Computer Science and Application 计算机科学与应用, 2025, 15(10), 24-33
Published Online October 2025 in Hans. https://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2025.1510247

Hans汉斯

基于BGNet边缘与上下文融合的改进

武新浩

西京学院计算机学院，陕西 西安

收稿日期：2025年9月4日；录用日期：2025年10月3日；发布日期：2025年10月10日

摘　要

伪装目标的检测是一项极具价值和挑战性的任务，针对现有深度学习网络对于伪装目标识别率不高、边缘识别不精准等问题，本文提出一种基于BGNet改进的边缘双注意力自适应融合与上下文融合的伪装目标识别模型(TGDNet)，具有更优的分割识别性能。本文使用CAMO、CHAMELEON、COD10K、NC4K数据集进行训练，与SINET、UGTR、LSR等神经网络模型对比测试，测试结果表明我们提出的TGDNet模型收敛速度快且稳定，在关键指标S-measure、E-measure、F-measure、MAE上均优于对比模型。

关键词

人工智能，伪装目标识别，Transformer，融合边缘引导，上下文融合

Improvement of Edge and Context Fusion Based on BGNet

Xinhao Wu

School of Computer Science, Xijing University, Xi'an Shaanxi

Received: September 4, 2025; accepted: October 3, 2025; published: October 10, 2025

Abstract

The detection of disguised targets is a highly valuable and challenging task. In response to the low recognition rate and inaccurate edge recognition of existing deep learning networks for disguised targets, a disguised target recognition model (TGDNet) based on improved edge dual-attention adaptive fusion

文章引用: 武新浩. 基于 BGNet 边缘与上下文融合的改进[J]. 计算机科学与应用, 2025, 15(10): 24-33.
DOI: 10.12677/csa.2025.1510247

and context fusion of BGNet is proposed in this paper to improve its segmentation and recognition performance. This article uses the CAMO, CHAMELEON, COD10K, and NC4K datasets for training, and compares them with neural network models such as SINET, UGTR, LSR, etc. The test results show that our proposed TGDNet model has fast and stable convergence speed, and outperforms the comparison model in key indicators, including S-measure, E-measure, F-measure, and MAE.

Keywords

Artificial Intelligence, Disguised Target Recognition, Transformer, Fusion Edge Guidance, Context Fusion

1. 引言

伪装目标检测作为一种重要视觉任务近期被广泛关注，其原因在于人们发现伪装目标检测具有极大的应用价值。例如：在物种发现领域，自然界动物通常会拥有天生的伪装使得物种检测人员极其不容易发现从而遗漏一些濒危或人类未发现物种；在军事领域，作战装备、人员、设施一般情况下均使用伪装喷涂进行伪装，使得在军事作战中作战人员不能直接发现作战单位，从而使得自身目标威胁。伪装目标检测在各行业和领域均拥有着极大的价值。然而，由于伪装目标检测与其他目标检测任务不同，被检测目标的种类不同导致其大小不一、形态不同，且纹理和边缘与背景高度相似，使得对此目标检测不易，因此伪装目标检测任务是一项颇具挑战性的任务。

2. 相关工作

在早期研究中，Fan 等人[1]首次构建了一种基于深度卷积神经网络的伪装目标检测模型 SINet。该模型借助搜索定位机制扩展感受野以优化目标预测，并通过融合深层与浅层特征，实现对有效信息的联合筛选。Sun 等人[2]提出了一种上下文感知交叉网络 C^2FNet，采用逐级融合深层特征并提取关键信息的方式，结合双分支融合模块与上下文建模机制，增强了模型对全局与局部特征的感知能力，从而提升了检测结果的精细化程度。Wang 等人[3]设计了一种交叉优化网络 D^2CNet，利用双分支结构进行多尺度特征提取与融合，并通过串联方式整合有效信息，逐步优化最终输出。Ren 等人[4]开发了一种纹理感知网络 TANet，聚焦于纹理感知的细化过程，通过提取多尺度特征图并利用纹理感知模块增强前景与背景之间的差异，以强化细节并抑制噪声干扰。

