

DFRNet:融合扩散-聚焦物理机制的 语义分割模型研究

黄依莎¹,姜林^{1,2*},管亚菲¹,张亚莎³,梁欣¹,曾伟豪¹,方晓萍⁴

(1. 湖南工商大学人工智能与先进计算学院,湖南长沙 410000;2. 湘江实验室,湖南长沙 410000;
3. 湖南工商大学智能工程与制造学院,湖南长沙 410000;4. 湖南工商大学数学与统计学院,湖南长沙 410000)

摘要: 针对图像语义分割任务中下采样过程导致的信息丢失问题,以及现有上采样方法在处理复杂场景中普遍存在的全局信息丢失、细节模糊、生成过程不稳定及信息冗余等局限,本文提出了一种融合物理扩散-聚焦机制的轻量级语义分割模型——DFRNet. 该模型引入了液体表面张力的扩散-聚焦机制,并进一步设计了动态上下文窗口选择(Dynamic context Window Selection, DWS)模块作为调节优化机制,从而实现了物理启发的能量传播上采样(Physics-Inspired Energy Propagation Upsampling, PIEPU)新方法. 该方法包含扩散、聚焦与调节三大机制,分别承担全局上下文信息扩展、关键区域特征增强与信息流动优化的功能,协同提升模型在复杂场景下的细粒度感知与语义一致性表达能力. 在7种类别14个数据集的验证表明:所提DFRNet在mIoU、 F_1 分数和Accuracy指标上均领先于其他先进模型,且在不同数据集上,mIoU提升幅度为0.165%~4.259%; F_1 分数提升为0.140%~2.888%;Accuracy提升为0.035%~1.386%,验证了方法在多样化任务环境下的鲁棒性与泛化能力. 本文模型参数量仅为3.34 MB,可满足轻量化实时应用.

关键词: 语义分割;上采样;扩散-聚焦;全局上下文

基金项目: 湘江实验室重大项目(No.23XJ01003, No.23XJ01009);湖南省教育厅科学研究重点项目(No.22A0441)

中图分类号: TP391 **文献标识码:** A **文章编号:** 0372-2112(2025)06-1755-16

电子学报 URL: <http://www.ejournal.org.cn> **DOI:** 10.12263/DZXB.20250186

第二十七届中国科协年会学术论文

DFRNet: A Semantic Segmentation Method Inspired with Physical Mechanism of Diffusion-Focus

HUANG Yi-sha¹, JIANG Lin^{1,2*}, GUAN Ya-fei¹, ZHANG Ya-sha³, LIANG Xin¹,
ZENG Wei-hao¹, FANG Xiao-ping⁴

(1. School of AI and Advanced Computing, Hunan University of Technology and Business, Changsha, Hunan 410000, China;

2. Xiangjiang Laboratory, Changsha, Hunan 410000, China;

3. School of Intelligent Engineering and Manufacturing, Hunan University of Technology and Business, Changsha, Hunan 410000, China;

4. School of Mathematics and Statistics, Hunan University of Technology and Business, Changsha, Hunan 410000, China)

Abstract: To address the information loss induced by downsampling in image semantic segmentation tasks, as well as the widespread limitations of existing upsampling methods: such as inadequate global perception, blurred fine-grained reconstruction, unstable generation processes, and redundant information handling in various scenarios, this paper proposes a lightweight semantic segmentation model, DFRNet, which incorporates a physics-inspired diffusion-focusing mechanism. Specifically, inspired by the surface tension of liquids, the model introduces a diffusion-focusing mechanism and designs a dynamic context window selection (DWS) module to optimize information flow, thereby implementing the physics-inspired energy propagation upsampling (PIEPU) framework. PIEPU comprises three core modules: diffusion, focusing, and regulation. These modules collaboratively enhance global contextual propagation, critical region feature reinforcement, and optimized information flow, thereby significantly improving fine-grained perception and semantic consistency across complex scenarios. Extensive experiments conducted on 14 datasets covering 7 semantic categories demonstrate that DFRNet consis-

tently achieves superior performance over state-of-the-art methods in terms of mean intersection over union (mIoU), F_1 score, and Accuracy. Specifically, mIoU improvements range from 0.165% to 4.259%, F_1 score gains span 0.140% to 2.888%, and Accuracy enhancements vary from 0.035% to 1.386% across diverse datasets. These results validate the robustness and generalization capability of the proposed approach. Notably, DFRNet has a model size of only 3.34 MB, making it suitable for lightweight real-time applications.

Key words: semantic segmentation; upsampling; diffusion focus; global context

Foundation Item(s): Major Project of Xiangjiang Laboratory (No.23XJ01003, No.23XJ01009); Key Scientific Research Project of Hunan Provincial Department of Education (No.22A0441)

1 引言

语义分割是计算机视觉领域的重要方向,旨在将图像像素分为特定类别,并给出具体语义标签^[1],可提升图像感知和理解能力.它已广泛应用于自动驾驶^[2]、工业自动化^[3]、医疗影像分析^[4]、卫星遥感图像分析^[5]、智能监控^[6]、增强现实和虚拟现实^[7]等领域.

主流语义分割技术主要依赖深度学习模型,尤其是卷积神经网络及其变体,它们能够学习图像的多层次特征,实现像素级别的精确分类.常用模型有 FCN^[8]、U-Net^[9-11]、DeepLab 系列^[12-14]、SegNet^[15,16]、PSPNet^[17]等.其中,U-Net因其高效的结构和优异的分割性能,成为广泛使用的语义分割模型之一.其编码器-解码器结构能够高效地捕捉图像的全局特征并恢复局部细节,跳跃连接进一步增强了细节信息的保留,提升了分割精度.此外,U-Net在小样本学习和数据不平衡问题上也表现出色^[18,19].

U-Net由编码器(encoder)、解码器(decoder)和跳跃连接(skip-connections)和输出层(output)四个部分组成^[9].在这四个部分中,解码器对于U-Net至关重要,它不仅恢复图像的空间分辨率,还通过跳跃连接融合编码器的多层次特征,结合低层次的细节信息和高层次的语义信息,提升分割精度,在保留全局语义信息的同时,精确地恢复图像中的如边缘、纹理等的局部细节.

在U-Net中,上采样是解码器的核心操作,用于恢复分辨率和细节信息.它通过转置卷积逐步恢复特征图的分辨率,使其与输入图像一致.上采样操作尤其有助于恢复边界区域的细节,提升分割性能.

