于确保 \mathbf{K} 与 \mathbf{Q} 的点积在一个合理的范围内。

类似于边缘引导分支,纹理引导分支以纹理特征 \mathbf{x}_t 和对象特征 \mathbf{F}_{i+1} 作为输入,生成受纹理特征引导增强后的对象特征 \mathbf{g}_i^t :

$$\begin{cases} \mathbf{g}_i^t = \text{Conv1}(\mathbf{V} \cdot \mathbf{A}_i^t) + \alpha_t \otimes \mathbf{x}_t; \\ \mathbf{A}_i^t = \text{Softmax}(\mathbf{Q} \cdot \mathbf{K}^T / \beta). \end{cases} \quad (5)$$

式中: α_t 为可学习的权重参数,用于自适应地对纹理特征进行补充; β 为可学习的缩放因子。

自增强分支仅以对象特征 \mathbf{F}_{i+1} 作为输入,通过自注意力计算的方式来捕捉图像的全局上下文信息,从而得到细化特征 \mathbf{g}_i :

$$\begin{cases} \mathbf{g}_i = \text{Conv1}(\mathbf{V} \cdot \mathbf{A}_i) + \alpha_s \otimes \mathbf{F}_{i+1}; \\ \mathbf{A}_i = \text{Softmax}(\mathbf{Q} \cdot \mathbf{K}^T / \gamma). \end{cases} \quad (6)$$

式中: α_s 为可学习的权重参数,用于自适应地对对象特征进行补充; γ 为可学习的缩放因子。

将 3 个分支中的细化特征进行拼接,从而获得多特征引导增强后的对象特征 \mathbf{G}_i 。 \mathbf{G}_i 与 \mathbf{f}_{R_i} 相加得到 \mathbf{F}_i 后被送到更浅层的 FGEM 中,从而生成更精细化的特征。同时, \mathbf{F}_i 在 1×1 卷积和进行上采样操作后生成粗略预测 $\{\mathbf{P}_i, i \in 1, 2, 3, 4, 5\}$, 通过 GT 对网络进行深层监督。其中,第 i 个层级的聚合特征 \mathbf{F}_i 可以写为

$$\begin{cases} \mathbf{F}_i = \mathbf{G}_i + \mathbf{f}_{R_i}; \\ \mathbf{G}_i = \text{Conv3}(\text{Cat}(\mathbf{g}_i^e, \mathbf{g}_i^t, \mathbf{g}_i)). \end{cases} \quad (7)$$

1.5 特征交互融合模块

现有的 COD 方法通常使用自上而下的特征融

合方式来聚合多层次特征,但高级特征中的对象语义会因此受到显著削弱。故本文提出了一种渐进融合的策略,通过 3 个特征交互融合模块 FIFM 来生成网络的最终预测结果。图 5 为特征交互融合模块。相较于简单的特征融合方法, FIFM 采用了一种注意力交互融合的方式来获取更有效、更全面的特征表示。

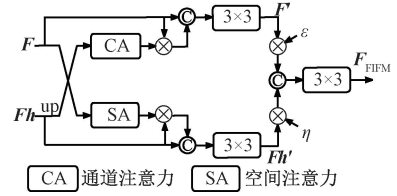


图 5 特征交互融合模块

Figure 5 Feature interaction fusion module

具体来说,对于输入的相邻特征,将低级特征记为 \mathbf{F} ,高级特征记为 \mathbf{Fh} 。为了充分探索相邻特征间的差异并挖掘互补信息,且同时关注高级特征中的通道信息和低级特征中的局部空间线索, FIFM 使用 \mathbf{F} 生成的空间注意力^[15]与 \mathbf{Fh} 进行逐元素相乘运算,并将得到的结果与 \mathbf{Fh} 进行拼接得到 \mathbf{Fh}' ,以补充 \mathbf{Fh} 中的空间信息。同样,也利用 \mathbf{Fh} 生成的通道注意力^[15]与 \mathbf{F} 进行逐元素相乘,将得到的结果与 \mathbf{F} 进行拼接得到 \mathbf{F}' ,以弥补 \mathbf{F} 中的通道信息的不足。在这种双向交互后,两分支中的特征都可以学习到来自另外分支中的信息。上述过程分别写为

