

# 融合 Transformer 和 CNN 的伪装目标检测

李沼洁<sup>1</sup>, 郭家毅<sup>1</sup>, 杨英东<sup>1</sup>, 毛国君<sup>2,3</sup>

(1. 厦门市公安局, 福建 厦门 361001;

2. 福建理工大学 计算机科学与数学学院, 福建 福州 350118;

3. 福建省大数据挖掘与应用技术重点实验室, 福建 福州 350118)

**摘要:** 针对复杂伪装场景下大多数方法难以有效定位、分割伪装目标等问题, 提出融合 Transformer 和卷积神经网络的伪装目标检测方法。首先, 为增强伪装目标特征表达能力, 提出了双分支编码-解码结构; 其次, 针对目标定位易出现偏差的问题, 设计了双分支边界感知模块; 为缓解伪装目标检测过程中出现误检和漏检等问题, 构建了一个交互式融合细化机制, 利用特征分组技术细化伪装目标的纹理, 提升复杂场景下模型检测精度。在3个公开数据集上进行了测试, 结果表明所提方法表现优异; 特别是, 相比于经典的 ZoomNet 算法, 所提方法在 CAMO 数据集上平均绝对误差降低了 4.76%, 结构相似性度量 S-measure 提高 1.59%。

**关键词:** 伪装目标检测; 双分支; 边界感知; 融合细化

中图分类号: TP391

文献标志码: A

文章编号: 2097-3853(2025)06-0557-07

## Camouflaged object detection by fusing Transformer and CNN

LI Zhaojie<sup>1</sup>, GUO Jiayi<sup>1</sup>, YANG Yingdong<sup>1</sup>, MAO Guojun<sup>2,3</sup>

(1. Xiamen Public Security Bureau, Xiamen 361001, China;

2. School of Computer Science and Mathematics, Fujian University of Technology, Fuzhou 350118, China;

3. Fujian Provincial Key Laboratory of Big Data Mining and Applications, Fuzhou 350118, China)

**Abstract:** Addressing the challenges of effectively locating and segmenting camouflaged objects in complex camouflage scenarios, where most methods falter, a Transformer and CNN fusion for camouflaged object detection is proposed. Firstly, to enhance the feature representation capability of camouflaged objects, a dual-branch encoder-decoder structure is introduced. Secondly, to address the problem of easy deviation in object localization, a dual-branch boundary perception module is designed; to alleviate the problems of false detection and missed detection in detecting camouflaged objects, an interactive fusion refinement mechanism is constructed. By using feature grouping techniques to refine the texture of camouflaged objects, the detection accuracy of models in complex scenes is improved. The model was tested on three challenging datasets, and the results show that the proposed method performs well. In particular, compared with the classic ZoomNet algorithm, the proposed method has reduced the MAE by 4.76% and the S-measure is increased by 1.59% on the CAMO dataset.

**Keywords:** camouflaged object detection; dual-branch; boundary perception; fusion refinement

近年来,随着深度学习技术在计算机视觉领域的不断创新和发展,伪装目标检测(camouflaged object detection, COD)技术逐渐成为研究的热点,

并广泛应用于医学外科领域<sup>[1-2]</sup>,自然灾害救援<sup>[3]</sup>以及艺术处理<sup>[4]</sup>等相关领域。

研究人员致力于提升伪装目标检测的效率,

收稿日期:2025-01-14

基金项目:国家自然科学基金资助项目(61773415)

第一作者简介:李沼洁(1998—),男,河南柘城人,硕士,研究方向:目标检测、语义分割等。

例如 Wang 等人<sup>[5]</sup>提出了 D2CNet 模型,采用密集串联的方式聚合多级特征,并引入注意力机制增强特征和细化局部信息。Ren 等人<sup>[6]</sup>设计多个纹理感知模块学习伪装目标和背景之间的差异,并计算每个位置的特征响应协方差矩阵提取纹理信息。李等人<sup>[7]</sup>提出一种基于 Transformer 的跨尺度交互学习伪装目标检测方法,通过高低层特征融合并引入全局上下文信息增强特征表达能力。Yang 等人<sup>[8]</sup>提出将卷积神经网络和 Transformer 结合,用概率表示伪装目标的不确定性,以关注不确定性区域。上述伪装目标检测方法虽然取得了一定的成果,但在复杂的伪装场景中难以准确识别出目标和背景噪声的纹理差异性。