现有上采样方法主要分为基于卷积和插值的上采样、基于扩散模型和生成模型的上采样以及基于多尺度特征融合的上采样三类.

(1)基于卷积和插值的上采样方法通过预定义的卷积操作或插值算法逐步恢复特征图的分辨率.Li等^[20]提出了双密集上采样,分两步实现2倍和4倍上采样;Zhou等^[21]在Deeplabv3+中结合双线性插值和卷积进行分层上采样;Xun等^[22]在脑肿瘤MRI图像分割任务中,使用双线性插值上采样方法;徐亮亮等^[23]在城市绿地遥感分类任务中,采用内容感知的上采样方法,根

据输入特征图的内容动态生成上采样核.U-Net中也是采用转置卷积或双线性插值来恢复特征图的分辨率^[9].但是这种方法在处理图像细节时缺乏动态调整能力,难以适应局部特征变化,导致全局信息丢失,同时也易导致平坦区域过于平滑和边界区域模糊.

(2)基于扩散模型的上采样方法通过生成模型的学习能力,逐步恢复高分辨率特征.Magnussen等^[24]提出基于去噪扩散概率模型,结合条件掩码引导生成过程;Xia等^[25]提出利用显式和隐式先验知识生成分割掩膜.然而这种方法训练难度高,生成过程不稳定,导致输出结果不一致或出现伪影,且对训练数据分布敏感,推理过程具有随机性.

(3)基于多尺度特征融合的上采样方法通过融合不同尺度的特征信息,逐步恢复高分辨率图像的细节.Chen等^[26]提出的EFDCNet通过高层语义信息指导下采样;张银胜等^[27]提出多层特征融合模块;张大锦等^[28]提出利用轮廓波变换的多尺度特性重构上采样.尽管这种方法融合多尺度特征,有效捕捉上下文信息,但是多尺度特征融合易导致信息冗余或特征冲突,影响分割精度.

针对现有上采样方法在处理复杂场景下存在的全局信息丢失、细节模糊、生成过程不稳定、信息冗余等问题,本文受液体表面张力启发,提出了一种基于扩散与聚焦机制的新型上采样方法,为复杂场景下的图像分割任务提供高效且鲁棒的解决方案.本文的贡献与创新点如下.

(1)扩散机制.为解决全局信息丢失的问题,本文在上采样初期及特征提取阶段引入扩散机制,通过扩大感受野,捕获更广泛的上下文信息,为后续的细节恢复提供可靠的语义支持.扩散机制有效解决了传统方法因固定感受野导致的全局信息丢失问题,显著提升了模型对复杂场景的建模能力.

(2)聚焦机制.为解决平滑效应导致的边界模糊问题,本文在图像的边缘、纹理等关键区域引入聚焦机制,优先强化这些区域的细节信息,确保细节信息在上采样过程中得到充分保留.聚焦机制有效解决了传统方法因平滑效应导致的细节模糊问题,显著提升了模型对边缘和纹理等细微结构的恢复能力.

(3)动态调节机制.为解决伪影和多尺度特征融合信息冗余的问题,本文提出动态调节机制,通过聚焦系数的大小动态调整窗口大小和空间分布权重,灵活调节扩散过程与聚焦过程的平衡.这种动态调节不仅优化了计算资源的分配,还减少了对不同任务需求的人工干预,显著提升了模型的适应性和鲁棒性.

2 本文动机

语义分割中,上采样模块通过低分辨率到高分辨率的恢复,旨在捕获特征空间细节,提高目标边界和局部结构的刻画能力.然而,主流上采样方法存在2个关键问题:一是因信息恢复不充分导致目标边缘模糊;二是因缺乏明确的特征选择机制导致未能聚焦关键区域

特征.

为了解决上述问题,本文从自然物理现象中获得启发,将液体表面张力的扩散-聚焦机制引入语义分割任务中.液体表面张力控制分子间的吸引与排斥,表现2种典型行为:当表面张力较小时,液体趋于扩散,表面变得平滑,这类似于语义分割上采样中的信息恢复过程,通过扩散机制可以增强特征的传递,避免因信息恢复不充分导致的边缘模糊;而当表面张力较大时,液体趋向聚焦,集中于局部区域,增强结构强度,这类似于语义分割中对关键区域增强关注的聚焦阶段,通过聚焦机制可以有效提升对关键区域特征的刻画精度.为此,本文将这一机制引入语义分割任务,液体表面张力与特征图之间的类比关系如图1所示(以水为例).

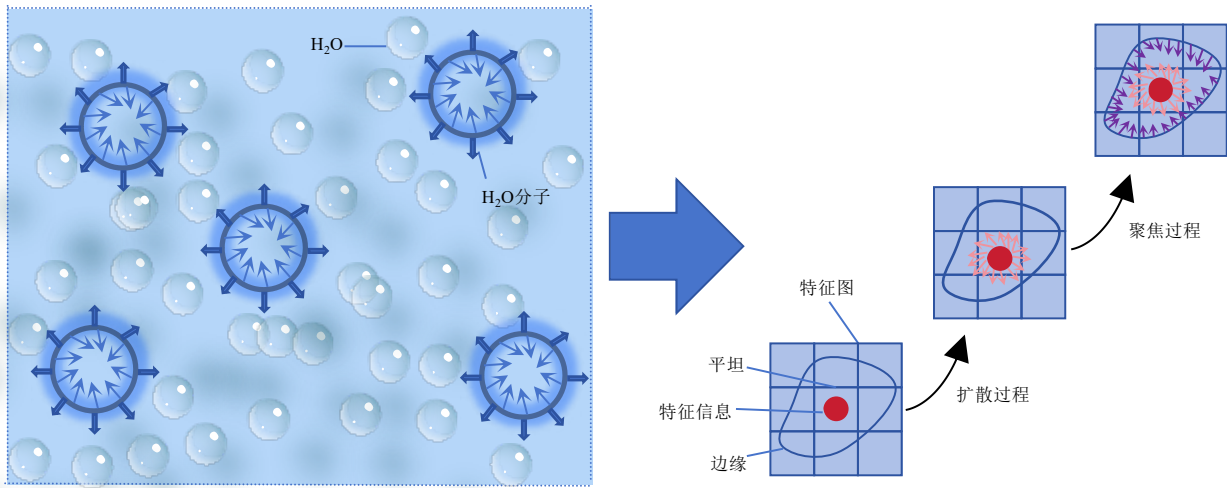


图1 液体表面张力与特征图对应关系

扩散强度由液体表面张力决定,表面张力越小,扩散效果越显著.扩散行为可以用扩散方程来描述,调节扩散强度能够影响特征图的恢复效果,使图像的全局信息更为平滑.液体的扩散可以通过以下方程来表示:

$$\frac{\partial u(x,t)}{\partial t} = D\nabla^2 u(x,t) \quad (1)$$

其中, $u(x,t)$ 是液体在空间位置 x 和时间 t 的状态(浓度、表面高度等); D 是扩散系数,表示液体的扩散能力,直接与表面张力相关; ∇^2 是拉普拉斯算子,表示空间上的扩散.