$$\begin{cases} \mathbf{F}' = \text{Conv3}(\text{Cat}(\mathbf{F}, \mathbf{F} \otimes \text{CA}(\mathbf{Fh}))); \\ \mathbf{Fh}' = \text{Conv3}(\text{Cat}(\mathbf{Fh}, \mathbf{Fh} \otimes \text{SA}(\mathbf{F}))). \end{cases} \quad (8)$$

$$\begin{cases} \text{SA}(\mathbf{F}) = \sigma(\text{Conv7}(\text{Cat}(\text{Avg}(\mathbf{F}), \text{Max}(\mathbf{F}))))); \\ \text{CA}(\mathbf{Fh}) = \sigma(\text{MLP}(\text{Avg}(\mathbf{Fh})) + \text{MLP}(\text{Max}(\mathbf{Fh}))). \end{cases} \quad (9)$$

式中: $\text{SA}(\cdot)$ 与 $\text{CA}(\cdot)$ 分别表示空间注意力和通道注意力; $\text{Avg}(\cdot)$ 和 $\text{Max}(\cdot)$ 分别表示全局平均池化和全局最大池化操作; $\text{MLP}(\cdot)$ 为共享多层感知机; $\text{Conv7}(\cdot)$ 为 7×7 卷积。

本文对两个分支中的特征进行加权融合,从而生成关键信息更丰富且噪声更小的聚合特征 \mathbf{F}_{FIFM} :

$$\mathbf{F}_{\text{FIFM}} = \text{Conv3}(\text{Cat}(\varepsilon \otimes \mathbf{F}', \eta \otimes \mathbf{Fh}')). \quad (10)$$

式中: ε 和 η 为两个可学习的权重参数,这两个参数使网络自适应地聚合两个相邻特征。

在 ETGENet 中,每个 FIFM 输出的聚合特征均将在 1×1 卷积和上采样操作后生成对应的预测图 $\{\mathbf{P}_i, i \in 6, 7, 8\}$, 本文以 \mathbf{P}_8 作为网络的最终预测。

1.6 损失函数

ETGENet 的损失函数包括 3 种监督,即对象监

督、边缘监督和纹理监督。受到文献[2]的启发,在对象监督中,本文使用二元交叉熵损失 L_{wBCE} 和加权交并比损失 L_{wIoU} 来对检测结果中的像素级差异和全局结构进行研究。这两种损失的计算方式与文献[16]中一致。

同时,受到文献[7]的启发,本文使用 Dice 损失 L_{Dice} 进行边缘监督,该损失可以更好地处理图像中不同类别间的像素数量不平衡的问题,故能够更有效地关注边缘部分的像素,有助于模型产生更加准确的边缘预测。对于纹理监督,采用文献[10]中的标准均方误差损失函数 L_{MSE} 。本文所提出的模型共包含 10 个预测输出,分别为 5 个粗略图 $P_1 \sim P_5$ 、3 个细化图 $P_6 \sim P_8$ 、1 个边缘图 P_e 和 1 个纹理图 P_t ,因此,网络的总体损失 L_{total} 可以表示为

$$L_{total} = \sum_{i=1}^8 L(P_i, GT) + \lambda L_{Dice}(P_e, G_e) + \lambda L_{MSE}(P_t, G_t); \quad (11)$$

$$L = L_{wBCE} + L_{wIoU} \quad (12)$$

式中: λ 为权衡参数,在实验中设置其值为 3。

2 实验设计

2.1 数据集

本文在 CAMO、COD10K 和 NC4K 这 3 个流行的公共基准数据集上评估了所提出的方法。其中,CAMO 共包含了 1 250 张从互联网上收集的伪装图像,涵盖了 8 个伪装类别,1 000 张图片被用于训练,剩余的 250 张图片被用于测试。COD10K 是目前包含图像数量最多的 COD 数据集,研究人员通常使用其中的 3 040 张伪装图像进行 COD 方法的训练,2 026 张进行测试。NC4K 是目前最大的 COD 测试

数据集,包含了 4 121 张伪装图像。

2.2 评估指标

COD 广泛认可且最常用的评估指标为结构度量(S)^[17]、加权 F 度量(F)^[18]、平均绝对误差(M)和自适应增强匹配度量(E)^[19]。其中, S 用于评价两个图像的结构内容间的相似性; F 为综合考虑精确率和召回率的指标; M 用于计算地面真值和预测图像间的绝对像素差异; E 用于评估预测图中全局信息和局部信息的精度。评估时, S 、 F 和 E 越大越好, M 越小越好。