生物学上认为,灵长类动物的眼睛不仅会出现视野重叠,还会出现双眼视差<sup>[9]</sup>,视觉中枢系统可以从双眼视差信号中提取产生立体视觉的信息,使观察者更加准确地判断物体的深度距离,进而在复杂环境中消除伪装。受此启发,本研究提出了一种融合 Transformer 和 CNN 的伪装目标检测方法 (fusion transformer and CNN for camouflaged object detection, FTCNet),设计了一种结合 Transformer

和卷积神经网络的双分支编码-解码网络。为了有效区分前景与背景,本研究构建了双分支边界感知模块 (dual-branch edge aware module, DEAM),并通过设计一种交互式融合细化机制 (Interactive fusion refinement mechanism, IFRM),进一步增强网络的特征表达能力。

## 1 FTCNet

FTCNet 的整体架构如图 1 所示。Transformer 编码器一共包含 12 层,可以对图像中的长短期依赖进行有效建模。通过对 Transformer 分支最后一层的输出使用层归一化,得到包含丰富全局语义信息的特征图  $x_t$ 。在 CNN 分支的编码阶段中,通过引入  $x_t$  来增强网络的语义信息表达能力。在解码阶段中,设计了一个双分支边界感知模块 DEAM,以生成高质量边界图,帮助模型定位和分割伪装目标。此外,本研究还设计了一个交互融合细化机制 IFRM,通过多尺度特征信息之间的交互融合,逐渐细化目标的纹理结构,进一步提升伪装目标检测效果。

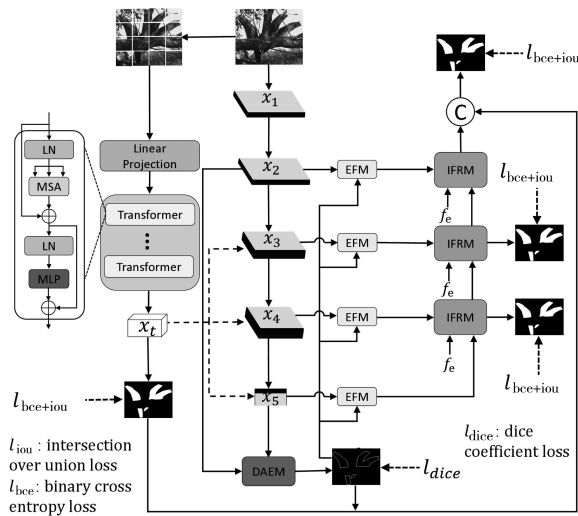


图 1 FTCNet 整体架构

Fig.1 Overall network architecture of FTCNet

### 1.1 双分支编码器-解码器

Transformer 擅长利用自注意力机制捕捉长距离依赖关系,能够有效获取全局语义信息。因此,本研究采用 Transformer 架构对图像进行编码,首先对输入图像  $X \in R^{C \times H \times W}$  ( $C$ ,  $H$  和  $W$  分别为图像的通道数,高度和宽度) 进行处理,将其均匀划分

为  $N=H/S \times W/S$  块,  $S$  设为 16。然后将  $N$  个图像块展平并馈送到输出维度为  $D$  的线性投影层中,得到  $Z=R^{D \times N}$  的原始序列。为了更好地利用空间先验知识,将可学习的位置信息嵌入序列  $Z$  中后作为 Transformer 编码器的输入。再将输入序列通过不同的线性映射生成查询矩阵  $Q$ , 键矩阵  $K$  和