当表面张力较大时,液体分子间的相互吸引力增强,液体趋向收缩并集中到特定区域.这种效应可以通过液体的曲率来描述,与图像中的目标区域的特征聚焦相似,有助于提升目标区域的分辨率与细节恢复.液体表面的聚焦行为通常由曲率流方程来建模,曲率流方程描述了表面在曲率作用下的演化,具体表示为

$$\frac{\partial u(x,t)}{\partial t} = \kappa(x,t)|\nabla u(x,t)| \quad (2)$$

其中, $\kappa(x,t)$ 是表面曲率,表示液体表面弯曲的程度;

$\nabla u(x,t)$ 是液体表面的位置梯度.

据此,本文引入液体表面张力机制,设计了一种具备扩散控制和聚焦调节能力的上采样策略,以更好地满足语义分割中不同区域的表达需求.扩散机制对应语义分割中的背景特征传播及上下文补充过程.在实际图像中,非关键区域(如背景、均匀纹理区域)虽不具备显著的边界特征,但承载着空间结构和上下文信息.而扩散机制通过调控扩散强度,增强特征在背景中的平滑传播能力,扩大上下文感知范围并增强底层特征表达,为复杂结构区域提供稳定的语义支撑.

聚焦机制则对应语义分割中的边缘强化及目标感知过程.图像中的关键区域(如目标边缘、器官轮廓、精细结构等)通常位于空间上的特定子区域,需要更高的分辨率和关注度来突出其语义价值.若在上采样过程中对所有区域一视同仁,易导致目标边缘模糊.而聚焦机制能够自适应地感知重要区域,并增强其语义表达与边缘清晰度,有效缓解传统方法在细节重建上的能力不足问题.

3 本文方法

3.1 总体框架

U-Net作为一种经典的图像分割架构,广泛应用于医学图像和遥感图像等领域.然而,随着任务复杂性和数据规模的增加,传统U-Net在处理多样化特征和高维数据时面临局限性.针对U-Net的不足,本文引入物理学中的扩散-聚焦机制,提出了一种融合物理扩散-聚焦

机制的语义分割模型DFRNet,在保留经典编码器-解码器架构及跳跃连接的基础上,增加了多尺度级联空间金字塔池化(Multi Scale CAscaded spatial Pyramid Pooling, MSCAPP)、物理启发能量传播上采样(Physics-Inspired Energy Propagation Upsampling, PIEPU)和动态上下文窗口选择(Dynamic context Window Selection, DWS)等创新模块,模型架构如图2所示.

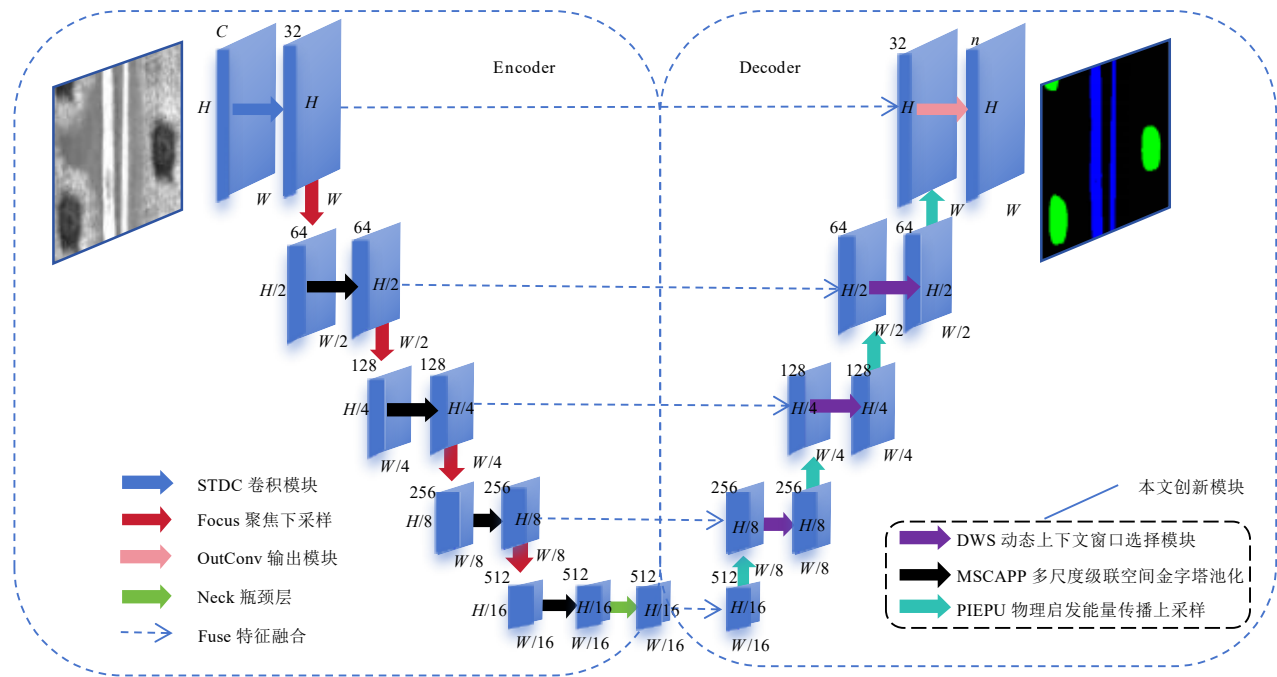


图2 DFRNet模型架构

3.2 MSCAPP模块

MSCAPP模块基于空间注意力和多尺度卷积,集成了多种卷积操作和池化策略,旨在增强特征图的多尺度信息并通过空间注意力机制提升重要区域的特征表示,如图3所示.

