文章编号：1007-2780(2026)08-1187-13

基于特征动态平衡与增强的通用分割算法

邓万宇，刘洋伍*
（西安邮电大学 计算机学院，陕西 西安 710121）

摘要：显著目标检测与伪装目标检测虽然视觉特征对立，但均对边界分割精度提出了极高要求。然而，现有方法在跨任务泛化时常因特征偏好单一而导致性能瓶颈。为此，本文提出了一种通用分割模型SAM2-UFFM。该模型以SAM2的Hiera骨干网络为编码器，通过冻结预训练参数并引入轻量级适配器，实现了参数高效微调，完整保留了强大的特征表达能力；摒弃了传统U-Net的架构，设计了AttUNet解码器，通过全尺度跳跃连接实现全尺度信息的高效融合，有效解决了信息流失问题；提出了特征平衡注意力模块，并结合特征互补精修模块，动态平衡全局语义与局部细节，随后针对性强化显著目标的全局感知与伪装目标的边界精细分割，同时辅以深度监督策略防止过分割。实验结果表明，在3个公开数据集上，该模型的综合表现优于现有先进方法。其中在COD和SOD数据集上的结构性度量相较于现有最佳模型分别提升1.02%和1.83%，平均绝对误差降低至0.035和0.018。此外，在钢材表面缺陷分割任务中IoU提升2.77%，有效验证了模型的跨领域泛化能力。

关　键　词：图像分割；显著目标；伪装目标；SAM2；注意力机制
中图分类号：TP391.4　　**文献标识码：**A　　**doi：**10.3724/CJLCD.2026-0115　　**CSTR：**32172.14.CJLCD.2026-0115

Universal segmentation algorithm based on dynamic feature balancing and enhancement

DENG Wanyu, LIU Yangwu*

(*School of Computer Science and Technology, Xi'an University of Posts and Telecommunications, Xi'an* 710121, *China*)

Abstract: Although salient object detection and camouflaged object detection exhibit opposite visual characteristics, both tasks impose stringent requirements on boundary segmentation accuracy. However, existing methods often suffer from limited cross-task generalization due to their reliance on task-specific feature preferences. To address this issue, we propose a universal segmentation model, SAM2-UFFM. The proposed framework adopts the Hiera backbone of SAM2 as the encoder, where the pretrained parameters are frozen and lightweight adapters are incorporated for parameter-efficient fine-tuning, thereby fully preserving the powerful feature representation capability of the foundation model. Furthermore, departing from the conventional U-Net architecture, we design an AttUNet decoder that employs full-scale

收稿日期：2026-06-27；**修订日期：**2026-07-18.
基金项目：陕西省科学家工程师创新团队项目(No. 2023KXJ-091)；秦创原总窗口“四链融合”计划重点项目(No. 2025PT-ZCK-67)
Supported by Shaanxi Scientist & Engineer Innovation Team (No. 2023KXJ-091); Key Project of the Qinchuangyuan General Window “Four-Chain Integration” Program (No. 2025PT-ZCK-67)
*通信联系人，E-mail：liuyangwu2023@163.com

skip connections to facilitate efficient multi-scale feature fusion and effectively alleviate information loss during feature propagation. In addition, a Feature Balance Attention module is introduced and integrated with a Feature Complementary Refinement module to dynamically balance global semantic information and local structural details. This design simultaneously enhances global perception for salient objects and promotes precise boundary delineation for camouflaged objects. Meanwhile, a deep supervision strategy is adopted to suppress over-segmentation and improve optimization stability. Experimental results on three public benchmark datasets demonstrate that the proposed method consistently outperforms state-of-the-art approaches. Specifically, on the COD and SOD benchmarks, the structural measure is improved by 1.02% and 1.83%, respectively, compared with the best existing methods, while the mean absolute error (MAE) is reduced to 0.035 and 0.018. Furthermore, on the steel surface defect segmentation task, the proposed model achieves a 2.77% improvement in IoU, further validating its superior cross-domain generalization capability.

Key words: image segmentation; salient object; camouflaged object; SAM2; attention mechanism

1 引 言

伪装目标检测(Camouflaged Object Detection, COD)与显著目标检测(Salient Object Detection, SOD)是图像分割领域中具有不同特征需求的两类典型任务。SOD旨在快速定位全局最突出的前景区域,强调模型的全局上下文感知能力;COD则需精准挖掘与背景高度融合的隐蔽目标。由于两类任务在全局语义与局部细节方面的关注重点存在差异,针对单一任务设计的模型通常难以在另一任务中保持良好的性能。因此,如何构建能够兼顾不同任务特征需求的通用分割模型,已成为计算机视觉领域的重要研究问题。

近年来,深度学习方法在各自领域均取得了显著进展。在COD方向,Fan等人提出了SINet[1]模型,该模型包含两个模块:搜索模块定位伪装目标,识别模块精确检测伪装目标,同时建立了首个大规模伪装目标数据集。Mei等人提出PFNet[2]模型,通过模拟捕食过程提高了伪装物体分割的精度。Liu等人提出OCENet[3]模型,通过干扰挖掘策略实现伪装目标的精确分割。Pang等人提出ZoomNet[4]模型,采用混合尺度三元组网络实现远近尺度协同检测,显著提高了多尺度伪装目标的分割质量。He等人提出FEDER[5]模型,通过特征分解与边缘重构模块在边界精度上取得了当时最佳表现。李明岩等人[6]设计边缘提取模块与跨层级特征聚合模块,利用边缘先验引导多级特征融合,进一步提升了伪装目标的边界分割精度。在SOD方向,主流模型更侧重于全局语义的建模与整合,以快速锁定整体显著区域。Qin等人提出的U^{2}Net[7]通过深度嵌套U结构实现了高效显著特征提取。Wu等人提出的EDN[8]通过边缘引导与细节增强模块在边界质量上进一步优化。Gao等人提出的MDSAM[9]通过多尺度细节增强机制结合SAM[10]显著提升了复杂场景表现。Yuan等人提出的UGRAN[11]通过不确定性引导的Transformer推理刷新了多项指标。近期,Xiong等人提出的SAM2-UNet[12]通过将SAM2[13]作为编码器结合经典U-Net解码器,并引入适配器进行微调,在COD和SOD上同时成为了当时的SOTA方法。

特征融合与注意力机制作为提升分割精度的核心技术,在其他细分领域也得到了广泛验证。邬开俊等人[14]面向城市道路语义分割,设计全局感知模块与多尺度特征融合模块,有效缓解了边界模糊与多尺度目标分割精度不足的问题。赵林等人[15]针对齿轮表面细微缺陷分割场景,提出多尺度特征增强模块与分块注意力机制,显著提升了模型对微小缺陷与局部纹理差异的感知能力。刘玉等人[16]在小样本语义分割任务中构建特征-掩码双重聚合模块,通过挖掘全局特征相似性有效减小了原型偏差,提升了新类目标的分割效果。上述跨领域的技术探索,为构建兼顾全局语义与局部细节的通用分割模型提供了重要的思路借鉴。