值矩阵  $V$ 。在  $Q$  和  $K$  之间使用矩阵乘法计算相似度,并通过 SoftMax 层进行归一化得到一组权重矩阵。最后,将权重与对应的 Value 进行加权求和。计算过程定义为:

$$\begin{cases} Q = F \times W_Q \\ K = F \times W_K \\ V = F \times W_V \end{cases} \quad (1)$$

$$\text{Attention}(Q, K, V) = \text{SoftMax}\left(\frac{QK^T}{\sqrt{d}}\right)V \quad (2)$$

其中,  $F$  表示的是输入特征,  $W_Q, W_K, W_V$  是具有可训练参数的权重,  $d$  为  $Q, K, V$  的维数。

对于最后一个编码器的输出,经过层归一化后再利用上采样操作将其恢复成分辨率为  $C \times H/32 \times W/32$  的特征图,以便与 CNN 分支提取到的特征图融合。

此外,对于另一个分支,本研究采用 Res2Net-50 作为主干网络,对输入图像  $X \in R^{C \times H \times W}$  ( $C, H$  和  $W$  分别为图像的通道数,高度和宽度)进行特征编码,得到 5 个不同尺度的特征图,分别表示为  $x_i, i = \{1, 2, \dots, 5\}$ 。将 Transformer 分支得到的特征图  $x_f$  引入到特征图  $x_3, x_4$  和  $x_5$  中,最后再通过设计的 DEAM 模块和 IFRM 机制进行辅助解码,不仅充分利用双分支编码器-解码器结构的互补性,还有效保留了丰富的细节信息和全局语义信息。

## 1.2 双分支边界感知

根据伪装特性得知,目标边界与背景的区别度较低会导致获取到的边界误差变大等问题。为此,本研究构建了双分支边界感知模块 DEAM,以更好地学习伪装目标的边界信息。双分支边界感知模块 DEAM 的结构如图 2 所示。

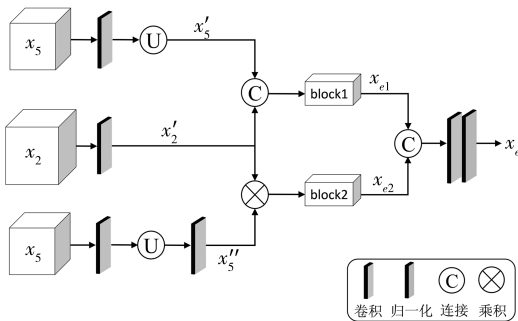


图 2 双分支边界感知模块

Fig.2 Dual-branch boundary perception module

在边界感知过程中,为了平衡噪声抑制和边缘细节保留之间的关系,选用边界细节信息充足的特征图  $x_2$  和语义信息充足的特征图  $x_5$ ,以更好地感知目标边界。对于第一个分支,首先使用核大小为  $1 \times 1$  的卷积分别降低特征图  $x_2$  和特征图  $x_5$  的通道数。然后将  $x_5$  一样大小的  $x'_5$ ,如式(3)所示。再将其沿通道维度拼接起来,从而保证通道维度上的边界信息充足,最后通过一个卷积块得到该分支的输出  $x_{e1}$ ,计算过程定义为式(4)。

$$x'_5 = \text{Up}(\text{Conv}(x_5)) \quad (3)$$

$$x_{e1} = \text{block1}(\text{cat}(x'_5, \text{Conv}(x_2))) \quad (4)$$

其中,Up 是上采样操作,Conv 是卷积操作,block1 表示的是一个卷积块,cat 表示的是沿通道维度拼接。

对于另一个分支,同上述方式将特征图  $x_5$  和  $x_2$  的通道数统一,如式(5)所示。由于特征图  $x_5$  已经上采样至  $x_2$  同样大小,所以本研究在两个特征图之间使用矩阵乘法来保证空间上的边界信息充足。然后通过一个卷积块得到该分支的输出  $x_{e2}$ 。计算过程定义如式(6)所示。