Mei 等人[5]提出了一种定位与聚焦相结合的伪装检测网络 PFNet，其结构包括定位模块(PM)和聚焦模块(FM)，整体借鉴仿生思想，模拟捕食者从全局定位到逐步聚焦的过程，通过串联多个聚焦模块以抑制误检和漏检，提升检测准确性。Fan 等人[6]在 SINet 的基础上进一步改进，提出 SINet-v2，引入了纹理增强模块(TEM)、邻接解码器(NCD)和分组反转注意力机制(GRA)，在搜索阶段强化纹理细节的感知，并通过语义一致性保持和深层特征聚焦进一步提升性能。Qin 等人[7]提出了一种边界感知网络 BASNet，采用 U 型编解码结构进行初步特征提取与定位，并借助边界结构化损失函数优化分割边缘。Fan 等人[8]设计了 PraNet，利用并行解码结构和注意力机制引导深层特征与边缘信息的精细融合，可适用于多种分割任务。

3. 本文方法

3.1. 模型网络结构

本文多尺度双重引导与 Transformer 上下文增强解码器模块整体网络如图 1 所示，整体架构在继承编码器–解码器的基础上，引入了多种设计。编码器采用预训练的 ConvNeXt 作为主干网络，该结构通过分层特征提取获得多尺度特征表示，包括富含细节信息的浅层特征 Stage1、Stage2 和蕴含语义信息的深层特征 Stage3、Stage4。这些特征为后续的边缘提取与融合提供了多层次的信息支持。边缘注意力模块(EAM)通过融合最具语义抽象性的 stage4 与包含丰富细节的 stage1，生成高质量边缘注意力图，作为引导信号贯穿整个网络，在特征融合与增强阶段，本文提出的多尺度双注意力引导融合模块 DGFM 被应用于编码器每一层级，将边缘先验信息与主干特征进行深度融合。该模块通过空间与通道双注意力机制实现精细的特征筛选与增强，使网络更加关注与目标边缘相关的区域和通道。在解码路径中，Transformer 上下文增强解码器 TCM 逐步融合浅层细节特征与深层语义特征，弥补了卷积操作的局部性缺陷。最终，网络通过多尺度预测头输出三个不同的结果及边缘预测输出。