现有方法在各自领域任务中取得了较好的性能,但在跨任务泛化方面仍存在明显局限。SOD

模型通常侧重全局语义建模与显著性抑制，过度依赖全局线索可能导致其忽略COD任务中的局部纹理差异，从而产生漏检和边界模糊等问题。相反，COD模型更加关注局部纹理差异与边界连续性，虽然能够精细刻画隐蔽目标的轮廓，但过度强调局部细节可能损害SOD任务中的目标整体一致性，导致内部空洞和碎片化过分割现象。即便是集成了SAM2[13]强大编码器的SAM2-UNet[12]，其相对简洁的U型解码结构在多尺度特征融合与动态特征调节方面仍存在一定局限性，难以同时兼顾两类任务。

综上所述，现有深度学习方法在各自的专业领域内表现出色，但面对具有差异化特征需求的任务时，便会暴露出明显的跨任务局限性，其本质在于缺少一种可在全局语义与局部细节间自适应调节的特征平衡机制。针对这一问题，本文提出一个分割模型SAM2-UFFM。该模型以SAM2的Hiera[17]骨干网络作为编码器，通过冻结预训练参数并引入轻量级适配器实现参数高效微调，在保留强大特征表达能力的同时降低训练成本，同时为了增强模型的多尺度表征能力，引入了RFB多尺度特征提取模块，利用不同扩张率的并行卷积分支捕获多尺度上下文信息，以适应不同尺度目标的分割需求。解码器设计了类UNet3+[18]的AttUNet架构，通过全尺度跳跃连接实现多层级特征的充分融合，减少深层网络传递过程中的空间信息损失。针对COD与SOD在特征需求上的差异，本文构建了一套特征调节机制：首先，通过特征平衡注意力模块（Feature BalancingAttention Module，FBAM）动态调整全局语义与局部细节的特征权重；随后，通过专为对立任务设计的特征互补精修模块，实现深层语义与浅层细节的针对性增强。此外，采用深度监督策略促进多层级特征学习与梯度传播，提高模型训练的稳定性。

本文模型在COD10K[1]和DUTS[19]数据集上的实验结果表明，SAM2-UFFM在COD和SOD任务中均取得了较好的性能。为进一步验证模型的跨领域泛化能力，本文将模型扩展应用于钢材表面缺陷数据集NEU-SEG[20]中，以测试其在复杂工业环境下的性能。

2　模型的构成及工作原理

2.1　整体网络架构

模型的整体网络架构如图1所示。给定输入图像$I\in R^{H\times W\times 3}$，其中$H$和$W$分别表示高度和宽度，Hiera输出4个级别的特征$X_i\in R^{C_i\times\frac{H}{2^{i+1}}\times\frac{W}{2^{i+1}}}$（$i=1,2,3,4$），通道维度为$C_i\in\{144,288,576,1\,152\}$。为实现参数高效微调，模型沿用SAM2-UNet[12]的

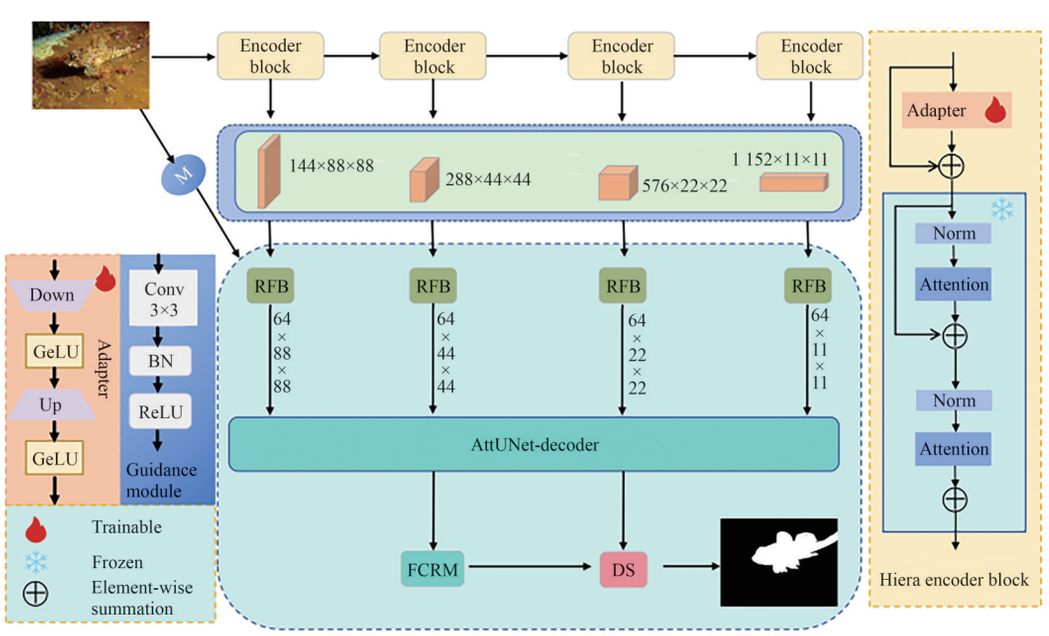


图1　SAM2-UFFM模型的整体网络架构

Fig. 1　Overall network architecture of the SAM2-UFFM model

瓶颈适配器设计，在 Hiera[17]的每个多尺度块之前进行插入。在提取编码器特征之后，采用 4 个 RFB 模块对特征进行处理，将通道数量减少到 64，并增强这些轻量级特征。在解码器部分，设计了 AttUNet 解码器，通过全尺度跳跃连接机制，将来自编码器所有层级的特征与解码器的每一个聚合节点进行融合，有效解决了传统 UNet[21]在特征传递中出现的信息流失问题。在 4 个解码层级中构建了多个特征聚合节点，采用双卷积块进行处理以确保特征的有效融合和非线性变换。之后，为处理融合后的复杂特征，模型在关键解码节点后插入特征平衡注意力模块（FBAM），依据当前任务类型动态调整权重。随后，引入特征互补精修模块（Feature Complementary Refinement Module，FCRM）进行最终的针对型增强。模型采用深度监督策略，从 3 个不同层级的解码节点输出分割结果，以促进梯度传播和特征学习，确保分割性能的稳定与精确。

2.2 AttUNet 解码器

针对 SAM2-Hiera[17]参数冻结的大规模预训练编码器，传统 U-Net[21]架构在特征融合阶段显露出固有的结构性缺陷。由于编码器参数冻结，编码器特征难以通过端到端训练进行充分适配，而传统 U-Net 相对简单的逐级上采样与多次非线性变换过程，在传递特征时极易丢失关键的浅层纹理与几何细节，难以支撑高精度的边界分割任务。鉴于 SAM2-Hiera 编码器已具备较强的多尺度表征能力，直接采用完整 UNet3+[18]解码结构可能导致特征冗余和局部响应过度增强，进而增加噪声干扰与过分割风险。基于此，本文在 UNet3+ 基础上进行剪枝与改进，提出了一种轻量化且高效的 AttUNet 解码器，其整体结构如图 2 所示，该图直观呈现了 AttUNet 解码器的四级层级排布、特征维度变化以及核心模块的连接逻辑。