$$x''_5 = \text{Conv}(\text{Up}(\text{Conv}(x_5))) \quad (5)$$

$$x_{e2} = \text{block2}(x''_5 \times \text{Conv}(x_2)) \quad (6)$$

其中,block2 表示的是一个卷积块。

最后,将两个分支的输出沿通道维度拼接,即可得到增强后的目标边界信息  $x_e$ 。

$$x_e = \text{cat}(x_{e1}, x_{e2}) \quad (7)$$

## 1.3 交互式融合细化

复杂场景中伪装目标的纹理大多与背景高度相似,严重影响模型检测精度,导致出现漏检和误判的问题。为此,本研究构建了一个交互融合细化机制 IFRM,如图 3 所示。

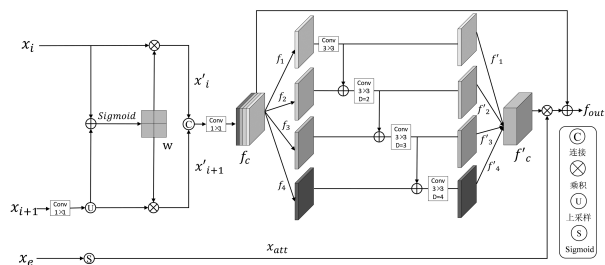


图 3 交互式融合细化

Fig.3 Interactive fusion refinement mechanism

以特征图  $x_4$  和  $x_5$  为例,首先将输入特征图

送到 EFM<sup>[10]</sup> 中,通过该组件中的注意力机制来关注特征图不同通道上最有用的信息;同时将来自双分支边界感知模块 DEAM 的边界语义信息注入到 EFM 组件中,以进一步增强对目标区域整体结构的特征表达能力。为了更好地利用边界信息区分伪装前景与背景,将增强后的特征  $x_4$ 、 $x_5$  以及边界信息  $x_e$  馈送到 IFRM 中。由于特征图的分辨率不同,首先通过双线性插值的方式将深层特征  $x_5$  上采样到和  $x_4$  同样大小。通过元素加法将其融合,并利用 Sigmoid 函数进行归一化操作得到权重  $w$ ,分别对  $x_4$ 、 $x_5$  加权。该过程可以定义为式(8)。

$$\begin{cases} x'_i = x_i \times \sigma(x_i + \text{Up}(x_{i+1})) \\ x'_{i+1} = x_{i+1} \times \sigma(x_i + \text{Up}(x_{i+1})) \end{cases} \quad (8)$$

其中,  $\sigma$  指的是 Sigmoid 函数。

为重组各通道之间的特征信息,使用通道分组技术将加权并拼接后的特征图按照通道维度平均分为 4 组分辨率相同的特征  $f_i$ ,  $i=1,2,3,4$ , 如式(9)所示。为了细化目标的纹理结构,对分组特征分别使用标准卷积和不同倍率的空洞卷积进行交互融合。第一组特征经过核大小为  $3 \times 3$  的标准卷积之后得到  $f'_1$ ,在  $f'_1$  和特征  $f_2$  之间使用元素加法;再对相加后的结果使用倍率为 2 的空洞卷积得到特征  $f'_2$ ,用同样的方法对其余特征两两相加,并分别使用倍率为 3 和 4 的空洞卷积得到特征  $f'_3$  和  $f'_4$ ,如式(10)所示。再将交互融合后的特征沿着通道维度拼接起来得到  $f'_c$ ,如式(11)所示。

$$\begin{aligned} f_c &= \text{conv}(\text{cat}(x'_i, x'_{i+1})) \\ f_i &= \text{split}(f_c), i=1,2,3,4 \end{aligned} \quad (9)$$

$$f'_i = \begin{cases} f_{\text{conv}^3}(f_i), i=1 \\ f_{\text{conv}^2}(f_i) + f'_1, i=2 \\ f_{\text{conv}^3}(f_i) + f'_2, i=3 \\ f_{\text{conv}^4}(f_i) + f'_3, i=4 \end{cases} \quad (10)$$