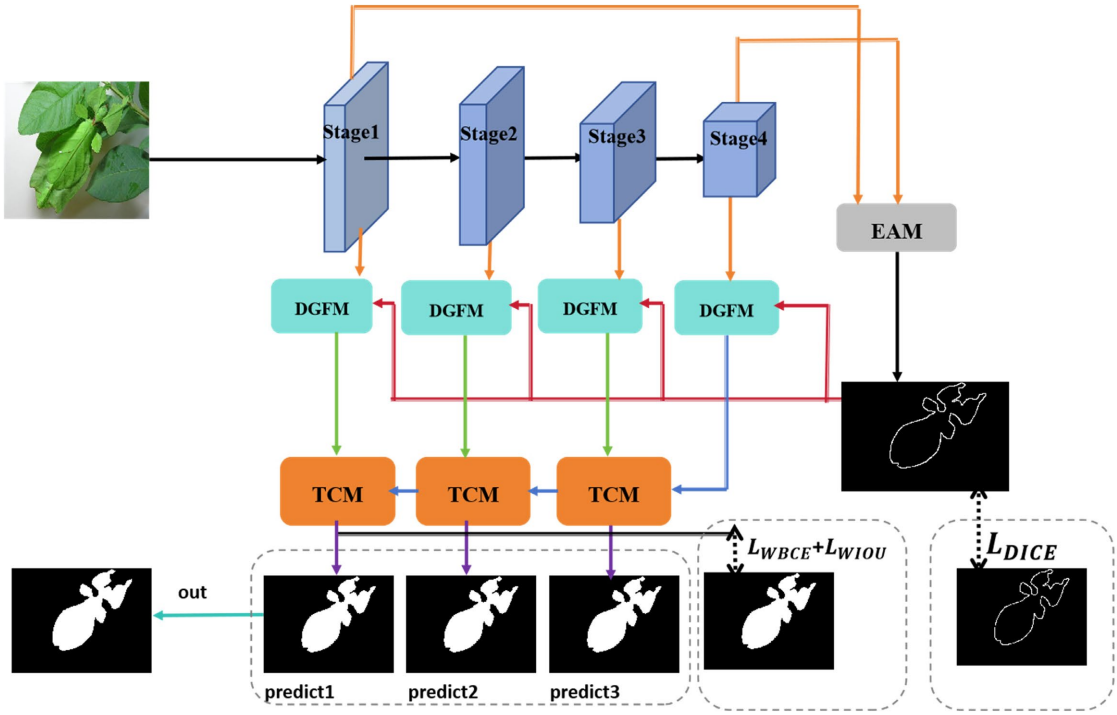

Figure 1. Network architecture diagram
图 1. 网络架构图

3.2. ConvNeXt 的特征提取模块

在深度卷积神经网络的演进中，梯度消失与计算爆炸是核心难题。ResNet 虽通过残差连接(如下图 2 所示)有效缓解了梯度回传的衰减问题，但其特征提取方式，在处理形态极端多变的伪装目标时，在感受野与多尺度捕获上仍存在局限。为此，本研究采用 ConvNeXt 作为主干网络。其核心单元(如下图 3 所示)在保留残差连接这一稳定优化路径的基础上，通过四项关键革新显著提升了多尺度特征表征能力：首先，采用倒置瓶颈结构，先以 1 × 1 卷积扩展通道维度，在更丰富的特征空间中进行后续计算；其次，引入大核深度卷积，将 3 × 3 卷积替换为 7 × 7 深度卷积，极大扩展了感受野以融合更广阔的上下文信息，这对感知与背景

融为一体的伪装目标至关重要；再次，精简激活函数与归一化层，仅在深度卷积前使用一层 LayerNorm，其后接 GELU 激活函数，此举增强了训练稳定性；最后，借鉴 Transformer 设计，大量使用通道更宽的卷积操作。这些革新使 ConvNeXt 在单一模块内高效实现了从局部细节到上下文的多尺度特征提取。

3.3. 边缘双注意引导

原 BGNet 方法直接将边缘检测图与卷积特征图相乘，虽然在一定程度上突出了边缘区域，但忽略了不同尺度特征对分割任务贡献的差异性，导致大量无关背景信息被保留，而关键的目标特征反而未被充分增强。此外，这类方法缺乏动态调整机制，无法根据图像内容自适应地强化重要特征并抑制噪声，在复杂背景或边缘模糊的场景中表现尤为不佳。使得网络难以在目标与背景高度融合的情况下实现准确分割。

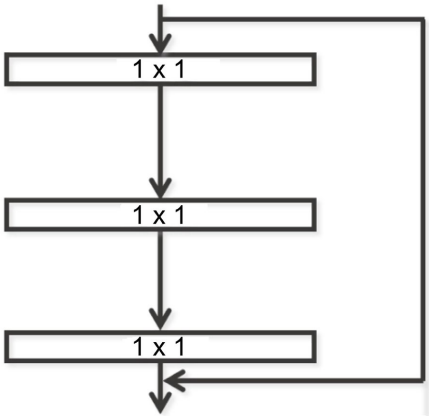

Figure 2. Residual structure
图 2. 残差结构

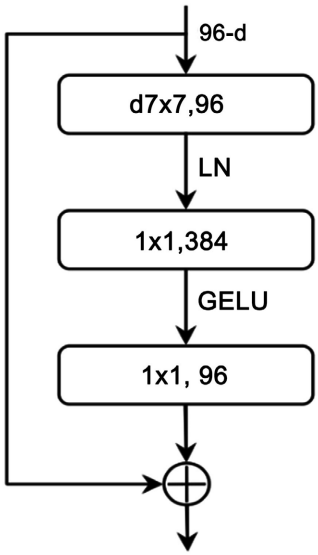

Figure 3. Inverted bottleneck structure
图 3. 倒置瓶颈结构

为克服上述局限性，本文提出了双注意力引导融合模块(DGFM)，如下图 4 所示，旨在实现边缘信息与多尺度主干特征之间的高效与深度融合。该模块通过协同利用空间注意力和通道注意力机制，使网络

能够动态地调整不同位置和通道的特征响应，从而增强对伪装目标关键区域的感知能力。具体而言，DGFM 接收来自主干的层级特征与边缘注意力图 att 作为输入。首先通过空间注意力分支对边缘图进行多尺度卷积，生成空间权重图，再将空间处理后的经由通道注意力，最终使用 1 × 1 卷积完成边缘双注意力动态引导。实验表明，此边缘引导方法优于直接相乘的硬注意力。