AttUNet 解码器摒弃了简单的层级传递，转而采用全尺度跳跃连接策略。在每个解码阶段，特征聚合节点不仅接收来自上一级解码层的深层语义信息，还通过自适应上采样融合编码器各层级的多尺度特征。该结构能够加强浅层细节与深层语义之间的信息交互，缓解下采样过程中空间信息的损失，并提升模型对不同尺度目标的表征能力。在各关键特征聚合节点之后嵌入 FBAM，以动态协调全局语义信息和局部细节信息。与此同时，指导模块 GM 利用原始图像中的高频细节辅助解码过程，增强模型对局部结构的感知能力，并抑制过分割现象。最后，通过深度监督策略对多个解码层级进行联合优化，确保分割性能的稳定与精确。

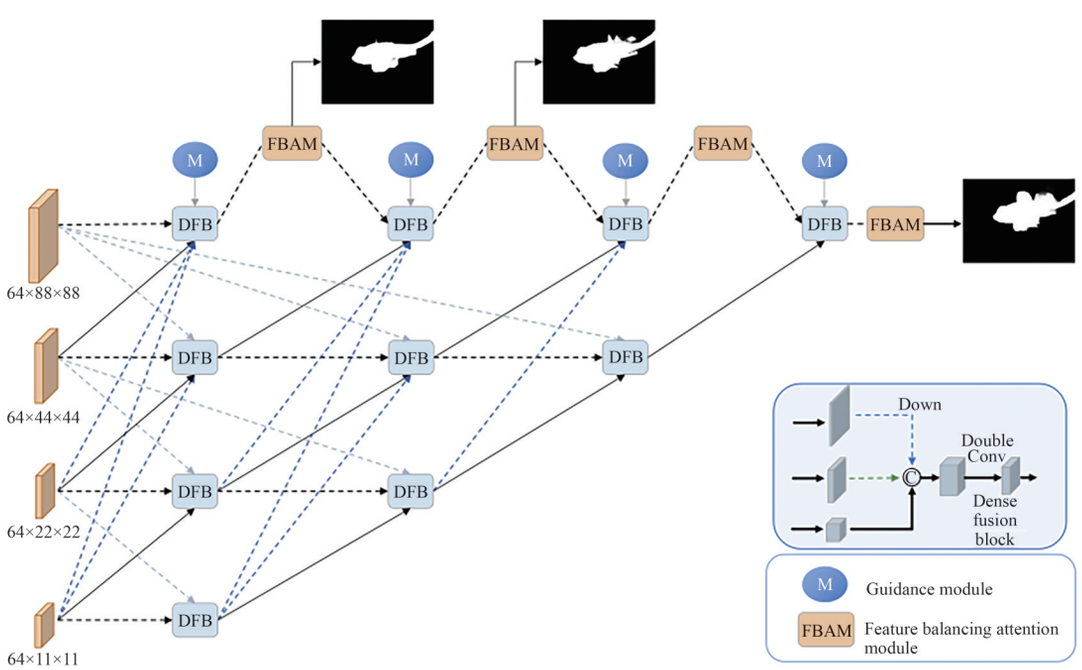


图 2 AttUNet 架构

Fig. 2 AttUNet architecture

2.3　特征平衡注意力模块

针对显著目标与伪装目标在特征响应的差异化需求，本文设计了特征平衡注意力模块(FBAM)，作为解码过程中的动态调和器，其整体结构如图3所示。采用通道与空间注意力双流并行机制，通过双独立通道注意力分支和多尺度空间注意力分支同时对输入特征进行协同加权，实现全局与局部的自适应协调。

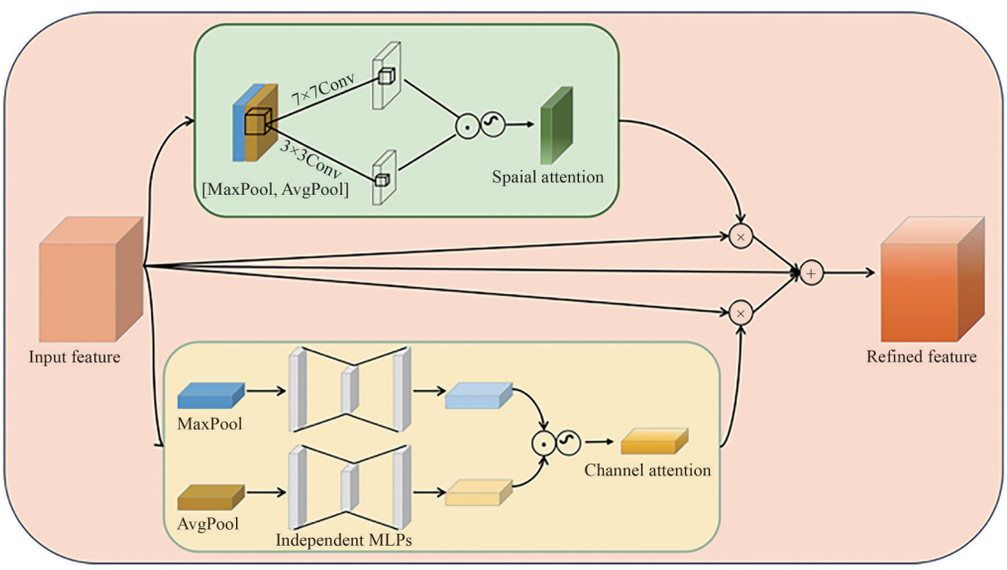


图3　FBAM架构

Fig. 3　FBAM architecture

通道注意力用于评估不同特征通道的重要性。为避免单一池化方式的局限性，本文采用全局平均池化(Global Average Pooling，GAP)和全局最大池化(Global Maximum Pooling，GMP)的组合，随后分别流经权重独立的双层MLP。

空间注意力模块则对原始输入特征沿通道维度取像素级均值与最大值并拼接为两通道特征图，随后并行送入7×7和3×3两个卷积分支。大卷积核负责捕获大范围上下文依赖，小卷积核则专注于保留局部细节与精确边缘信息，两分支的输出逐元素相加融合生成最终的空间注意力图，确保了模型在增强全局定位的同时，进一步提高了伪装目标的边界分割精度。

最终，通道加权特征、空间加权特征以及原始输入特征通过可学习的Softmax权重进行加权融合，得到增强后的输出特征。FBAM模块确保了通道和空间增强的解耦，并通过学习最优的融合权重，明确增强了关键特征的判别力，缓解了传统注意力机制难以同时平衡SOD和COD对特征要求的难题，对跨任务性能提升起到重要作用。FBAM的计算公式如下：

$$M_c(F)=\sigma\big(\mathrm{MLPavg}(\mathrm{AvgPool}(F))+\mathrm{MLPmax}(\mathrm{MaxPool}(F))\big),\tag{1}$$

$$F_{\mathrm{channel}}=F\odot M_c(F),\tag{2}$$

$$A_{\mathrm{input}}=\mathrm{Concat}(\mathrm{Avg}(X,\mathrm{dim}=1),\mathrm{Max}(X,\mathrm{dim}=1)),\tag{3}$$

$$M_s(F)=\sigma\big(\mathrm{BN}(\mathrm{Conv7}\times 7(A)+\mathrm{Conv3}\times 3(A))\big),\tag{4}$$

$$F_{\mathrm{spatial}}=F\odot M_s(F),\tag{5}$$

$$W=\mathrm{Softmax}([W_c,W_s,W_o]),\tag{6}$$

$$\mathrm{TAM}(F)=W_1\cdot F_{\mathrm{channel}}+W_2\cdot F_{\mathrm{spatial}}+W_3\cdot F,\tag{7}$$