首先,输入特征图经过 1×1 卷积降维,随后进行归一化与激活,增强特征图的表达能力.然后执行局部卷积和全局卷积,局部卷积捕捉细粒度的局部信息,全局卷积通过3种不同尺度的卷积核(3×3 、 5×5 、 7×7)捕捉不同范围的全局上下文信息.此外,模块通过级联池化操作,利用不同大小的池化核(5×5 、 9×9 、 13×13)获取多尺度空间信息,增强了模型对不同空间尺度的适应能力.在特征提取阶段完成后,通过拼接将各个尺度的特征图融合,形成一个包含多尺度信息的高维特征图,并通过 1×1 卷积压缩通道数,输出最终的特征图.最后,空间注意力机制对特征图进行加权,强化重要区域,抑制不相关区域,提升关键特征的代表能力.

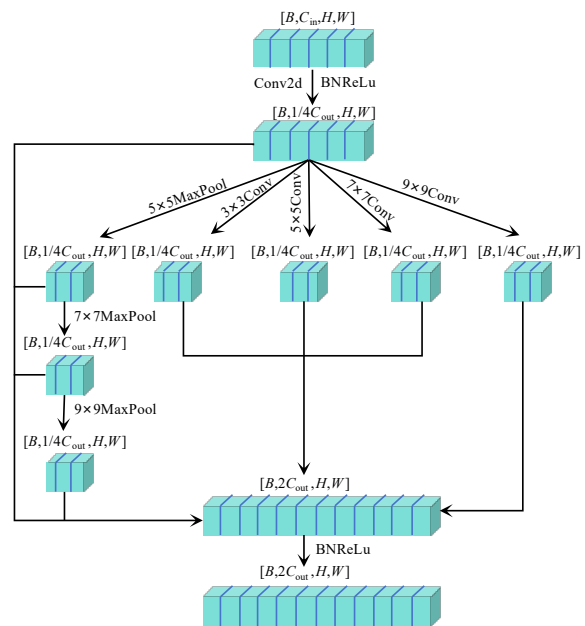


图3 MSCAPP模块

3.3 PIEPU 模块

PIEPU 模块是对经典 U-Net 结构进行改进的上采样单元,旨在提升图像生成和恢复的精度.该模块通过引入扩散和聚焦机制,模型能够更加自适应地调整不同区域的特征图信息,从而增强特征传递有效性和细节恢复能力.具体流程如图 4 所示.

PIEPU 的扩散机制通过迭代局部更新特征图,有效传播信息并增强图像细节恢复.其核心基于图像的局部梯度和二阶导数信息,动态调整每个像素的扩散强度.首先,扩散过程通过计算图像的拉普拉斯算子衡量局部像素值的变化.拉普拉斯算子 $\nabla^2 x$ 描述了像素值在空间上的变化程度,能够有效捕捉图像中的边缘信息,其计算的是像素周围邻域的差异,计算公式如下:

$$\nabla^2 x = \frac{\partial^2 x}{\partial x^2} + \frac{\partial^2 x}{\partial y^2} \quad (3)$$

拉普拉斯算子的计算通过卷积操作完成,拉普拉斯算子近似为

$$\text{Laplace}(x) = \sum_{i=-1}^1 \sum_{j=-1}^1 w(i,j)x(i,j) - 4x(0,0) \quad (4)$$

通过这种方式,拉普拉斯算子提供了图像局部区域的变化信息,作为扩散过程的基础,对应式(1)中的 ∇^2 .

然后通过高斯二阶导数(LoG)来进一步处理图像,LoG通过对图像进行高斯平滑:

$$\text{LoG}(x) = \nabla^2(G(x) * x) \quad (5)$$

其中, $G(x)$ 是高斯滤波器; * 表示卷积操作; x 是输入图像.这一过程有助于去除噪声并增强图像的边缘特征,为后续的扩散过程提供更加准确的梯度信息.

PIEPU 的聚焦机制核心思想是通过梯度变化动态调整扩散过程.具体而言,梯度较大的区域通常代表着边缘或其他重要的结构信息,这些区域需要被保留,因此在这些区域扩散过程会被抑制.相反,梯度较小的区域通常对应平坦区域,这些区域的信息在下采样过程中可能会丢失,因此需要通过聚焦来恢复.该机制依赖于图像的梯度 k_1 和二阶导数 k_2 , 计算聚焦系数 g , 对应式(2)中的 $\kappa(x,t)$, 其计算公式为

$$g(\nabla u, \nabla^2 u) = \exp\left(-\left(\frac{|\nabla u|}{k_1}\right)^2\right) \cdot \sigma\left(\frac{\nabla^2 u}{k_2}\right) \quad (6)$$

其中, $|\nabla u|$ 是图像的梯度幅值, $\nabla^2 u$ 是图像的拉普拉斯算子.通过计算梯度变化,模型能够自动增强梯度较大的区域,平滑和扩展梯度较小的区域,保护了特征图的边缘,同时对低细节区域进行平滑处理,有效减少上采样过程中可能出现的伪影.

扩散和聚焦机制的选择取决于输入特征图在每个位置的重要性的信息变化度,即依据特征图局部的变化情况来判断是否需要扩散以平滑低变化区域,或是否需要聚焦以增强高变化区域的局部特征.在高变化区域倾向于聚焦;在平坦区域或低变化区域,则更适合扩散,平滑不重要的区域.最终得到一个融合的特征图,既保留了细节,又确保了整体的平滑度和结构.