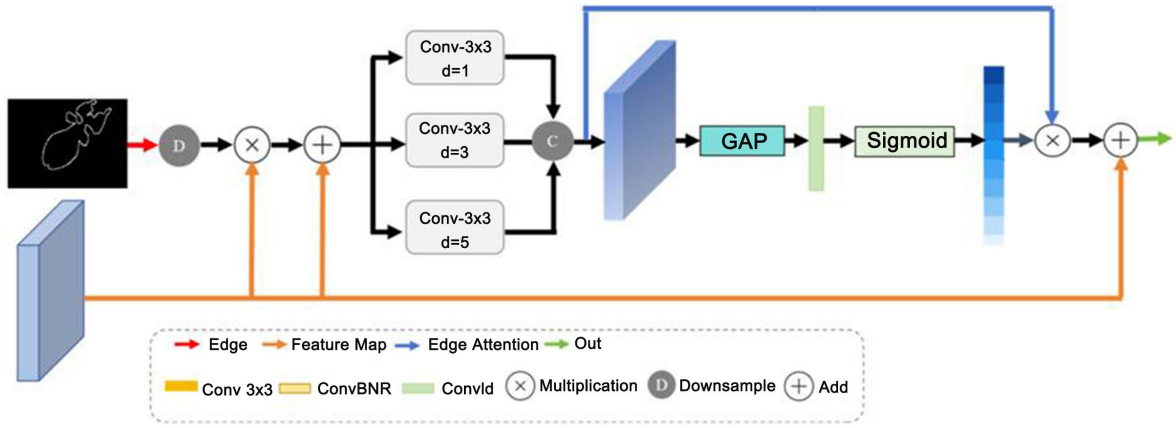

Figure 4. Edge dual attention guidance module
图 4. 边缘双注意力引导模块

3.4. Transformer 上下文增强模块

尽管卷积神经网络在图像分割任务中取得了显著成功，但其固有的局部性限制使其在建模长距离依赖关系方面存在明显不足，这一问题在伪装目标检测中尤为突出。许多基于 CNN 的方法通过堆叠卷积层或使用空洞卷积来扩大感受野，但这些方法往往计算成本高昂且效率低下，难以真正建立全局上下文关联。此外，一些方法尝试引入非局部模块或金字塔池化模块来捕获多尺度上下文，但这些模块通常缺乏对序列化信息的有效建模能力。

为克服上述局限性，本文提出了 Transformer 上下文增强模块(TCM)，如下图 5 所示，是本文模型中用于提升全局语义感知能力的关键组件。该模块利用 Transformer 架构中的自注意力机制，有效地建模图像中的长距离依赖关系，帮助网络在复杂背景中准确识别与分割伪装目标。TCM 的设计用于替代传统的卷积融合模块。首先对输入的高级特征图和上级低级特征图进行特征降维，随后进行词嵌入进入多头注意力机制进行融合，再进行 3 × 3 卷积得到向下逐步融合的特征图，另外再进行 1 × 1 的卷积和 sigmoid 激活函数得到本层级预测效果图。实验表明，Transformer 融合比使用多尺度空洞卷积有着更强的上下文捕获融合能力。

3.5. 损失函数

本文的损失函数是多阶段边缘融合的损失函数，如下式(1)，其中 $loss_{o_i}$ 为 i 层的对象掩码损失(其中每个层级为 WBCE 损失与 WIOU 损失相加)，$loss_{edge_w}$ 为在边缘加权的边缘损失(其中 w 为 DICE 损失)。

$$Loss_{total} = loss_{o1} + loss_{o2} + loss_{o3} + 3.5loss_{edge_w} \tag{1}$$

4. 实验部分

4.1. 数据集

本文采用四个伪装目标公开基准数据集如图 6 (CAMO, CHAMELEON, COD10K, NC4K)进行实验并

且评估本文方法，其中 CAMO 数据集对应于现实世界中的动物和人类，由 1250 张图像组成(训练集 1000 张图像，测试集 250 张图像)，CHAMELEON 拥有 76 张带有手动注释的目标级 ground-truth (GTs)的图像，COD10K 是目前最大规模的伪装目标检测数据集之一，包含多种类型的伪装物体，专为伪装目标检测任务设计，共计 4121 张图像。NC4K 是在自然场景拍摄的伪装目标数据集，拥有 4121 张图像。