其中：$F\in R^{B\times C\times H\times W}$为输入特征图；AvgPool(·)、MaxPool(·)为自适应全局平均和最大池化，输出$R^{B\times C\times 1\times 1}$；$\mathrm{MLP}_{\mathrm{avg}}$、$\mathrm{MLP}_{\mathrm{max}}$为两个权重完全独立的双层全连接网络，降维比为16；$\odot$为逐元素相乘，$\sigma(\cdot)$表示将通道注意力值归一化到(0,1)范围，作为权重因子；F_{channel}和F_{spaial}分别是经过通道注意力和空间注意力增强后的特征图；BN(·)的作用是对卷积结果进行归一化；W_c、W_s、W_o分别对应通道加权特征、空间加权特征和原始特征的

可学习初始权重。

2.4 特征互补精修模块

在FBAM完成特征动态平衡的基础上，模型构建了先平衡、后针对性增强的系统化解决方案，设计了特征互补精修模块(Feature Complementary Refinement Module，FCRM)，由位置注意力模块[22](Position Attention Module，PAM)和边界注意力模块(Boundary Attention Module，BAM)组成，分别从全局空间依赖和局部边界细节两个方面对特征进行精修。PAM主要用于捕捉长距离依赖与全局上下文整合，对于强调整体完整性的SOD，PAM能够将离散的特征碎片聚合为连贯主体，产生强增强效应；BAM则用于高频信息的过滤与边界轮廓的雕刻，用于优化COD难以辨认的边界精度。

位置注意力模块旨在通过模拟自注意力机制，建立特征图上任意两个位置之间的空间依赖关系，从而克服传统卷积操作的局部感受野限制。如图4所示，给定一个局部特征$A\in R^{C\times H\times W}$，首先将其输入到卷积层中以生成两个新的特征图$B$和$C$，其中$\{B,C\}\in R^{C\times H\times W}$。然后将它们重塑为$R^{C\times N}$，其中$N$是像素的数量。之后，对特征$C$的转置和$B$进行矩阵乘法，并应用一个Softmax层来计算空间注意力图$S\in R^{N\times N}$，同时将原始特征A输入另一个卷积层，生成一个新的特征图$D\in R^{C\times H\times W}$，并将其权重重塑为$R^{C\times N}$。对$D$和注意力图$S$的转置进行矩阵乘法，并将结果重塑回$R^{C\times H\times W}$。最后，将该结果乘以一个可学习的尺度参数$\alpha$，并与原始特征$A$进行元素级的求和操作，得到最终的输出特征$E\in R^{C\times H\times W}$。PAM的计算公式如下：

$$S_{ji}=\frac{\exp(B_i\cdot C_j)}{\sum_{i=1}^{N}\exp(B_i\cdot C_j)},\tag{8}$$

$$E_j=\alpha\sum_{i=1}^{N}(S_{ji}D_i)+A_j.\tag{9}$$

S_{ji}的意义：衡量第i个位置的特征对第j个位置的影响程度。特征表示越相似的两个位置，它们之间的相关性越高，S_{ji}越大。E_j的意义：最终特征E的每个位置是所有位置的特征根据空间注意力图S_{ji}进行加权求和的结果，并结合了原始特征A_j。α的意义：尺度参数α初始化为0，并通过学习逐渐分配更多的权重。

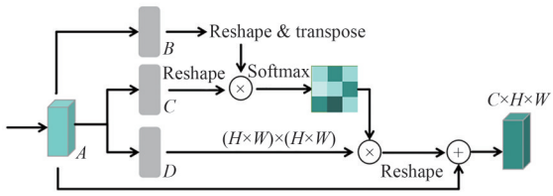


图4 位置注意力模块

Fig. 4 Position attention module

边界注意力模块专注于特征图的局部细节和边缘区域。如图5所示，给定输入特征图$X\in R^{C\times H\times W}$，输入边界检测分支和特征增强分支。首先，对输入特征施加一系列的卷积、归一化和激活操作。该分支使用通道缩减后，再通过两个3×3卷积层提炼边界特征。最后，通过一个1×1卷积层将通道数降为1，并应用Sigmoid激活函数，获得单通道的边界注意力图$M\in R^{1\times H\times W}$，其中$M$的值表示该位置是边界的可能性。同时，特征增强分支对输入X应用3×3卷积、批量归一化和ReLU激活函数，旨在提取更具表现力的特征表示。最后，将边界注意力图M转换为加权因子$(1+M)$。将加权因子与增强特征图F进行元素级乘法，生成加权特征F_{out}。随后，将加权特征F_{out}与原始输入特征图X进行残差连接，输出最终的增强特征图。BAM的计算公式如下：

$$M=\sigma\Big(\mathrm{Conv1\times1}\Big(\mathrm{ReLU}\Big(\mathrm{BN}\Big(\mathrm{Conv3\times3}(\cdots\mathrm{Conv}_{3\times3}(X))\Big)\Big)\Big)\Big),\tag{10}$$

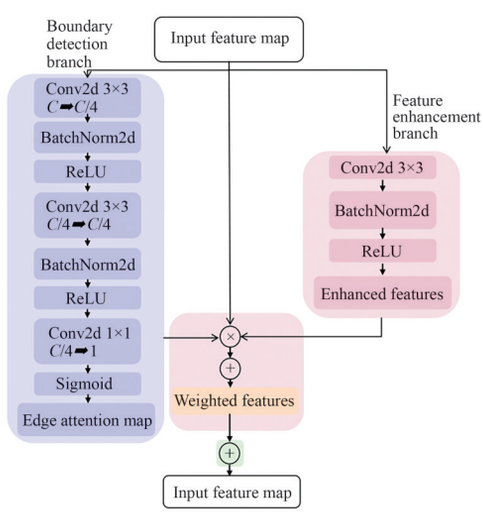


图5 边界注意力模块

Fig. 5 Boundary attention module

$$F_{\text{enhanced}} = \text{ReLU}\left(\text{BN}\left(\text{Conv}_{1\times 1}(X)\right)\right), \quad (11)$$

$$F_{\text{out}} = X + F_{\text{enhanced}} \odot (1 + M), \quad (12)$$

其中：σ代表Sigmoid激活函数，$\text{Conv}_{k\times k}$代表$k\times k$卷积操作，BN代表批量归一化，$\odot$表示元素级乘法。