2.3 实验细节

本文使用 PyTorch 深度学习框架,所有的实验均在 Nvidia GeForce RTX 3090 GPU(24 GB 显存)上运行。所有输入图像的大小设置为 384×384 ,批量大小为 24。实验中使用 Adam 优化器,初始学习率设置为 $1e-4$,每 50 个周期学习率除以 10,整个训练过程共包含 100 个周期。

3 实验结果与分析

3.1 定量评估

为了展示所提方法的优越性,本文将其与 12 个 COD 方法进行了比较,表 1 为定量评估结果。由表 1 可知,ETGNet 在所有评估指标上均取得了优异的性能。与同样使用了边缘感知的方法(EAMNet、BGNet 和 BSA-Net)和纹理感知的方法(DGNet)相比,ETGNet 取得了明显的性能提升,其中,在最大的测试数据集 NC4K 上,ETGNet 的 F 平均提升了 5.2 百分点,这一性能提升主要归功于本文提出的 FGEM 和 FIFM 模块,它们通过有效整合边缘与纹理线索来引导特征优化,从而显著提高了预测精度。

表 1 定量评估结果

Table 1 Quantitative evaluation results

基线模型	CAMO				COD10K				NC4K			
	S	E	F	M	S	E	F	M	S	E	F	M
SINet ^[2]	0.751	0.834	0.606	0.100	0.771	0.797	0.551	0.051	0.808	0.882	0.723	0.058
LSR ^[3]	0.775	0.847	0.673	0.088	0.814	0.865	0.666	0.035	0.833	0.893	0.739	0.053
BSA-Net ^[8]	0.796	0.859	0.717	0.079	0.818	0.894	0.699	0.034	0.841	0.903	0.771	0.048
BGNet ^[7]	0.812	0.876	0.749	0.073	0.831	0.902	0.722	0.033	0.851	0.911	0.788	0.044
ZoomNet ^[4]	0.820	0.878	0.752	0.066	0.838	0.892	0.729	0.029	0.853	0.904	0.784	0.043
DGNet ^[10]	0.839	0.900	0.769	0.057	0.822	0.876	0.693	0.033	0.857	0.907	0.784	0.042
PENet ^[20]	0.828	0.890	0.771	0.063	0.831	0.908	0.723	0.031	0.855	0.912	0.795	0.042
EAMNet ^[9]	0.831	0.890	0.763	0.064	0.839	0.907	0.733	0.029	0.862	0.916	0.801	0.040
FDNet ^[5]	0.836	0.886	0.777	0.066	0.857	0.918	0.763	0.028	0.865	0.911	0.803	0.042
FEDER ^[21]	0.802	0.870	0.738	0.071	0.822	0.900	0.716	0.032	0.847	0.910	0.789	0.044
FSPNet ^[6]	0.856	0.919	0.799	0.050	0.851	0.900	0.735	0.026	0.879	0.923	0.816	0.035
DINet ^[22]	0.821	0.883	—	0.068	0.832	0.901	—	0.031	0.856	0.910	—	0.043
ETGNet	0.863	0.930	0.826	0.047	0.861	0.924	0.777	0.023	0.882	0.936	0.838	0.032

3.2 定性评估

图6显示了本文模型与较先进的方法以及同样引入了边缘或纹理信息的方法的定性评估结果。由图6可以看出,ETGENet能够检测出更完整的对象结构且能抑制噪声的干扰。同时,其预测结果的边缘细节还原度更好,并能更准确定位和分割目标区域,拥有更优越的性能。

3.3 消融实验

为了验证引入边缘和纹理信息以及两个关键模块FGEM和FIFM的有效性,本文在COD10K与NC4K数据集上对本文方法ETGENet进行了消融实

验。表2为消融实验定量评估结果。表2中,对于基线模型a,本文删除了全部的附加模块(即ETPD、FGEM和FIFM),仅使用RFB模块处理骨干特征,并通过相加操作来自上而下地聚合多级特征生成最终预测;模型b、d、f在模型a的基础上分别引入边缘信息、纹理信息以及边缘和纹理信息,这些信息通过相加的方式对对象特征进行引导并生成预测;模型c、e、g将模型b、d、f中相加的引导方式替换为所提出的FGEM;模型h在模型a的基础上将自上而下的特征融合方式替换为FIFM;模型i则是最终模型。