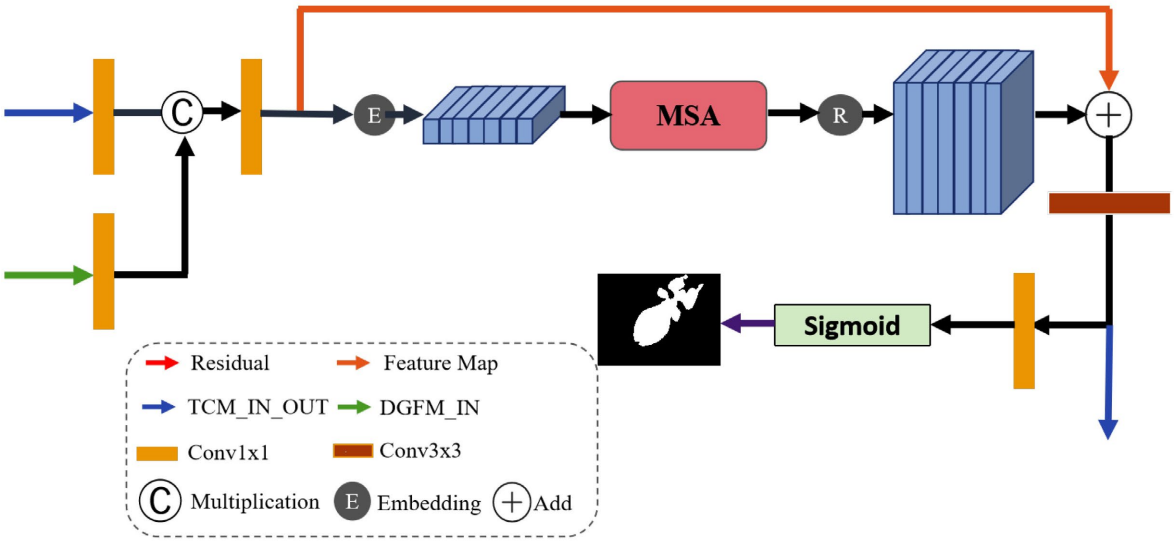

Figure 5. Attention fusion module
图 5. 注意力融合模块

Figure 6. Preview image of the dataset
图 6. 数据集预览图

4.2. 数据集预处理

Figure 7. Edge generation graph
图 7. 边缘生成图

模型需要边缘引导的监督，需要生成对应图像的边缘图，本文采用系统化的边缘提取流程来生成高质量的真实边缘标注图。具体而言，对于数据集中提供的二值化真实分割掩模 GT，使用 OpenCV 库的 Canny 边缘检测算法进行初步边缘提取，以获得像素级精度的初始边缘。为进一步增强边缘的视觉

显著性与结构连续性，随后应用形态学膨胀操作，使用 3 × 3 的全 1 卷积核进行单次迭代处理，有效增加边缘线宽以提高其在后续模型训练中的可识别性。所有生成的边缘图均保存为与原掩模同名的 PNG 格式文件，并统一存储于专用边缘标注目录，确保数据组织结构的规范性与一致性。该预处理方法在保留细节特征的同时强化了边缘结构，为模型学习显著目标边界提供了高质量的监督信息，如图 7 所示。

4.3. 评估方法

本文使用四种常见伪装目标识别领域标准的度量方式：S-measure、E-measure、MAE、F-measure。其中 MAE 用于评估预测图和真实图的绝对平均误差，计算公式如(2)所示：

$$\mathrm{MAE}=\frac{1}{H*W}\sum_{i=1}^{W}\sum_{j=1}^{H}\left|\mathrm{predict}(i,j)-\mathrm{GT}(i,j)\right| \tag{2}$$

E-measure [9]通过考虑全局像素级的统计信息和局部像素的匹配信息，来全面评估图像分割或目标检测算法的性能，计算公式如(3)所示：

$$\mathrm{E}_{\phi}=\frac{1}{W\times H}\sum_{i=1}^{W}\sum_{j=1}^{H}\phi_{\mathrm{FM}}(i,j) \tag{3}$$