$$f'_c = \text{cat}(f'_1, f'_2, f'_3, f'_4) \quad (11)$$

其中, split 表示按特征通道数分组操作。

最后,通过残差连接得到最终输出  $f_{\text{out}}$ ,该过程可以用式(12)定义。

$$f_{\text{out}} = f_c + f'_c \times \sigma(x_e) \quad (12)$$

## 1.4 损失函数

该模型的损失函数主要包括对伪装目标监督

和对目标边界监督两种。对于伪装目标监督,采用二值交叉熵(BCE)损失评估预测结果和真实值之间的差异性是当前的主流做法。但这种监督方法过于关注像素之间的差异性,没有充分考虑每个像素之间的相关性。因此,本研究结合加权的二值交叉熵损失和加权的交并比损失进行监督,如式(13)所示。

$$L = L_{\text{BCE}}^w(G_m, P_m) + L_{\text{IOU}}^w(G_m, P_m) \quad (13)$$

其中,  $G_m$  表示的是真值图,  $P_m$  表示的是预测图,  $L$  表示损失函数,  $L_{\text{BCE}}^w$  和  $L_{\text{IOU}}^w$  分别表示加权的 BCE 损失函数和加权的 IOU 损失函数。

对于目标边界,本研究采用 dice<sup>[10]</sup> 损失来充分平衡正负样本,以预测出更精细的边界图。此外,还对中间结果和 Transformer 分支的输出实施了深度监督,最终总的损失  $L_{\text{total}}$  的计算如式(14)所示。

$$L_{\text{total}} = \sum_{i=1}^4 (L_{\text{BCE}}^w(G_m, P_m) + L_{\text{IOU}}^w(G_m, P_m) + \lambda L_{\text{dice}}(G_e, P_e)) \quad (14)$$

其中,  $G_e$  表示的是边界真值图,  $P_e$  表示的是预测生成的边界图,  $L_{\text{dice}}$  表示 dice 损失函数,  $\lambda$  是一个权重参数,本研究在实验中将其设置为  $\lambda=3$ 。

## 2 实验结果与讨论

### 2.1 数据集

本研究使用 CAMO<sup>[11]</sup> 数据集和 COD10K<sup>[12]</sup> 数据集中共 4 040 张用于训练的图像对提出的模型进行训练。此外,将 FTCNet 算法在 CAMO<sup>[11]</sup>, CHAMELEON<sup>[12]</sup> 和 NC4K<sup>[13]</sup> 3 个公开的数据集上进行测试。CAMO 数据集共包含 1 250 张图像,其中 250 张用于测试,涵盖了自然伪装和人工伪装的图像;CHAMELEON 数据集虽然只包含 76 张图像,但图像中的目标与背景对比度极低,且存在拟态伪装等具有挑战性的图像,有助于验证模型的泛化能力;NC4K 数据集共包含 4 121 张用于测试的图像,为伪装目标检测算法提供了丰富的测试样本。

### 2.2 评价指标

为准确评估 FTCNet 的性能,本研究选用平均绝对误差(MAE)、结构相似性度量 S-measure、F-measure 和 E-measure 等 4 个主流评价指标进行评估,公式为:

$$\left\{ \begin{array}{l} \text{MAE} = \frac{1}{W \times H} \sum_{x=1}^W \sum_{y=1}^H |P(x,y) - G(x,y)| \\ S_{\alpha} = \alpha \times S_o + (1 - \alpha) \times S_r \\ F_{\beta} = \frac{(1 + \beta^2) \text{Precision} \times \text{Recall}}{\beta^2 \text{Precision} + \text{Recall}} \\ E_m = \frac{1}{W \times H} \sum_{i=1}^W \sum_{j=1}^H \emptyset(i, j) \end{array} \right. \quad (15)$$

其中,  $P(x,y)$  代表预测的伪装图,  $G(x,y)$  代表真值图;  $S_o$  用来计算目标感知, 而  $S_r$  用来计算区域

的特征,  $\alpha$  表示的是在区间  $[0, 1]$  的权值; 根据实验分别将  $\alpha$  设置为 0.5,  $\beta^2$  设置为 0.3;  $\emptyset(\cdot)$  表示对角矩阵。

### 2.3 实验结果分析

为了验证所提方法的有效性, 将 FTCNet 与 7 种最先进的算法在 3 个公开数据集上进行对比。在保持参数一致的前提下使用相同的训练数据集和测试数据集对模型进行训练和测试, 并在同等软硬件环境下实施, 量化比较结果如表 1 所示。