3　实验与分析

3.1　数据集

为全面评估模型在不同视觉任务中的适应性与泛化能力，选取3个具有代表性的基准数据集开展实验。DUTS[19]/DUT-OMRON[23]数据集是目前广泛使用的通用显著目标检测数据集之一，它包含10 553张训练图像和5 019张测试图像。其子集DUT-OMRON进一步增加了难度，包含更为复杂的背景和细小物体，用以严格检验模型在复杂场景下的泛化能力。COD10K[1]数据集是迄今为止规模最大的伪装目标检测基准数据集之一，包含了5个大类和69个子类，共计5 066张图像（3 040张训练图像和2 026张测试图像）。通过在SOD和COD两类具有不同特征需求的数据集上进行统一评估，可以验证SAM2-UFFM对不同视觉特征的适应能力。引入钢材表面缺陷数据集NEU-SEG[20]，该数据集来源于NEU-DET，选取其中不同尺度的3类缺陷：斑块（Patches）、夹杂（Inclusion）以及划痕（Scratches）。将样本按分类以8∶2比例随机划分训练集与测试集，分别进行综合测试以及各分类测试。

3.2　实验细节

模型使用PyTorch 2.4.1，在NVIDIA GeForce RTX 3080TiGPU上进行所有实验。采用预训练的SAM2Hiera-large模型作为编码器。输入图像被调整为352×352的尺寸，并采用两种数据增强策略：随机垂直和水平翻转。在训练过程中，批处理大小设定为12，并在训练过程中使用Adam优化器，初始的学习率设置为2e−4，模型共训练20轮。

3.3　评估指标

实验采用多项国际公认的评价指标对模型性能进行评估，包括交并比（IoU）[24]、结构性度量（S_a）[25]、F度量（F_β）[26]、加权F指标（F_β^w）[26]、平均绝对误差（MAE）[27]、E指标（E_ϕ）[28]，确保从不同维度客观反映模型的综合表现。

交并比（IoU）用于衡量预测分割区域与真实标注区域之间的重叠程度，如式（13）所示：

$$\text{IoU} = \frac{\text{TP}}{\text{TP} + \text{FN} + \text{FP}}, \quad (13)$$

其中，TP、FP和FN分别表示第i个类别的真正例、假正例和假反例。

结构性度量S_a评估预测结果与标注图像之间的区域级和对象级结构相似性，如式（14）所示：

$$S_a = \alpha \cdot S_o(P, G) + (1 - \alpha) \cdot S_r(P, G), \quad (14)$$

其中：S_o和S_r分别表示对象级和区域级的结构相似性；根据其他研究中的经验，这里的α设置为0.5。

F度量（F_β）通过计算精确率与召回率的加权调和平均，以综合评估预测结果的准确性与完整性，如式（15）所示：

$$F_\beta = \frac{(1 + \beta^2) \times \text{Precision} \times \text{Recall}}{\beta^2 \times \text{Precision} + \text{Recall}}, \quad (15)$$

其中：Precision和Recall分别表示精确率和召回率；根据其他研究中的经验，在显著性目标检测中β^2通常设置为0.3。

加权F度量（F_β^w）在传统F度量的基础上，为靠近分割边界的难例像素分配更高权重，使得评估更侧重于结构边界的精确性，如式（16）所示：

$$F_\beta^w = \frac{(1 + \beta^2) \times \text{Precision}_w \times \text{Recall}_w}{\beta^2 \times \text{Precision}_w + \text{Recall}_w}, \quad (16)$$

其中，Precision_w和Recall_w分别为基于像素权重图计算得到的加权精确率与加权召回率。

E指标（E_ϕ）使用一个增强对齐矩阵（O_{FM}）联合计算图像级的统计信息和像素级的匹配信息，可以同时衡量预测图的整体完整性和局部精确性，如式（17）所示：

$$E_\phi = \frac{1}{w \times h} \sum_{x=1}^{w} \sum_{y=1}^{h} \phi^{\text{FM}}(x, y), \quad (17)$$

其中：W表示图像宽度，H表示图像高度，ϕ_{FM}为在位置(x, y)处的增强对齐值。

平均绝对误差（MAE）直接计算预测显著性图与真实二值标注图之间每个像素的绝对误差均值，能够直观反映预测的总体误差水平，如式（18）所示：

$$\text{MAE} = \frac{1}{W \times H} \sum_{x=1}^{W} \sum_{y=1}^{H} |P(x, y) - G(x, y)|, \quad (18)$$

其中：W表示图像宽度，H表示图像高度，P为预测图，G为真实标注图。

3.4 实验结果及对比

将SAM2-UFFM与COD和SOD领域最先进的方法进行比较，包括SINet[1]、PFNet[2]、OCENet[3]、MDSAM[9]、UGRAN[11]等。为确保评估的公平性与客观性，所有对比实验均直接采用原作者开源的预测图，并使用统一的公式进行指标计算。

在SOD任务中，针对现有模型在复杂背景干扰下容易出现主体分割不完整或边缘模糊的瓶颈，SAM2-UFFM展现出了显著的优越性。表1实验结果表明，SAM2-UFFM在DUTS[19]和DUT-OMRON[23]两个测试集上，S_a和MAE二个关键指标均位列第一，而E_ϕ略低于UGRAN模型，主要原因是UGRAN引入了不确定性引导的多尺度细化分支，专门针对SOD任务优化边界；本文模型侧重跨任务通用平衡，未针对单一任务做定制化强化，后续可引入不确定性分支进一步提升边界精度。

表1 不同模型在DUTS和DUT-OMRON上3种评价指标的定量结果

Tab. 1 Quantitative results of different models for three evaluation metrics on DUTS and DUT-OMRON

基线模型	DUTS			DUT-OMRON		
	S_a↑	E_ϕ↑	MAE↓	S_a↑	E_ϕ↑	MAE↓
U2Net	0.874	0.884	0.044	0.847	0.872	0.054
EDN	0.892	0.925	0.035	0.850	0.877	0.049
MENeT	0.905	0.937	0.028	0.850	0.891	0.045
MDSAM	0.920	0.949	0.024	0.878	0.910	0.039
UGRAN	0.931	0.964	0.022	0.878	0.911	0.045
SAM2-UNet	0.934	0.959	0.020	0.884	0.912	0.039
Ours	**0.936**	**0.961**	**0.019**	**0.893**	**0.920**	**0.035**

注：粗体数字表示最优值

在DUTS数据集上，本文模型取得了最优结果，S_a为0.936，E_ϕ为0.959，MAE为0.019。即使在背景复杂的DUTS-OMRON数据集上，模型同样保持了领先优势，模型取得了S_a(0.893)、E_ϕ(0.920)以及MAE(0.035)的优异成绩。

从图6可以看出，在面对极小目标(第1行)和占据全图的大目标(第2行)时，得益于编码器多尺度特征的有效提取与AttUNet全尺度跳跃连接的聚合，有效避免了局部特征的丢失或全局视野的局限，分割掩码完整。在目标与环境对比度极低(第4行)或背景纹理杂乱(第3行、第5行)的环境中，其他基线模型普遍出现了严重的误判或过分割现象，而SAM2-UFFM凭借特征平衡与精修模块，能够精准抑制背景噪声，清晰地剥