S-measure [10]通过结合对象级别的全局结构相似性(基于 SSIM)和区域级别的局部结构一致性(基于区域增强对齐)来综合评价，计算公式如(4)所示：

$$S_{\alpha}=\alpha\times S_{O}+(1-\alpha)S_{r} \tag{4}$$

F-measure [11]通过引入一个基于空间权重的策略来改进传统的 F-measure：强调了对分类不确定性更高的边缘区域的评估重要性，计算公式如(5)所示：

$$F_{\beta}^{w}=\frac{(1+\beta^{2})\,Precision^{w}\times Recall^{w}}{\beta^{2}\times Precision^{w}+Recall^{w}} \tag{5}$$

4.4. 实验细节

本文使用 Pytorch 2.5.1、CUDA 14.1、vGPU-32 GB (32 GB)显卡进行训练与评估，所有图像经过图形标准化将其数据分布调整到以 0 为中心且尺度为 416 × 416 与训练模型 ConvNeXt 兼容。批量设置为 10，训练周期为 30 个 epcoh，并且采用学习率为 1e−4 的 ADAM 优化器。

4.5. 实验结果

4.5.1. 定量比较

为了更直观地评估不同算法在伪装目标检测任务中的性能差异，本文对所提方法及其他主流方法生成的预测结果进行了可视化对比。如图 8 所示，与传统的伪装目标检测方法相比，近年来提出的伪装目标检测模型在目标定位和识别能力方面普遍表现更优，能够较为准确地识别出伪装目标的潜在区域。本文所提出的 TGDNet 网络(图中红框标注部分)不仅在复杂背景中稳健地探测出伪装目标的隐藏区域，还能够清晰、完整地重构出目标的细节轮廓。

本文实验还在迷彩伪装人员数据集上做了可视化的对比，如图 9 可以更清楚地看到我们的模型在伪装目标区域识别与边缘性上也有着比基准模型更优秀的分割性能(矩形红色部分)，也表现了本模型有着伪装目标跨数据集的识别特性。

4.5.2. 定性比较

为了更加客观地评价本文提出模型的有效性，结合 4.3 小节评价指标，我们在 CAMO、COD10K、NC4K 数据集分别做了近年来伪装目标检测领域当中模型的横向对比，见表 1，可以看出本文提出的模型优于基准模型和对比模型。