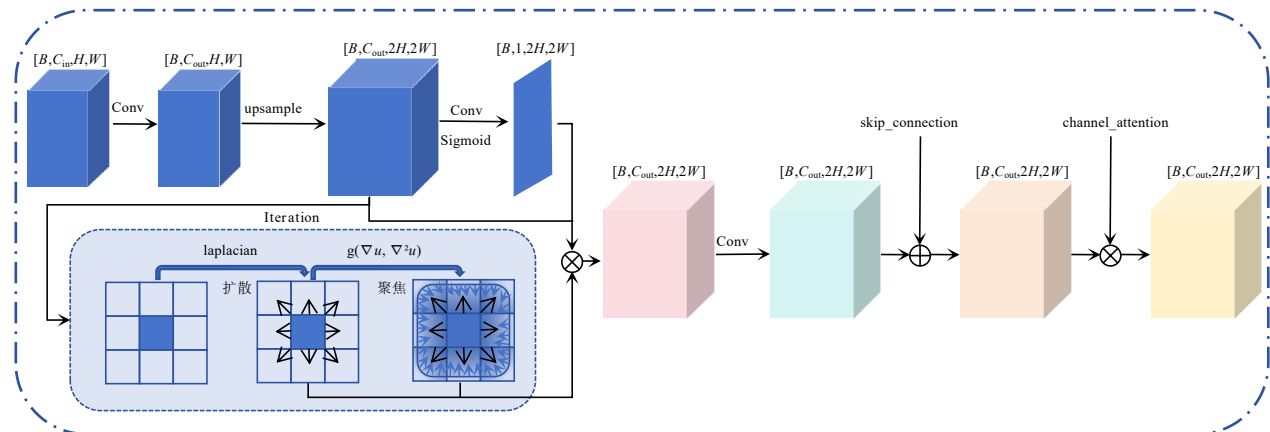


图 4 PIEPU 模块流程

3.4 DWS 模块

DWS 模块通过引入调节机制对扩散-聚焦机制进行动态调节,其核心思想在于根据输入特征的不同情况动态选择使用扩散机制或聚焦机制,并在扩散与聚焦过程中进行相应的调整.该机制包括通道注意力机制、动态窗口大小选择以及稀疏上下文建模 3 个关键部分,旨在动态地选择和调整不同区域的上下文窗口大

小及其重要性,从而有效捕捉图像中的细节信息和上下文关系.具体流程如图 5 所示.

通道注意力机制通过卷积层实现通道级别的动态调整,增强模型对关键信道的关注.模型首先通过卷积操作学习每个通道的重要性,并通过激活函数生成一个通道权重图.这样,模型能够根据输入特征的变化自动调整各个通道的权重,强化关键通道的信息表达,抑

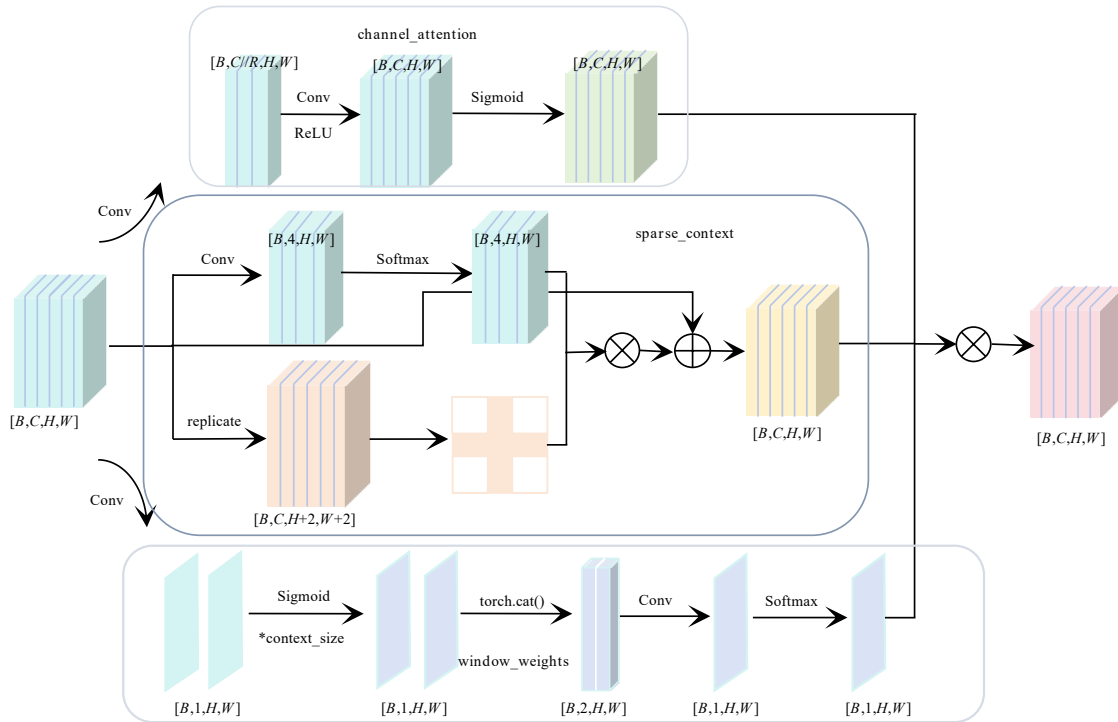


图5 DWS模块流程

制无关通道的影响. 在图像的不同区域, DWS模块引入了2个并行的卷积分支来预测每个空间位置的窗口大小. 这2个分支分别生成一个权重图, 表示图像中每个区域的上下文窗口大小. 通过对这2个分支输出的窗口大小权重图进行聚合, 模型能够动态选择最适合当前区域的上下文窗口大小. 在图像特征的进一步处理过程中, DWS模块生成一个包含4个通道的重要性图, 每个通道分别表示图像中不同区域(左上、左下、右上、右下)的上下文重要性. 这一重要性图通过卷积操作生成, 并结合输入特征图的边界扩展进行加权特征融合. 边界扩展操作通过复制边界像素来处理边缘效应, 确保在处理图像边界时不会丢失关键信息. 加权特征融合则根据重要性图的权重调整不同区域的特征贡献, 使得模型能够更加精细地捕捉图像中的关键区域信息. 通过稀疏上下文建模, 模型根据重要性图对不同区域分配不同的权重, 只关注对最终任务有显著影响的特征区域, 避免不必要的计算浪费, 提升模型的处理速度.

调节机制通过一个动态学习的参数 α 控制扩散过程的速率. 这个参数基于特征图的局部信息进行调整, 使得在特征图中每个像素点的扩散速率与其局部环境相适应. 通过卷积层生成的扩散速率图会随着特征图的更新而变化. 在高变化区域, 调节机制减少扩散速率, 强化边缘信息. 在低变化区域, 扩散速率会适当增加, 确保细节得到平滑处理. 通过调节扩散系数, 调节机制能保留图像细节的同时, 平滑低变化区域, 防止伪

影的产生. 此外, 扩散系数的动态计算使得扩散过程能够自适应地调整, 在不同类型的区域中扩散速率能够得到最优控制.