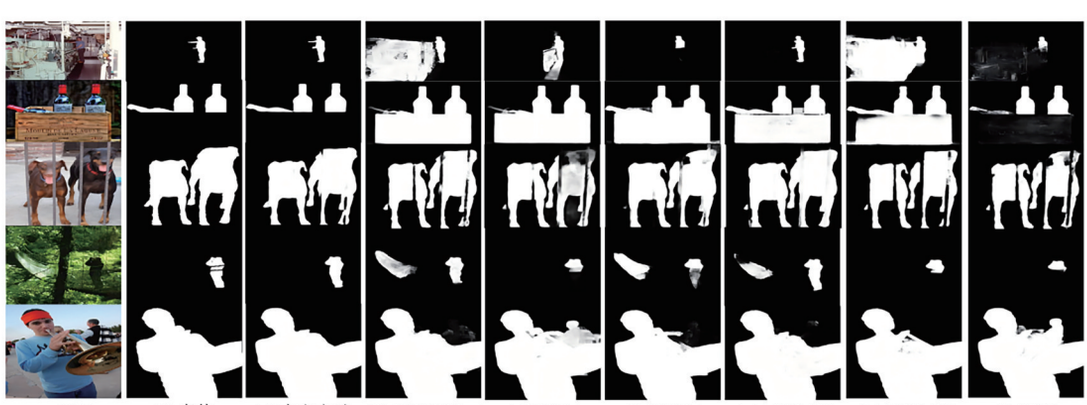


图6 本文方法与显著目标检测模型的可视化对比(DUTS数据集图例)

Fig. 6 Visual comparison of our method with SOD models (Legend of DUTS dataset)

离出目标轮廓。在图 7 中，面对多目标并存（第 3 行）和复杂前景目标（第 4～5 行），本文模型同样展现出优秀的细节雕刻能力。模型不仅准确识别了所有显著实例，还能精准捕捉目标的纤细分支与复杂边界，其预测图的内部连贯性与边缘锐利度均优于其他对比模型。

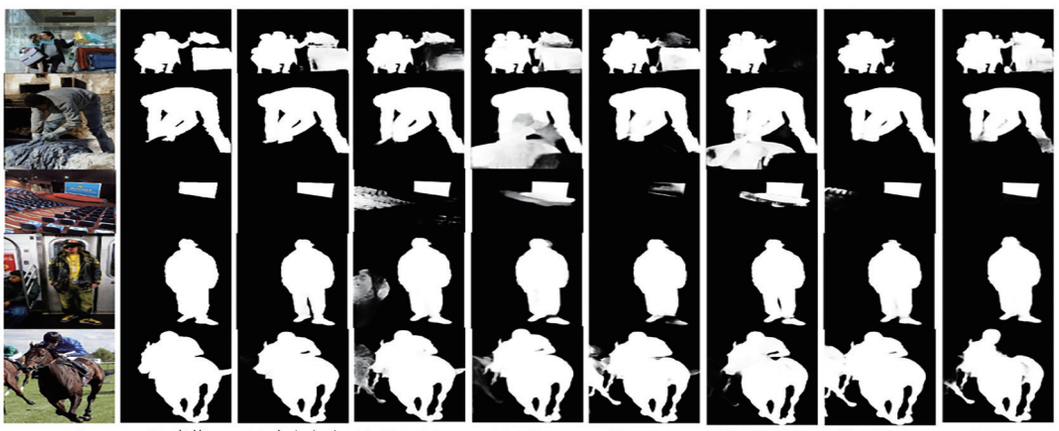


图 7　本文方法与显著目标检测模型的可视化对比（DUT-OMRON 数据集图例）

Fig. 7　Visual comparison of our method with SOD models (Legend of DUT-OMRON dataset)

COD10K[1]数据集专注于具有高度伪装特性的目标分割，要求模型能够识别并分割出与背景高度融合的目标。

实验结果显示，SAM2-UFFM 在 COD10K 数据集上的所有 4 项指标均达到了最佳性能（见表 2）。其中，S_a 为 0.896，F_β 为 0.826，E_ϕ 为 0.950，MAE 低至 0.018。

如图 8 所示，在面对与背景纹理高度融合且边缘极度模糊的目标（第 1～2 行）时，传统模型极易发生严重的漏检或边缘粘连，SAM2-UFFM 得益于 FBAM 在解码阶段动态调配全局语义与局部细节权重的机制，成功锁定了目标的整体方位，有效滤除了背景噪声的干扰。无论大尺度目标（第 3 行）还是小尺度目标（第 4 行），SAM2-UFFM 均能完整输出目标掩码，未出现传统模型中常见的局部欠分割或内部空洞现象。最后，即使当目标遭遇环境部分遮挡（第 5 行）时，本文模型依然能够凭借强大的边界特征强化机制，精准定位并剥离目标足部等残缺区域。

表 2　不同模型在 COD10K 上对 4 种评价指标的定量结果

Tab. 2　Quantitative results of different models for four evaluation metrics on COD10K

基线模型	COD10K			
	S_a↑	F_β↑	E_ϕ↑	MAE↓
SINet	0.776	0.667	0.864	0.043
PFNet	0.800	0.676	0.877	0.040
C2FNet	0.813	0.686	0.890	0.036
OCENET	0.832	0.745	0.890	0.032
ZoomNet	0.838	0.766	0.888	0.029
FEDER	0.844	0.784	0.911	0.029
SAM2-UNet	0.880	0.789	0.936	0.021
Ours	**0.896**	**0.826**	**0.950**	**0.018**

注：粗体数字表示最优值

NEU-SEG[20]数据集专注于复杂的工业钢材表面缺陷分割任务，涵盖了显著性斑块、不规则夹杂以及极具挑战性的细微划痕，对模型的多尺度感知能力和纹理抗干扰能力提出了高要求。

将本文模型与传统分割模型 DeeplabV3+[29]，PSPNet[30]做对比。如表 3 和表 4 实验结果显示，SAM2-UFFM 模型在 NEU-SEG 数据集上的所有评估指标均取得了最佳性能。特别是在综合测试的 IoU 指标上，模型达到 0.853，相较于当前最先进的次优基准模型 SAM2-UNet[12]（0.835）有显著提升。此外，衡量预测误差的 MAE 指标降低至 0.011。在子类测试实验中，SAM2-UFFM 表现同样出色：对于边缘极难界定的夹杂类缺陷，IoU 达到 0.820；对于划痕类缺陷，IoU 达到 0.822。

如图 9 所示，面对视觉特征极度微弱、与金属



图 8 本文方法与伪装目标检测模型的可视化对比(COD10K 数据集图例)

Fig. 8 Visual comparison of our method with COD models (Legend of COD10K dataset)

表 3 不同模型在 NEU-SEG 数据集上的综合性能与斑块类缺陷分割结果对比

Tab. 3 Comparison of segmentation results for comprehensive performance and patch defects of different models on NEU-SEG dataset

基线模型	Overall			Patches		
	IoU↑	S_a↑	MAE↓	IoU↑	S_a↑	MAE↓
PSPNet	0.747	0.868	0.020	0.873	0.920	0.027
UNet++	0.763	0.871	0.019	0.895	0.931	0.022
DeepLabV3+	0.809	0.909	0.014	0.904	0.945	0.019
SAM2-UNet	0.830	0.915	0.013	0.907	0.944	0.020
Ours	**0.853**	**0.923**	**0.011**	**0.917**	**0.945**	**0.017**

注:粗体数字表示最优值

表 4 不同模型在 NEU-SEG 数据集上的划痕与夹杂类缺陷分割结果对比

Tab. 4 Comparison of segmentation results for scratches and inclusion defects of different models on NEU-SEG dataset