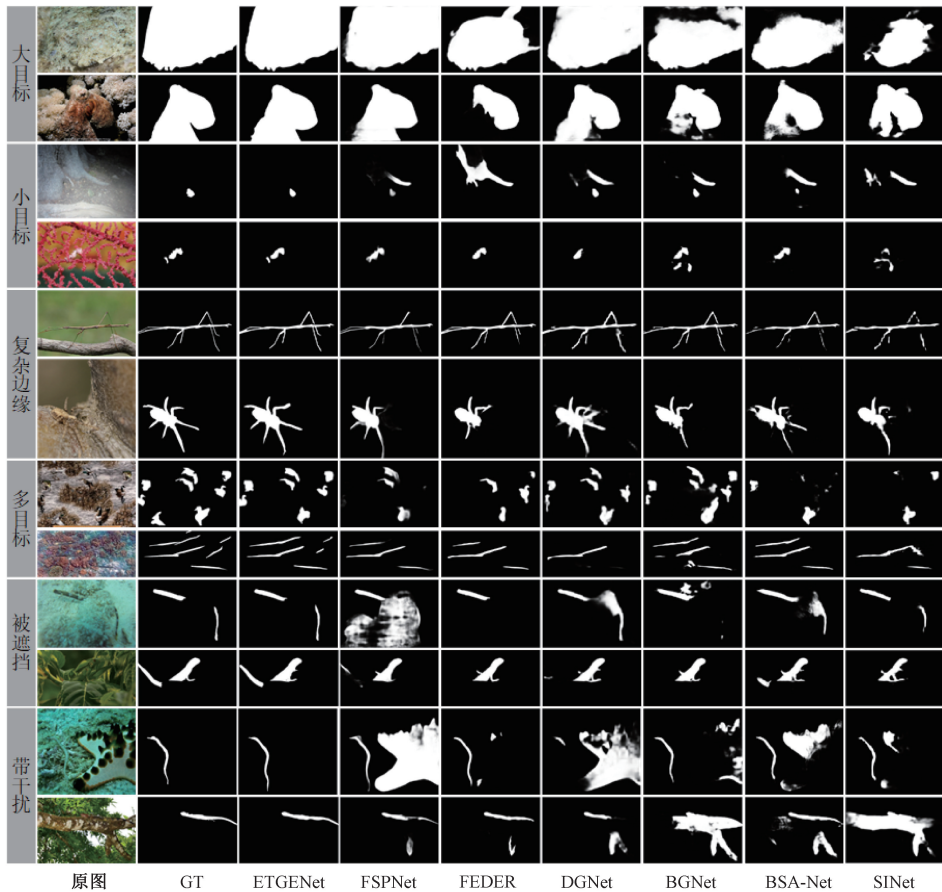


图6 定性评估结果

Figure 6 Qualitative evaluation results

表2 消融实验定量评估结果

Table 2 Quantitative evaluation results of ablation experiments

模型编号	实验策略	COD10K				NC4K			
		<i>S</i>	<i>E</i>	<i>F</i>	<i>M</i>	<i>S</i>	<i>E</i>	<i>F</i>	<i>M</i>
a	Baseline	0.843	0.907	0.744	0.027	0.872	0.928	0.820	0.035
b	Baseline+EPD	0.850	0.911	0.754	0.026	0.875	0.928	0.825	0.035
c	Baseline+EPD+FGEM	0.853	0.917	0.764	0.024	0.877	0.932	0.830	0.034
d	Baseline+TPD	0.846	0.914	0.750	0.025	0.874	0.930	0.824	0.034
e	Baseline+TPD+FGEM	0.850	0.917	0.759	0.024	0.875	0.931	0.828	0.034
f	Baseline+ETPD	0.854	0.914	0.761	0.025	0.878	0.931	0.829	0.034
g	Baseline+ETPD+FGEM	0.857	0.920	0.770	0.024	0.880	0.934	0.835	0.033
h	Baseline+FIFM	0.853	0.918	0.766	0.025	0.878	0.934	0.833	0.033
i	ETGENet	0.861	0.924	0.777	0.023	0.882	0.936	0.838	0.032

3.3.1 边缘与纹理信息的有效性

本文首先评估了引入边缘与纹理信息对模型的贡献。如表 2 所示,模型 b、d 和 f 在所有评估指标上均优于模型 a,同时模型 f 相较于模型 b、d 实现了更好的整体性能。图 7 为引入边缘和纹理信息的消融实验对比。从图 7 可以看出,相比于基线模型 a,在添加边缘信息后,预测结果中的边缘细节更加准确;在添加纹理信息后,网络预测中的假阴性情况得到了明显改善;而同时添加边缘和纹理信息后,网络获得了边缘细节更充足且对象更完整的检测结果。