4 实验分析

4.1 数据集介绍

为了全面评估本文所提出的语义分割模型DFR-Net的泛化能力, 本文使用了包括钢材类、医学类、生物学类、遥感类、道路类、月球探索类及农业类7类, 共计14个数据集进行实验测试. 相比已有研究^[29-38], 本文所使用的数据集在数量和类别覆盖上均更为广泛.

具体而言, 钢材类图像背景纹理复杂, 缺陷区域细小、边缘模糊; 医学类图像区域形态多变、类间相似度高, 存在严重混淆问题; 生物学类任务中目标密集、结构细碎, 易发生粘连或漏检; 遥感类任务中尺度变化剧烈、地物边界复杂, 考验多尺度建模能力; 道路类图像常伴随运动模糊、遮挡与光照变化, 存在目标断裂与边界模糊问题; 月球探索类图像灰度分布单一、对比度低, 边缘信息稀疏; 农业类中植被形态复杂, 类间纹理相似, 干扰因素多.

这些数据集在纹理复杂度、边界清晰度、尺度分布及语义差异性等方面具有广泛覆盖性, 能够有效反映模型在多任务场景下的适应性与鲁棒性, 充分验证DFRNet在应对不同类型语义分割问题上的普适性和稳定性. 各数据集的训练集和验证集划分情况见表1.

表 1 数据集划分情况

数据集类别	数据集名称	Total	Train	Valid	分割类别
钢材类	NEU_Seg ^[36]	4 470	3 630	840	4
	FSSD12 ^[39]	3 368	2 694	674	2
医学类	Synapse Multi-organ Segmentation Dataset ^[30]	2 211	1 769	442	8
	ACDC ^[30]	943	754	189	4
生物学类	2018 Data science bowl ^[32]	670	536	134	2
	BF-C2DL-HSC ^[40]	3 528	2 822	706	2
遥感类	2021LoveDA ^[38]	4 191	3 353	838	8
	WHU Building Dataset - Aerial Imagery Dataset ^[41]	8 188	6 550	1 638	2
道路类	TP-dataset ^[42]	1 391	1 113	278	2
	KITTI Road Detection ^[43]	289	231	58	2
月球探索类	LuSNAR ^[44]	2 196	1 757	439	5
	人工月球景观	5 656	4 525	1 131	4
农业类	PlantSeg v2 ^[45]	7 909	6 326	1 583	2
	2016 Sugar Beets ^[46]	1 321	1 057	264	2

注:人工月球景观数据集网址:<https://www.cvmart.net/dataSets/detail/733>.

4.2 实验设置

本文基于 Pytorch 搭建框架,所有实验都基于 RTX 4090 GPU (24 GB) 和 Python 3.8 实现,学习率为 1×10^{-6} , batch size 为 16, 优化器采用 RMSprop, 在类别像素上采用交叉熵损失函数和 Dice 损失函数. 本文实验共分 2 个部分:(1) 对比实验,为验证本文模型的有效性,将本文模型与 10 个先进模型在各数据集上定量对比;(2) 消融实验,为验证本文模块的有效性,将各模块在各数据集上定量对比. 对比实验的 epoch 设置为 200 轮,消融实验的 epoch 设置为 50 轮. 为了便于比较,本文所有对比实验和消融实验都在相同的训练和测试设置下进行,与多数语义分割方

法采用的指标一致,本文实验共使用了 5 个评价指标,分别是 mIoU (Mean Intersection over union)、 F_1 分数 (F_1)、召回率 (Recall)、精度 (Precision)、准确率 (Accuracy).

4.3 对比实验

为全面验证本文所提出模型在多类别图像语义分割任务中的性能优势,本文进行一系列对比实验. 表 2—8 分别展示了本文模型与其他先进模型在 7 种类别 14 个数据集上的对比实验结果.(评价指标各项的数据均为百分比表示)表中加粗的表示在不同网络模型下的最佳数据,↑ 与 ↓ 分别表示与最佳数据相比较本文所提出改进模型的数据是增长还是下降.

表 2 钢材类数据集上的模型语义分割结果对比

模型	Param/MB	NEU_Seg/%					FSSD12/%				
		mIoU ↑	F_1 ↑	Recall ↑	Precision ↑	Acc ↑	mIoU ↑	F_1 ↑	Recall ↑	Precision ↓	Acc ↑
U-Net ^[9]	7.76	82.332	90.232	89.529	90.963	90.593	79.591	88.636	84.976	92.624	84.976
DeepLab-V3 ^[12]	54.60	85.032	91.807	91.974	91.646	98.280	80.385	89.126	87.411	90.910	96.949
SegNet ^[15]	29.45	82.167	90.131	90.711	89.562	97.856	80.628	89.275	87.257	91.389	97.001
FCN ^[8]	32.95	84.369	91.442	91.988	90.919	98.246	81.042	89.528	88.804	90.265	97.028
PSPNet ^[17]	53.32	85.328	92.027	92.113	91.948	98.350	82.082	90.159	89.443	90.887	97.208
MEWUNet ^[47]	140.27	84.005	91.253	91.008	91.509	98.099	81.426	89.762	88.189	91.393	97.122
UNetV2 ^[35]	25.15	78.571	87.855	87.907	87.828	97.427	76.630	86.769	85.108	88.495	96.287
PIDNet-S ^[48]	7.62	82.549	90.356	89.802	90.950	97.989	80.074	88.935	88.294	89.584	96.857
DDRNet-23-S ^[49]	5.70	82.286	90.201	89.518	90.924	97.963	79.568	88.622	87.271	90.015	96.794
ACC-UNet ^[50]	4.26	83.976	91.234	91.252	91.227	98.111	77.404	87.263	84.785	89.891	96.460
Ours	3.34	87.365	93.203	92.672	93.756	98.595	82.598	90.470	89.480	92.353	97.303