基线模型	Scratches			Inclusion		
	IoU↑	S_a↑	MAE↓	IoU↑	S_a↑	MAE↓
PSPNet	0.692	0.885	0.017	0.677	0.828	0.017
UNet++	0.756	0.885	0.012	0.632	0.795	0.022
DeepLabV3+	0.775	0.910	0.011	0.749	0.873	0.011
SAM2-UNet	0.800	0.908	0.010	0.798	0.894	0.008
Ours	**0.822**	**0.915**	**0.008**	**0.820**	**0.908**	**0.007**

注:粗体数字表示最优值

背景纹理高度融合的细微划痕(第 1 行,第 3 行),传统模型极易受噪声干扰而发生分割断裂或漏检。而 SAM2-UFFM 对高频特征的敏锐捕捉,成功滤除了金属底纹噪声,保证了狭长缺陷特征的连续性与边缘锐利度。对于形态不规则、甚至内部存在复杂孔洞的大尺度夹杂物与斑块(第 2 行),得益于 PAM 建立的全局长距离依赖关系,模型有效克服了局部感受野受限带来的内部空洞或欠分割现象,完整且精细地分割出了缺陷的真实边缘轮廓。

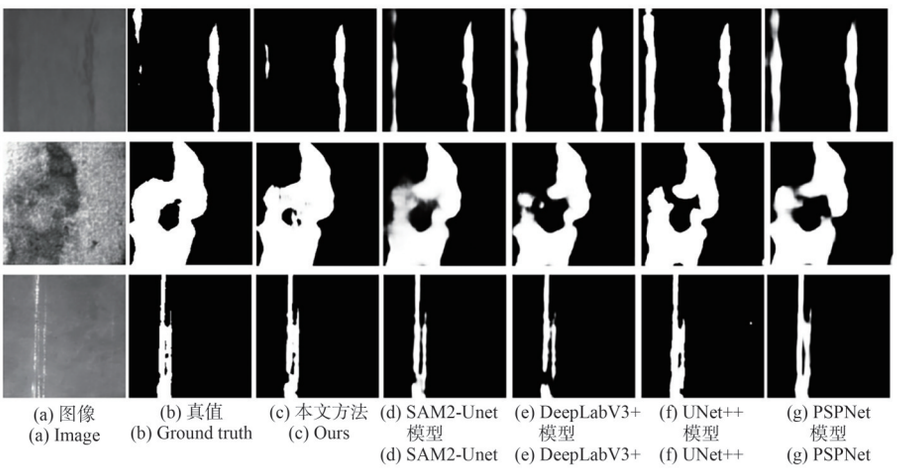


图 9　本文方法与其他分割模型的可视化对比（NEU-SEG 数据集图例）

Fig. 9　Visual comparison of our method with other methods (Legend of NEU-SEG dataset)

3.5　消融实验

为验证所提各组件的有效性，本节在 COD10K[1]、DUT-OMRON[23]数据集上开展系统性消融实验。基线模型仅采用冻结的 SAM2-Hiera[17]主干配合轻量 RFB 模块与简单逐级上采样解码器。当引入特征平衡注意力模块后，模型具备在全局语义与局部细节之间动态分配注意力的能力，COD10K 与 DUT-OMRON 上的 S_α 分别提升 0.93% 和 1.63%。进一步加入特征互补精修模块后，模型能够针对“显著性”与“伪装性”这两类对立线索分别施加强化，位置注意力确保显著目标的整体一致性，边界注意力精准捕捉伪装目标的微弱边缘，S_α 再次大幅提升 2.55% 和 2.68%，MAE 在 COD10K 上降至 0.020。各配置方案的性能数据如表 5 所示，可清晰反映各组件的独立性能与模块间的协同增益。

表 5　不同模块的有效性分析

Tab. 5　Effectiveness analysis of different modules

Configurations	COD10K				DUT-OMRON		
	S_α↑	F_β↑	E_ϕ↑	MAE↓	S_α↑	E_ϕ↑	MAE↓
Baseline	0.861	0.786	0.921	0.026	0.858	0.896	0.043
Baseline＋FBAM	0.869	0.798	0.925	0.024	0.872	0.902	0.040
Baseline＋FCRM	0.883	0.814	0.932	0.022	0.881	0.910	0.039
Baseline＋FBAM＋FCRM	0.889	0.820	0.940	0.020	0.887	0.914	0.037
Baseline＋FBAM＋FCRM＋AttUNet	0.896	0.826	0.950	0.018	0.893	0.920	0.035

二者协同作用时产生正向增益，COD10K 上的 S_α 累计提升 3.25%，体现出动态平衡与针对性增强。最终替换为 AttUNet 全尺度解码器后，COD10K 与 DUT-OMRON 的 S_α 进一步提升 4.06% 和 4.08%。上述结果充分证明：特征平衡注意力模块赋予模型动态平衡特征能力，特征互补精修模块提供对立任务的针对性增强，而 AttUNet 解码器则通过全尺度特征融合充分挖掘多层级特征信息，三者形成互补，共同提升了模型的整体性能。

4　结　　论

针对显著目标与伪装目标分割任务中存在的特征需求矛盾，本文提出了一种基于特征动态平衡的通用分割框架 SAM2-UFFM。通过构建参数高效的 AttUNet 解码体系，利用全尺度跳跃连接成功打破了层级特征壁垒。为解决矛盾任务的特征平衡难题，提出了作为特征平衡器的特征平衡注意力模块，用于动态平衡全局语义与局部细节。为了实现显著-伪装针对性增强，设计了专

注于对立任务的特征互补精修模块，通过位置注意力强化显著目标的全局感知，边界注意力强化伪装目标的边界精细分割。最后采用深度监督策略，从多层级输出分割结果，确保模型分割稳定与精确。实验结果表明，SAM2-UFFM在伪装目标检测数据集COD10K上S_α提升至0.896，MAE降低至0.018；显著目标检测数据集DUT-OMRON上S_α提升至0.893，MAE降低至0.035。基于NEU-SEG钢材表面数据集的泛化实验进一步验证了SAM2-UFFM在复杂背景中依旧保持着较强的检测能力。综上所述，该模型在兼顾全局定位与局部细节上的出色性能，为同时高效处理显著与伪装目标的通用分割模型提供了有效技术方案。

参 考 文 献：

[1] FAN D P, JI G P, SUN G L, *et al*. Camouflaged object detection [C]//*Proceedings of the* 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*). Seattle: IEEE, 2020: 2774-2784.

[2] MEI H Y, JI G P, WEI Z Q, *et al*. Camouflaged object segmentation with distraction mining [C]//*Proceedings of the* 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*). Nashville: IEEE, 2021: 8768-8777.

[3] LIU J W, ZHANG J, BARNES N. Modeling aleatoric uncertainty for camouflaged object detection [C]//*Proceedings of the* 2022 *IEEE/CVF Winter Conference on Applications of Computer Vision* (*WACV*). Waikoloa: IEEE, 2022: 2613-2622.

[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]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*). New Orleans: IEEE, 2022: 2150-2160.