表 1 FTCNet 与 7 种伪装目标检测算法对比

Tab.1 Comparison of FTCNet with seven camouflaged object detection algorithms

方法	CAMO				CHAMELEON				NC4K			
	$S_{\alpha}$	$F_{\beta}$	MAE	$E_m$	$S_{\alpha}$	$F_{\beta}$	MAE	$E_m$	$S_{\alpha}$	$F_{\beta}$	MAE	$E_m$
C2FNet <sup>[14]</sup>	0.796	0.719	0.080	0.854	0.888	0.828	0.032	0.935	0.838	0.762	0.049	0.897
SINet <sup>[12]</sup>	0.751	0.606	0.100	0.771	0.869	0.740	0.044	0.891	0.808	0.723	0.058	0.871
FAPNet <sup>[15]</sup>	0.815	0.734	0.076	0.865	0.893	0.825	0.028	0.940	0.851	0.775	0.047	0.899
OCENet <sup>[16]</sup>	0.802	0.723	0.080	0.852	0.897	0.833	0.027	0.940	0.853	0.785	0.045	0.902
PreyNet <sup>[17]</sup>	0.790	0.708	0.077	0.842	0.895	0.844	0.028	0.952	—	—	—	—
CamoFocus <sup>[18]</sup>	0.812	0.752	0.071	0.873	0.898	0.849	0.027	0.953	0.847	0.788	0.043	0.909
ZoomNet <sup>[19]</sup>	0.820	0.752	0.066	0.877	0.902	0.845	0.023	0.943	0.853	0.784	0.043	0.896
FTCNet	0.833	0.762	0.063	0.886	0.904	0.850	0.025	0.941	0.862	0.790	0.042	0.909

根据表 1 中的数据可知, 在涵盖自然伪装和人为伪装图像的 CAMO 数据集上, 与 ZoomNet、PreyNet 和 OCENet 相比, 结构相似性度量  $S_{\alpha}$  分别提升了 1.59%、5.44% 和 3.87%, 说明所提方法有效缓解了漏检等问题; 此外, 平均绝对误差 MAE 分别降低了 4.76%、22.22% 和 26.98%, 进一步表明 FTCNet 可以更准确地定位并分割出图像中的前景和背景。在包含较多拟态伪装的 NC4K 数据集上, 与 ZoomNet、OCENet 和 FAPNet 相比,  $E_m$  分别提高了 1.45%、0.78% 和 1.11%, 也表明 FTCNet 可以准确检测出复杂场景下的伪装目标。

为了进一步证明 FTCNet 的性能, 进行了一系列更详细的视觉对比实验, 相应的可视化对比如图 4 所示。从图 4 可见, FTCNet 模型预测的伪装

图结构完整、清晰, 且边界更加平滑。

如图 4 第 1 行所示, FTCNet 可以更准确地检测出被水生植物覆盖的鳄鱼, 而其他方法都在不同程度上存在漏检的问题。从第 2、3 行可见, 对于人工制造的透明目标, 大多方法难以从复杂的背景中区分出前景目标, 而 FTCNet 可以捕捉更丰富的语义信息, 并能够有效整合细节信息, 获得相对完整的预测图。第 4、5、6 行中均存在动物拟态隐藏的图像, 其他方法预测出的结果存在漏检、边界不清晰甚至是误判等问题, 而 FTCNet 通过模拟双目视差效应, 准确定位伪装目标位置。此外, 通过学习边界信息并将其引入到交互融合细化机制中, FTCNet 可以得到结构完整清晰、边界平滑的预测结果。综上, FTCNet 可以有效应对各种复杂伪装场景的挑战。