综上所述,本文所提出的 DFRNet 在上述 7 类任务的评价指标中大部分均表现为提升,表明其具备良好的跨领域泛化能力. 与大多数方法仅在 2-3 类任务中验

证的做法不同,DFRNet 在更多类别任务中都获得稳定改进,凸显了模型在异质视觉特征下的鲁棒性. 由于本模型采用深度可分离卷积替代 U-Net 中的标准卷积,

表3 医学类数据集上的模型语义分割结果对比

模型	Param/MB	Synapse Multi-organ Segmentation/%					ACDC/%				
		mIou \uparrow	F_1 \uparrow	Recall \uparrow	Precision \uparrow	Acc \uparrow	mIou \uparrow	F_1 \uparrow	Recall \uparrow	Precision \uparrow	Acc \uparrow
U-Net ^[9]	7.76	78.185	87.010	86.185	87.979	94.050	84.409	91.399	90.564	92.269	91.103
DeepLab-V3 ^[12]	54.60	83.675	90.744	89.116	92.514	99.699	85.699	92.181	92.141	92.223	99.477
SegNet ^[15]	29.45	72.771	82.457	78.171	88.921	99.557	83.784	91.018	90.711	91.341	99.400
FCN ^[8]	32.95	82.547	90.093	90.125	90.115	99.649	85.363	91.978	91.218	92.764	99.481
PSPNet ^[17]	53.32	83.074	90.300	90.709	89.926	99.687	86.421	92.613	93.614	91.681	99.503
MEWUNet ^[47]	140.27	81.746	89.362	87.621	91.287	99.683	81.635	89.712	90.436	89.030	99.313
UNetV2 ^[35]	25.15	68.226	79.596	77.773	81.682	99.415	49.026	64.794	60.922	69.349	97.660
PIDNet-S ^[48]	7.62	80.201	88.494	89.372	87.762	99.623	84.213	91.304	91.494	91.132	99.422
DDRNet-23-S ^[49]	5.70	79.584	88.057	87.510	88.684	99.620	83.481	90.864	90.224	91.523	99.388
ACC-UNet ^[50]	4.26	82.328	89.908	90.519	89.340	99.681	74.068	84.719	84.115	85.484	99.047
Ours	3.34	87.934	93.402	92.854	93.988	99.766	86.974	92.952	93.949	92.993	99.547

表4 生物学类数据集上的模型语义分割结果对比

模型	Param/MB	2018 Data science bowl/%					BF-C2DL-HSC/%				
		mIou \uparrow	F_1 \uparrow	Recall \uparrow	Precision \uparrow	Acc \uparrow	mIou \uparrow	F_1 \uparrow	Recall \uparrow	Precision \downarrow	Acc \uparrow
U-Net ^[9]	7.76	86.693	92.872	92.247	93.506	92.247	90.213	94.854	95.644	94.078	95.644
DeepLab-V3 ^[12]	54.60	84.041	91.328	90.705	91.960	90.705	77.580	87.375	92.336	82.919	92.336
SegNet ^[15]	29.45	82.195	90.227	92.546	88.022	92.546	75.795	86.231	88.246	84.305	88.246
FCN ^[8]	32.95	75.691	86.164	85.174	87.177	85.174	65.167	78.910	82.657	75.489	82.657
PSPNet ^[17]	53.32	79.325	88.470	88.649	88.293	96.428	78.829	88.161	93.207	83.634	93.207
MEWUNet ^[47]	140.27	84.416	91.550	90.434	92.693	90.434	81.604	89.870	91.686	88.124	91.686
UNetV2 ^[35]	25.15	83.281	90.878	90.404	91.357	90.404	75.398	85.973	90.496	81.881	90.496
PIDNet-S ^[48]	7.62	74.396	85.318	84.129	86.542	84.129	64.568	78.470	80.817	76.255	80.817
DDRNet-23-S ^[49]	5.70	74.826	85.601	85.311	85.893	85.311	64.918	78.727	81.949	75.750	81.949
ACC-UNet ^[50]	4.26	85.668	92.281	91.183	93.405	91.183	90.401	94.958	95.396	94.524	95.396
Ours	3.34	87.891	94.093	92.720	93.845	96.720	90.654	95.098	96.423	93.809	96.423

表5 遥感类数据集上的模型语义分割结果对比

模型	Param/MB	2021LoveDA/%					WHU Building Dataset - Aerial Imagery/%				
		mIou \uparrow	F_1 \uparrow	Recall \uparrow	Precision \uparrow	Acc \uparrow	mIou \uparrow	F_1 \uparrow	Recall \uparrow	Precision \uparrow	Acc \uparrow
U-Net ^[9]	7.76	60.075	75.048	74.930	75.306	77.270	87.198	93.161	92.358	93.979	92.358
DeepLab-V3 ^[12]	54.60	56.795	71.734	70.954	72.904	75.963	84.569	91.640	90.302	93.017	90.302
SegNet ^[15]	29.45	53.695	68.886	66.940	71.413	73.662	82.176	90.216	88.377	92.134	88.377
FCN ^[8]	32.95	59.615	74.199	73.085	75.495	77.857	79.895	88.824	88.240	89.416	88.240
PSPNet ^[17]	53.32	62.960	76.858	76.617	77.149	79.771	85.364	92.104	91.536	92.679	91.536
MEWUNet ^[47]	140.27	57.550	72.431	71.641	73.506	75.497	86.524	92.775	91.914	93.653	91.914
UNetV2 ^[35]	25.15	54.019	69.291	68.379	70.562	72.914	82.545	90.438	88.942	91.986	88.942
PIDNet-S ^[48]	7.62	52.718	68.088	65.228	71.921	71.574	78.535	87.977	87.344	88.619	87.344
DDRNet-23-S ^[49]	5.70	55.492	70.630	68.405	73.132	72.765	78.837	87.664	86.442	88.922	86.442
ACC-UNet ^[50]	4.26	56.548	71.530	69.611	73.897	74.524	87.695	93.444	92.959	93.934	92.959
Ours	3.34	63.125	79.746	78.856	79.254	79.931	89.363	94.383	93.938	94.832	93.938

Fuse 特征融合替代跳跃连接,以及 PIEPU 上采样替代转置卷积及 DWS 模块替代固定卷积窗口等优化策略,所以本文模型参数量仅为 3.34 MB,远低于其他现有模型.