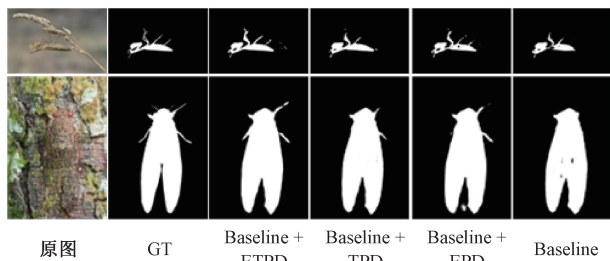


图 7 引入边缘和纹理信息的消融实验对比

Figure 7 Comparison of ablation experiments of introducing edge and texture information

3.3.2 特征引导增强模块的有效性

为了验证 FGEM 的有效性,本文对模型 b 和 c、模型 d 和 e 以及模型 f 和 g 的实验结果进行了深入分析。由表 2 可知,在添加了 FGEM 后,各模型的检测性能均有所提升,这是由于该模块能够利用引导注意力进行更充分的特征引导。

图 8 为特征引导增强模块的消融实验对比。从图 8(a)~8(c)的定性评估结果可以看出,在模型 b 的基础上添加 FGEM 后,ETGENet 生成的检测结果中与对象相关的边缘区域更加完整、精细;在模型 d 的基础上添加 FGEM 后,可以减少直接引入纹理信息所导致的假阴性或假阳性错误;在模型 f 的基础上添加 FGEM 后,检测结果更完整、噪声更小以及在边缘细节上表现更好,这也充分证明了设计的模型可以更好地利用边缘和纹理信息辅助检测任务,从而提升分割质量。

3.3.3 特征交互融合模块的有效性

本文使用两组不同的网络来分析 FIFM 所起到的作用,分别为模型 a 和 h、模型 g 和 i。由表 2 可知,包含 FIFM 的模型(h 和 i)可以在渐进的相邻层级聚合过程中获得更好的 COD 性能。图 9 为特征交互融合模块的消融实验对比。从图 9(a)可以看出,添加 FIFM 后,网络的预测结果更加精细和完整;从图 9(b)可以看出,当在模型 i 的基础上去除