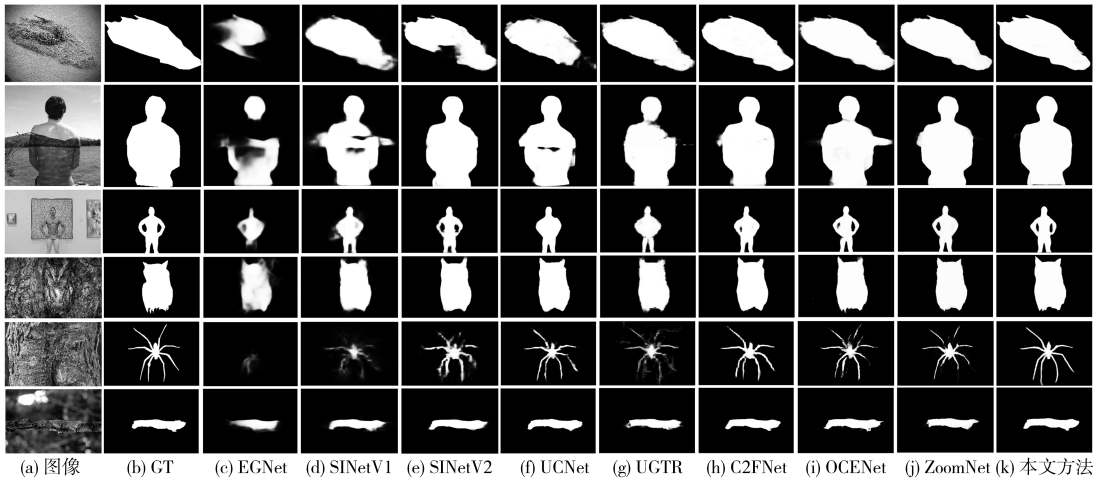


图 4 可视化对比结果展示

Fig.4 Visualization of comparison results

## 2.4 消融实验

为了研究 FTCNet 中不同关键组件的有效性,分别在 CAMO 数据集和 NC4K 数据集上设计了消融实验。对于基线模型 B1,实验中仅使用 Transformer 进行复杂场景中伪装目标的检测。对于基线模型 B2,选择删除所提出的 DEAM 和 IFRM,将 Transformer 分支的输出和 CNN 分支的输出简单聚合。在 CAMO 数据集和 NC4K 数据集上的消融实验研究结果如表 2 所示。

从表 2 可知,与基线模型 B1 相比,基线模型 B2 在两个数据集上的量化指标均不占优势,特别是,在 CAMO 数据集上  $F_{\beta}$  降低了 2.16%,这是因

为简单的特征聚合可能引入噪声。与基线模型 B2 相比,在添加 DEAM 模块后,CAMO 数据集上的  $F_{\beta}$  提升了 2.75%,表明提出的 DEAM 模块可以有效帮助模型在复杂场景中定位和分割伪装目标。此外,为验证 IFRM 的有效性,删除其余组件和边缘辅助信息,仅使用交互融合细化机制。从表 2 中第 4 行可以看出,在更加复杂的 NC4K 数据集上,加入 IFRM 后,与基线模型 B2 相比  $S_{\alpha}$  提高了 1.30%,说明多尺度特征信息的交互融合细化可以帮助模型更准确地识别前景与背景纹理差异性。当同时引入 DEAM 和 IFRM 时,模型性能获得进一步提升,在各项度量指标上均达到最优。

表 2 在 CAMO 数据集和 NC4K 数据集上的消融实验研究

Tab.2 Experimental study of ablation on CAMO dataset and NC4K dataset

M	Method	CAMO				NC4K			
		$S_{\alpha}$	$F_{\beta}$	MAE	$E_m$	$S_{\alpha}$	$F_{\beta}$	MAE	$E_m$
1	B1	0.807	0.742	0.066	0.875	0.844	0.779	0.046	0.907
2	B2	0.801	0.726	0.079	0.858	0.846	0.773	0.051	0.901
3	B2+DEAM	0.810	0.746	0.072	0.866	0.852	0.784	0.046	0.907
4	B2+IFRM	0.803	0.737	0.076	0.869	0.857	0.778	0.044	0.905
5	B2+DEAM+ IFRM	0.833	0.762	0.063	0.886	0.862	0.790	0.042	0.909