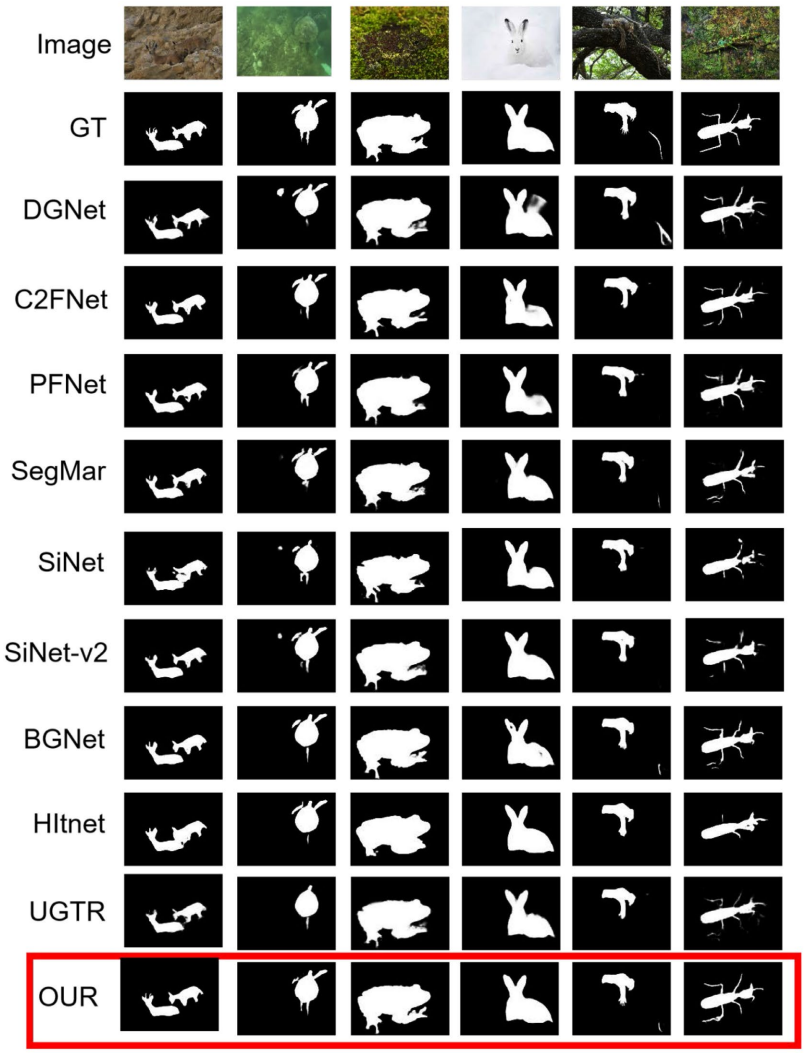

Figure 8. Qualitative comparative experiment
图 8. 定性对比实验

Figure 9. Cross-dataset effect
图 9. 跨数据集效果

Table 1. Comparative experiment
表 1. 对比实验

	CAMO				COD10K				NC4K			
eval	S↑	E↑	F↑	M↓	S↑	E↑	F↑	M↓	S↑	E↑	F↑	M↓
SiNet	0.745	0.804	0.644	0.092	0.776	0.864	0.631	0.043	0.808	0.871	0.723	0.058
PFNet	0.782	0.841	0.695	0.085	0.800	0.877	0.660	0.040	0.829	0.887	0.745	0.053
S-MGL	0.772	0.806	0.664	0.089	0.811	0.844	0.654	0.037	0.829	0.862	0.731	0.055
R-MGL	0.775	0.812	0.673	0.088	0.814	0.851	0.666	0.035	0.833	0.867	0.739	0.053
UGTR	0.784	0.821	0.683	0.086	0.817	0.852	0.665	0.036	0.839	0.874	0.746	0.052
LSR	0.787	0.838	0.696	0.080	0.804	0.880	0.673	0.037	0.840	0.895	0.766	0.048
C2FNet	0.796	0.854	0.719	0.080	0.813	0.890	0.686	0.036	0.838	0.897	0.762	0.049
JCSOD	0.800	0.859	0.728	0.073	0.809	0.884	0.684	0.035	0.841	0.898	0.771	0.047
BGNet	0.812	0.870	0.749	0.073	0.831	0.901	0.722	0.033	0.851	0.907	0.788	0.044
Ours	**0.863**	**0.916**	**0.825**	**0.049**	**0.860**	**0.921**	**0.784**	**0.023**	**0.881**	**0.930**	**0.840**	**0.031**

4.6. 消融实验

在本小节中，我们通过合理的设计消融实验来验证所提算法中各个模块的合理性和有效性。所有的消融实验均和前文的配置环境和参数保持不变，在 CAMO、COD10K、NC4K 这四个数据集上进行消融实验，如下表 2 所示，同时，依旧选用 MAE、F-measure 和 S-measure 作为评价指标，可以看到在添加 TCM 模块或添加 DGFM 模块指标均有提升，在 TCM + DGFM (TGDNet)指标达到最大值。

Table 2. Melting experiment
表 2. 消融实验

	CAMO				COD10K				NC4K			
eval	S↑	E↑	F↑	M↓	S↑	E↑	F↑	M↓	S↑	E↑	F↑	M↓
Base	0.820	0.830	0.712	0.07	0.793	0.799	0.614	0.039	0.833	0.839	0.715	0.055
+TCM	0.860	0.909	0.823	0.050	0.854	0.916	0.779	0.025	0.872	0.923	0.833	0.034
Ours	**0.863**	**0.916**	**0.825**	**0.049**	**0.860**	**0.921**	**0.784**	**0.023**	**0.881**	**0.930**	**0.840**	**0.031**