FIFM 得到模型 g 时,网络的预测中容易丢失对象信息或者产生假阳性错误,这进一步验证了所提出的 FIFM 的有效性。

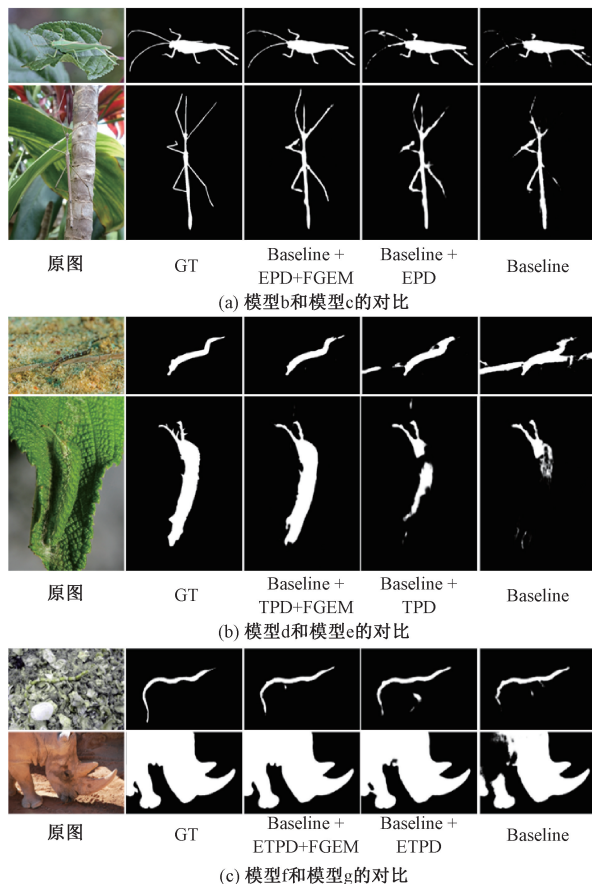


图 8 特征引导增强模块的消融实验对比

Figure 8 Comparison of ablation experiments of FGEM

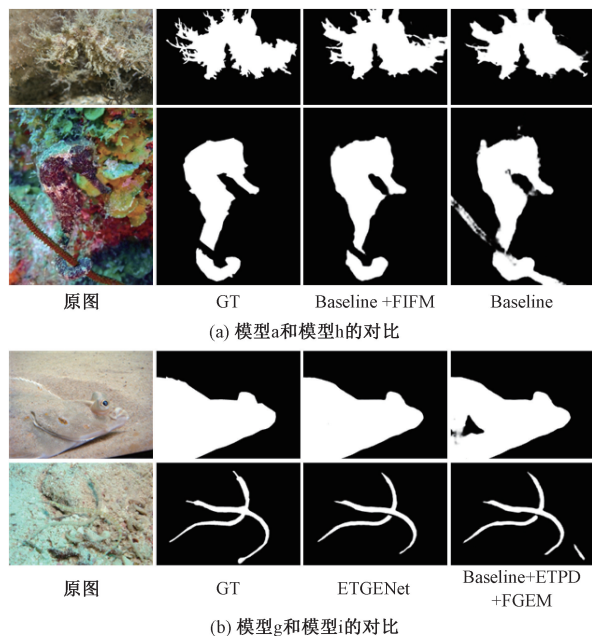


图 9 特征交互融合模块的消融实验对比

Figure 9 Comparison of ablation experiments of FIFM

4 结论

本文提出了一种边缘-纹理引导增强网络 ET-GENet 来进行精确的伪装目标检测。具体来说,网络中的边缘与纹理感知解码器可以有效获取与伪装对象相关的边缘和纹理信息。同时,所提出的特征引导增强模块能够通过多分支的并行处理和引导注意力细化操作使得网络更充分地利用所提取的多种类型对象线索,并增强网络的抗干扰能力。此外,特征交互融合模块能够通过交互和加权融合策略对不同层级的网络特征进行有效聚合,从而生成更完整的网络预测结果。在 3 个公共数据集上与其他方法的对比结果表明,所提出的 ETGENet 表现出更好的检测性能。在未来的工作中,将对网络进行轻量化研究,以实现更加高效的伪装目标检测。

参考文献:

- [1] LE T N, NGUYEN T V, NIE Z L, et al. Anabranched network for camouflaged object segmentation[J]. *Computer Vision and Image Understanding*, 2019, 184: 45-56.
- [2] FAN D P, JI G P, SUN G L, et al. Camouflaged object detection[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway:IEEE, 2020: 2774-2784.
- [3] LV Y Q, ZHANG J, DAI Y C, et al. Simultaneously localize, segment and rank the camouflaged objects[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway:IEEE, 2021: 11586-11596.
- [4] PANG Y W, ZHAO X Q, XIANG T Z, et al. Zoom in and out: a mixed-scale triplet network for camouflaged object detection [C] // 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2022: 2150-2160.
- [5] SONG Y X, LI X Y, QI L. Camouflaged object detection with feature grafting and distractor aware[C]//2023 IEEE International Conference on Multimedia and Expo. Piscataway:IEEE, 2023: 2459-2464.
- [6] HUANG Z, DAI H, XIANG T Z, et al. Feature shrinkage pyramid for camouflaged object detection with transformers[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2023: 5557-5566.
- [7] SUN Y J, WANG S, CHEN C, et al. Boundary-guided camouflaged object detection[EB/OL]. (2022-07-02) [2024-11-20]. <https://doi.org/10.48550/arXiv.2207.00794>.
- [8] ZHU H W, LI P, XIE H R, et al. I can find you! boundary-guided separated attention network for camouflaged object detection [C] // Proceedings of the 36th AAAI Conference on Artificial Intelligence. Palo Alto: AAAI, 2022: 3608-3616.
- [9] SUN D Y, JIANG S Y, QI L. Edge-aware mirror network for camouflaged object detection[C]//2023 IEEE International Conference on Multimedia and Expo. Piscataway: IEEE, 2023: 2465-2470.
- [10] JI G P, FAN D P, CHOU Y C, et al. Deep gradient learning for efficient camouflaged object detection [J]. *Machine Intelligence Research*, 2023, 20(1): 92-108.
- [11] WANG W H, XIE E Z, LI X, et al. PVT v2: improved baselines with pyramid vision transformer[J]. *Computational Visual Media*, 2022, 8(3): 415-424.
- [12] LIU S T, HUANG D, WANG Y H. Receptive field block net for accurate and fast object detection [EB/OL]. (2017-11-21) [2024-11-20]. <https://doi.org/10.48550/arXiv.1711.07767>.
- [13] 田旭, 彭飞, 刘飞, 等. 基于金字塔特征与边缘优化的显著性对象检测[J]. *郑州大学学报(工学版)*, 2022, 43(2): 35-43.
TIAN X, PENG F, LIU F, et al. Salient object detection based on pyramid features and edge optimization [J]. *Journal of Zhengzhou University (Engineering Science)*, 2022, 43(2): 35-43.
- [14] ZAMIR S W, ARORA A, KHAN S, et al. Restormer: efficient transformer for high-resolution image restoration [C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2022: 5718-5729.
- [15] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[EB/OL]. (2018-07-17) [2024-11-20]. <https://doi.org/10.48550/arXiv.1807.06521>.
- [16] WEI J, WANG S H, HUANG Q M. F³Net: fusion, feedback and focus for salient object detection[C]// Proceedings of the 34th AAAI Conference on Artificial Intelligence. Palo Alto: AAAI, 2020: 12321-12328.
- [17] CHENG M M, FAN D P. Structure-measure: a new way to evaluate foreground maps[J]. *International Journal of Computer Vision*, 2021, 129(9): 2622-2638.
- [18] MARGOLIN R, ZELNIK-MANOR L, TAL A. How to evaluate foreground maps [C]//2014 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2014: 248-255.
- [19] FAN D P, GONG C, CAO Y, et al. Enhanced-alignment measure for binary foreground map evaluation [EB/OL]. (2018-05-26) [2024-11-20]. <https://doi.org/10.48550/arXiv.1805.10421>.
- [20] LI X F, YANG J X, LI S H, et al. Locate, refine and restore: a progressive enhancement network for camou-