## 3 结束语

本研究提出了一种融合 Transformer 和 CNN 的伪装目标检测方法 FTCNet,通过构建双分支编码器-解码器结构模拟视差效果,充分提取图像

中的全局信息和局部信息,并采用自顶向下的方式融合多层次特征以完善伪装目标的结构。虽然该方法显著提高了模型检测精度,但其参数量较大,今后将主要研究更加轻量化的伪装目标检测算法,进一步提高模型效率。

## 参考文献:

- [1] HE J Y, WU X, JIANG Y G, et al. Hookworm detection in wireless capsule endoscopy images with deep learning[J]. IEEE Transactions on Image Processing, 2018, 27(5): 2379–2392.
- [2] FAN D P, ZHOU T, JI G P, et al. Inf-net: automatic COVID-19 lung infection segmentation from CT images[J]. IEEE Transactions on Medical Imaging, 2020, 39(8): 2626–2637.
- [3] MISHRA B, GARG D, NARANG P, et al. Drone-surveillance for search and rescue in natural disaster[J]. Computer Communications, 2020, 156: 1–10.
- [4] CHU H K, HSU W H, MITRA N J, et al. Camouflage images[J]. ACM Transactions on Graphics, 2010, 29(4): 1–8.
- [5] WANG K, BI H, ZHANG Y, et al. D2C-Net: a dual-branch, dual-guidance and cross-refine network for camouflaged object detection[J]. IEEE Transactions on Industrial Electronics, 2021, 69(5): 5364–5374.
- [6] REN J, HU X, ZHU L, et al. Deep texture-aware features for camouflaged object detection[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 33(3): 1157–1167.
- [7] 李建东, 王岩, 曲海成. 基于 Transformer 的跨尺度交互学习伪装目标检测[J]. 计算机系统应用, 2024, 33(2): 115–124.
- [8] YANG F, ZHAI Q, LI X, et al. Uncertainty-guided transformer reasoning for camouflaged object detection[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 4146–4155.
- [9] BARLOW H B, BLAKEMORE C, PETTIGREW J D. The neural mechanism of binocular depth discrimination[J]. The Journal of Physiology, 1967, 193(2): 327–342.
- [10] SUN Y J, WANG S, CHEN C, et al. Boundary-guided camouflaged object detection[EB/OL] (2022-07-02) [2024-10-07]. <https://arxiv.org/abs/2207.00794v1>.
- [11] LE T N, NGUYEN T V, NIE Z L, et al. Anabranch network for camouflaged object segmentation[J]. Computer Vision and Image Understanding, 2019, 184: 45–56.
- [12] FAN D P, JI G P, SUN G, et al. Camouflaged object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 2777–2787.
- [13] LYU 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 (CVPR). Nashville, TN, USA: IEEE, 2021: 11586–11596.
- [14] SUN Y J, CHEN G, ZHOU T, et al. Context-aware cross-level fusion network for camouflaged object detection[EB/OL] (2021-05-26) [2024-10-07]. <https://arxiv.org/abs/2105.12555v1>.
- [15] ZHOU T, ZHOU Y, GONG C, et al. Feature aggregation and propagation network for camouflaged object detection[J]. IEEE Trans Image Process, 2022, 31: 7036–7047.
- [16] LIU J W, ZHANG J, BARNES N. Modeling aleatoric uncertainty for camouflaged object detection[C]//2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). Waikoloa, HI, USA: IEEE, 2022: 2613–2622.
- [17] ZHANG M, XU S, PIAO Y R, et al. PreyNet: preying on camouflaged objects[C]//Proceedings of the 30th ACM International Conference on Multimedia. Lisboa, Portugal: ACM, 2022: 5323–5332.
- [18] KHAN A, KHAN M, GUEAIEB W, et al. CamoFocus: enhancing camouflage object detection with split-feature focal modulation and context refinement[C]//2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). Waikoloa, HI, USA: IEEE, 2024: 1423–1432.
- [19] 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 (CVPR). New Orleans, LA, USA: IEEE, 2022: 2150–2160.

(责任编辑: 方素华)