5. 结论

本文针对伪装目标检测任务中目标与背景相似度高、边缘细节易丢失等挑战，提出了一种基于边缘双注意力与上下文融合的改进模型 TGDNet。该模型以 ConvNeXt 为主干网络，充分利用其多尺度特征提取能力，并创新性地设计了双注意力引导融合模块(DGFM)与 Transformer 上下文增强模块(TCM)。DGFM 通过协同空间与通道注意力机制，实现了边缘信息与层级特征的动态自适应融合，有效强化了目标边缘区域的响应。TCM 则通过引入 Transformer 的自注意力机制，克服了 CNN 在建模长距离依赖关系上的局限性，显著提升了模型对全局语义上下文信息的感知与融合能力。

在四个公开基准数据集(CAMO, CHAMELEON, COD10K, NC4K)上的大量实验表明，本文所提出的 TGDNet 模型在多项关键评价指标(S-measure, E-measure, F-measure, MAE)上均优于现有的主流方法，展现了更优的分割精度、更清晰的边缘细节和更强的模型泛化能力。消融实验进一步验证了各个核心模块的有效性。本研究为复杂场景下的伪装目标精准检测提供了一种有效的解决方案，后续工作将探索模型

在移动端的轻量化部署及其在相关领域(如医学图像分割、缺陷检测)的应用潜力。

参考文献

[1] Fan, D., Ji, G., Sun, G., Cheng, M., Shen, J. and Shao, L. (2020) Camouflaged Object Detection. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Seattle, 13-19 June 2020, 2774-2784. https://doi.org/10.1109/cvpr42600.2020.00285

[2] Sun, Y., Chen, G., Zhou, T., Zhang, Y. and Liu, N. (2021) Context-Aware Cross-Level Fusion Network for Camouflaged Object Detection. *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence*, Montreal, 19-27 August 2021, 1025-1031. https://doi.org/10.24963/ijcai.2021/142

[3] Wang, K., Bi, H., Zhang, Y., Zhang, C., Liu, Z. and Zheng, S. (2022) D^2C-Net: A Dual-Branch, Dual-Guidance and Cross-Refine Network for Camouflaged Object Detection. *IEEE Transactions on Industrial Electronics*, **69**, 5364-5374. https://doi.org/10.1109/tie.2021.3078379

[4] Ren, J.J., Hu, X.W., Zhu, L., *et al.* (2021) Deep Texture-Aware Features for Camouflaged Object Detection. *IEEE Transactions on Circuits and Systems for Video Technology*, **33**, 1157-1167.

[5] Mei, H., Ji, G., Wei, Z., Yang, X., Wei, X. and Fan, D. (2021) Camouflaged Object Segmentation with Distraction Mining. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Nashville, 20-25 June 2021, 8768-8777. https://doi.org/10.1109/cvpr46437.2021.00866

[6] Fan, D., Ji, G., Cheng, M. and Shao, L. (2022) Concealed Object Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **44**, 6024-6042. https://doi.org/10.1109/tpami.2021.3085766

[7] Qin, X., Zhang, Z., Huang, C., Gao, C., Dehghan, M. and Jagersand, M. (2019) BASNet: Boundary-Aware Salient Object Detection. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Long Beach, 15-20 June 2019, 7471-7481. https://doi.org/10.1109/cvpr.2019.00766

[8] Fan, D., Ji, G., Zhou, T., Chen, G., Fu, H., Shen, J., *et al.* (2020) PraNet: Parallel Reverse Attention Network for Polyp Segmentation. In: Martel, A.L., *et al.*, Eds., *Medical Image Computing and Computer Assisted Intervention—MICCAI* 2020, Springer, 263-273. https://doi.org/10.1007/978-3-030-59725-2_26

[9] Fan, D., Gong, C., Cao, Y., Ren, B., Cheng, M. and Borji, A. (2018) Enhanced-Alignment Measure for Binary Foreground Map Evaluation. *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*, Stockholm, 13-19 July 2018, 698-704. https://doi.org/10.24963/ijcai.2018/97

[10] Fan, D., Cheng, M., Liu, Y., Li, T. and Borji, A. (2017) Structure-Measure: A New Way to Evaluate Foreground Maps. 2017 *IEEE International Conference on Computer Vision* (*ICCV*), Venice, 22-29 October 2017, 4558-4567. https://doi.org/10.1109/iccv.2017.487

[11] Margolin, R., Zelnik-Manor, L. and Tal, A. (2014) How to Evaluate Foreground Maps. 2014 *IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, 23-28 June 2014, 248-255. https://doi.org/10.1109/cvpr.2014.39