- flaged object detection [C] // Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence. Cape Town; International Joint Conferences on Artificial Intelligence Organization, 2023; 1116–1124.
- [21] HE C M, LI K, ZHANG Y C, et al. Camouflaged object detection with feature decomposition and edge reconstruction [C] // 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway; IEEE, 2023; 22046–22055.
- [22] ZHOU X F, WU Z C, CONG R M. Decoupling and integration network for camouflaged object detection [J]. IEEE Transactions on Multimedia, 2024, 26; 7114–7129.

Edge-texture Guided Enhancement Network for Camouflaged Object Detection

WEI Mingjun^{1,2}, CHEN Xiaoru¹, LIU Ming¹, LIU Yazhi^{1,2}, LI Hui¹

(1. College of Artificial Intelligence, North China University of Science and Technology, Tangshan 063210, China; 2. Hebei Provincial Key Laboratory of Industrial Intelligent Perception, Tangshan 063210, China)

Abstract: Camouflaged object detection (COD) is facing significant challenges due to the high similarity between target objects and their background, such as blurred edge predictions, incomplete detection results, and interference caused by the insufficient use of edge and texture information. To address the issues of current COD, a novel edge-texture guided enhancement network (ETGENet) was proposed to further improve the performance of COD through explicit and sufficient edge-texture guidance strategies. Firstly, a key feature guided enhancement module (FGEM) was used in ETGENet, which could use parallel feature refinement branches to process and enhance object features. The guide branch could obtain object features by guiding attention correlation with edge and texture cues to enhance the network's understanding of object details and suppress noise interference. While the self-enhancement branch could use the self-attention mechanism to refine the characteristics of camouflaged objects from a global perspective. Secondly, a feature interaction fusion module (FIFM) was also proposed to progressively fuse adjacent features. FIFM could utilize the attention interaction mechanism and weighted fusion strategy to learn complementary information between features to generate more complete predicted map. Experiments on three public datasets CAMO, COD10K, and NC4K demonstrate that the proposed network outperformed state-of-the-art methods in the field across metrics such as structure measure S , adaptive enhanced matching measure E , weighted F -measure, and mean absolute error M . Notably, on the largest test set, NC4K, the weighted F -measure surpassed the best-performing method among the 12 advanced COD methods, FSPNet, by 2.2 percentage points.

Keywords: camouflaged object detection; edge information; texture information; feature guide; feature interaction