[5] HE C M, LI K, ZHANG Y C, *et al*. Camouflaged object detection with feature decomposition and edge reconstruction [C]//*Proceedings of the* 2023 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*). Vancouver: IEEE, 2023: 22046-22055.

[6] 李明岩，吴川，朱明. 基于边缘增强和特征融合的伪装目标分割[J]. 液晶与显示，2024，39(1)：48-58.
LI M Y, WU C, ZHU M. Camouflaged object segmentation based on edge enhancement and feature fusion [J]. *Chinese Journal of Liquid Crystals and Displays*, 2024, 39(1): 48-58. (in Chinese)

[7] QIN X B, ZHANG Z C, HUANG C Y, *et al*. U^2-Net: going deeper with nested U-structure for salient object detection [J]. *Pattern Recognition*, 2020, 106: 107404.

[8] WU Y H, LIU Y, ZHANG L, *et al*. EDN: salient object detection via extremely-downsampled network [J]. *IEEE Transactions on Image Processing*, 2022, 31: 3125-3136.

[9] GAO S X, ZHANG P P, YAN T Y, *et al*. Multi-scale and detail-enhanced segment anything model for salient object detection [C]//*Proceedings of the* 32*nd ACM International Conference on Multimedia*. Melbourne: ACM, 2024: 9894-9903.

[10] KIRILLOV A, MINTUN E, RAVI N, *et al*. Segment anything [C]//*Proceedings of the IEEE/CVF International Conference on Computer Vision* (*ICCV*). Paris: IEEE, 2023: 3992-4003.

[11] YUAN Y, GAO P, DAI Q, *et al*. Uncertainty-guided refinement for fine-grained salient object detection [J]. *IEEE Transactions on Image Processing*, 2025, 34: 2301-2314.

[12] XIONG X Y, WU Z H, TAN S Y, *et al*. SAM2-UNet: segment anything 2 makes strong encoder for natural and medical image segmentation [J]. *Visual Intelligence*, 2026, 4(1): 2.

[13] RAVI N, GABEUR V, HU Y T, *et al*. SAM 2: segment anything in images and videos [C]//*Proceedings of the Thirteenth International Conference on Learning Representations*. Singapore: OpenReview.net, 2025.

[14] 邬开俊，张治瑞，汪滢，等. 全局感知与多尺度特征融合的城市道路语义分割[J]. 光学 精密工程，2025，33(14)：2262-2277.
WU K J, ZHANG Z R, WANG Y, *et al*. Urban road semantic segmentation with global awareness and multi-scale feature fusion [J]. *Optics and Precision Engineering*, 2025, 33(14): 2262-2277. (in Chinese)

[15] 赵林，马思奇，方一鸣，等. 基于多尺度特征融合与分块注意力的齿轮表面缺陷分割算法[J]. 光学 精密工程，2025，33(22)：3536-3548.
ZHAO L, MA S Q, FANG Y M, *et al*. Gear surface defect segmentation algorithm based on multi-scale feature

fusion and block-wise attention [J]. *Optics and Precision Engineering*, 2025, 33(22): 3536-3548. (in Chinese)

[16] 刘玉,于明,朱叶. 基于双重聚合和自合并网络的小样本图像语义分割[J]. 液晶与显示,2024,39(10):1421-1430.
LIU Y, YU M, ZHU Y. Bi-aggregation and self-merging network for few-shot image semantic segmentation [J]. *Chinese Journal of Liquid Crystals and Displays*, 2024, 39(10): 1421-1430. (in Chinese)

[17] RYALI C, HU Y T, BOLYA D, *et al*. Hiera: a hierarchical vision Transformer without the bells-and-whistles [C]// *Proceedings of the International Conference on Machine Learning*. Honolulu: PMLR, 2023, 202: 29441-29454.

[18] HUANG H M, LIN L F, TONG R F, *et al*. UNet 3+ : a full-scale connected UNet for medical image segmentation [C]. ICASSP 2020—2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Barcelona: IEEE, 2020: 1055-1059.

[19] WANG L J, LU H C, WANG Y F, *et al*. Learning to detect salient objects with image-level supervision [C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*). Honolulu: IEEE, 2017: 3796-3805.

[20] HE Y, SONG K C, MENG Q G, *et al*. An end-to-end steel surface defect detection approach via fusing multiple hierarchical features [J]. *IEEE Transactions on Instrumentation and Measurement*, 2020, 69(4): 1493-1504.

[21] RONNEBERGER O, FISCHER P, BROX T. U-Net: convolutional networks for biomedical image segmentation [C]// *Proceedings of* 18*th International Conference on Medical Image Computing and Computer-Assisted Intervention-MICCAI* 2015. Munich: Springer, 2015: 234-241.

[22] FU J, LIU J, TIAN H J, *et al*. Dual attention network for scene segmentation [C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*). Long Beach: IEEE, 2019: 3141-3149.

[23] YANG C, ZHANG L H, LU H C, *et al*. Saliency detection via graph-based manifold ranking [C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Portland: IEEE, 2013: 3166-3173.

[24] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation [C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*). Boston: IEEE, 2015: 3431-3440.

[25] FAN D P, CHENG M M, LIU Y, *et al*. Structure-measure: a new way to evaluate foreground maps [C]// *Proceedings of the IEEE International Conference on Computer Vision* (*ICCV*). Venice: IEEE, 2017: 4558-4567.

[26] MARGOLIN R, ZELNIK-MANOR L, TAL A. How to evaluate foreground maps [C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Columbus: IEEE, 2014: 248-255.

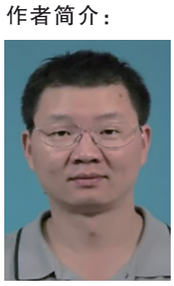
[27] PERAZZI F, KRÄHENBÜHL P, PRITCH Y, *et al*. Saliency filters: contrast based filtering for salient region detection [C]. 2012 IEEE Conference on Computer Vision and Pattern Recognition. Providence: IEEE, 2012: 733-740.

[28] 范登平,季葛鹏,秦雪彬,等. 认知规律启发的物体分割评价标准及损失函数[J]. 中国科学:信息科学,2021,51(9):1475-1489.
FAN D P, JI G P, QIN X B, *et al*. Cognitive vision inspired object segmentation metric and loss function [J]. *Scientia Sinica Informationis*, 2021, 51(9): 1475-1489. (in Chinese)

[29] CHEN L C, ZHU Y K, PAPANDREOU G, *et al*. Encoder-decoder with Atrous separable convolution for semantic image segmentation [C]//*Proceedings of the* 15*th European Conference on Computer Vision-ECCV* 2018. Munich: Springer, 2018: 833-851.

[30] ZHAO H S, SHI J P, QI X J, *et al*. Pyramid scene parsing network [C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu: IEEE, 2017: 6230-6239.

作者简介:



邓万宇,男,博士,教授,2010年于西安交通大学获得博士学位,主要从事计算机视觉及图像处理方面的研究。E-mail:dengwanyu@126.com



刘洋伍,男,硕士研究生,2023年于重庆三峡学院获得学士学位,主要从事图像处理算法方面的研究。E-mail:liuyangwu2023@163.com