具体地, mIou 是衡量模型全局分割性能的核心指标,其提升表明本文模型在捕捉全局上下文信息方面具有显著优势. 在 mIou 这一关键指标上,本文模型在医学类 Synapse Multi-organ Segmentation 数据集的 mIou

表 6 道路类数据集上的模型语义分割结果对比

单位: %

模型	Param/MB	TP-dataset/%					KITTI Road Detection/%				
		mIoU \uparrow	F_1 \uparrow	Recall \downarrow	Precision \uparrow	Acc \uparrow	mIoU \uparrow	F_1 \uparrow	Recall \uparrow	Precision \uparrow	Acc \uparrow
U-Net ^[9]	7.76	85.051	91.922	92.204	91.642	92.204	81.700	89.928	90.670	89.199	90.670
DeepLab-V3 ^[12]	54.60	82.965	90.689	87.485	94.137	87.485	87.929	93.577	94.683	92.496	94.683
SegNet ^[15]	29.45	60.298	75.233	65.534	88.300	65.534	84.099	91.363	90.121	92.639	90.121
FCN ^[8]	32.95	78.171	87.748	85.164	90.494	85.164	88.013	93.624	93.940	93.311	93.940
PSPNet ^[17]	53.32	90.478	95.001	94.181	95.836	94.181	89.683	94.561	94.416	94.707	94.416
MEWUNet ^[47]	140.27	80.973	89.486	90.210	88.774	90.210	85.549	92.212	94.327	90.190	94.327
UNetV2 ^[35]	25.15	68.649	81.410	76.506	86.986	76.506	76.949	86.973	87.387	86.563	87.387
PIDNet-S ^[48]	7.62	79.847	88.794	87.149	90.503	87.149	84.330	91.499	89.801	93.262	89.801
DDRNet-23-S ^[49]	5.70	80.473	89.180	86.265	92.299	86.265	85.460	92.160	92.467	91.855	92.467
ACC-UNet ^[50]	4.26	86.041	92.497	91.336	93.688	91.336	85.970	92.456	92.403	92.508	92.403
Ours	3.34	91.383	95.951	93.587	96.794	95.567	90.290	94.897	94.926	94.869	94.926

表 7 月球探索类数据集上的模型语义分割结果对比

模型	Param/MB	LuSNAR/%					人工月球景观/%				
		mIoU \uparrow	F_1 \uparrow	Recall \uparrow	Precision \uparrow	Acc \uparrow	mIoU \uparrow	F_1 \uparrow	Recall \uparrow	Precision \uparrow	Acc \uparrow
U-Net ^[9]	7.76	92.590	96.071	95.534	96.627	98.592	70.954	81.883	77.564	87.219	92.248
DeepLab-V3 ^[12]	54.60	93.733	96.705	96.495	96.920	98.871	68.498	79.829	74.801	86.278	91.422
SegNet ^[15]	29.45	86.860	92.717	91.746	93.745	98.846	63.047	75.079	69.839	82.174	96.393
FCN ^[8]	32.95	93.656	96.653	96.196	97.122	98.766	66.967	78.615	73.653	85.200	91.084
PSPNet ^[17]	53.32	94.243	96.989	97.135	96.837	99.008	70.576	81.502	76.631	87.671	92.100
MEWUNet ^[47]	140.27	92.986	96.290	95.467	97.148	99.359	69.556	80.699	76.176	86.327	97.034
UNetV2 ^[35]	25.15	87.631	93.179	92.257	94.127	98.904	59.641	71.858	66.691	79.346	95.997
PIDNet-S ^[48]	7.62	91.836	95.631	94.811	96.495	98.300	67.373	78.915	74.331	84.764	91.351
DDRNet-23-S ^[49]	5.70	91.618	95.511	95.468	95.561	98.360	66.901	78.541	73.539	85.030	91.035
ACC-UNet ^[50]	4.26	95.050	97.430	97.435	97.427	99.547	69.756	80.838	76.602	85.987	97.095
Ours	3.34	96.599	98.257	98.005	98.512	99.686	73.512	82.986	80.569	88.127	97.356

表 8 农业类数据集上的模型语义分割结果对比

模型	Param/%	PlantSeg v2/%					2016 Sugar Beets/%				
		mIoU \uparrow	F_1 \uparrow	Recall \uparrow	Precision \uparrow	Acc \uparrow	mIoU \uparrow	F_1 \uparrow	Recall \uparrow	Precision \uparrow	Acc \uparrow
U-Net ^[9]	7.76	53.221	69.469	68.656	70.301	68.656	86.951	93.020	92.981	93.059	92.981
DeepLab-V3 ^[12]	54.60	57.436	72.964	72.739	73.192	88.520	75.956	86.335	84.732	88.000	84.732
SegNet ^[15]	29.45	55.209	71.141	68.603	73.875	88.147	75.870	86.280	86.001	86.561	97.671
FCN ^[8]	32.95	58.316	73.671	72.888	74.470	88.905	76.043	86.391	87.816	85.012	87.816
PSPNet ^[17]	53.32	61.391	76.077	76.184	75.971	89.796	80.100	88.950	86.977	91.015	86.977
MEWUNet ^[47]	140.27	59.560	74.655	74.976	74.337	89.158	84.271	91.464	90.459	92.492	98.562
UNetV2 ^[35]	25.15	55.843	71.666	70.671	72.689	88.099	81.462	89.784	89.029	90.552	98.275
PIDNet-S ^[48]	7.62	55.504	71.386	72.179	70.610	87.677	76.266	86.535	85.978	87.100	85.978
DDRNet-23-S ^[49]	5.70	53.449	69.664	69.893	69.437	87.036	76.038	86.388	84.833	88.001	84.833
ACC-UNet ^[50]	4.26	58.156	73.542	71.827	75.342	88.994	85.222	92.021	92.860	91.198	98.629
Ours	3.34	62.895	78.163	77.912	79.737	89.916	87.789	93.497	92.978	94.734	98.902

提升最大,相较于其他取得最佳性能的先进模型提升了4.259%。 F_1 分数的提升表明模型在精度和召回率之间取得了更好的平衡,能够更准确地分割细节区域. 本文模型在 2021LoveDA 数据集和 Synapse Multi-organ

Segmentation 数据集提升较大,分别为 2.888% 和 2.658%。Precision 和 Accuracy 的提升表明模型更能准确地分割目标区域.

为进一步验证本文模型在分割精度和细节处理上

的优越性,将本文模型和基线模型(U-Net)在各数据集上的一些分割结果实例展示如图6所示,每一行分别代表每一类数据集,用虚线框突出明显区别区域.从图中可以直观地看出,基线模型存在全局信息丢失、过分割和分类错误等问题,如第1行、第4行和第6行的左边虚

线区域显示,缺陷中间都出现了空缺,说明基线模型在低分辨率特征恢复过程中出现全局信息缺失;第1行右边和第2行左边区域中本应为背景的区域被错误分类为目标类别;第2行和第7行右边部分存在有不同类别没有被分割.

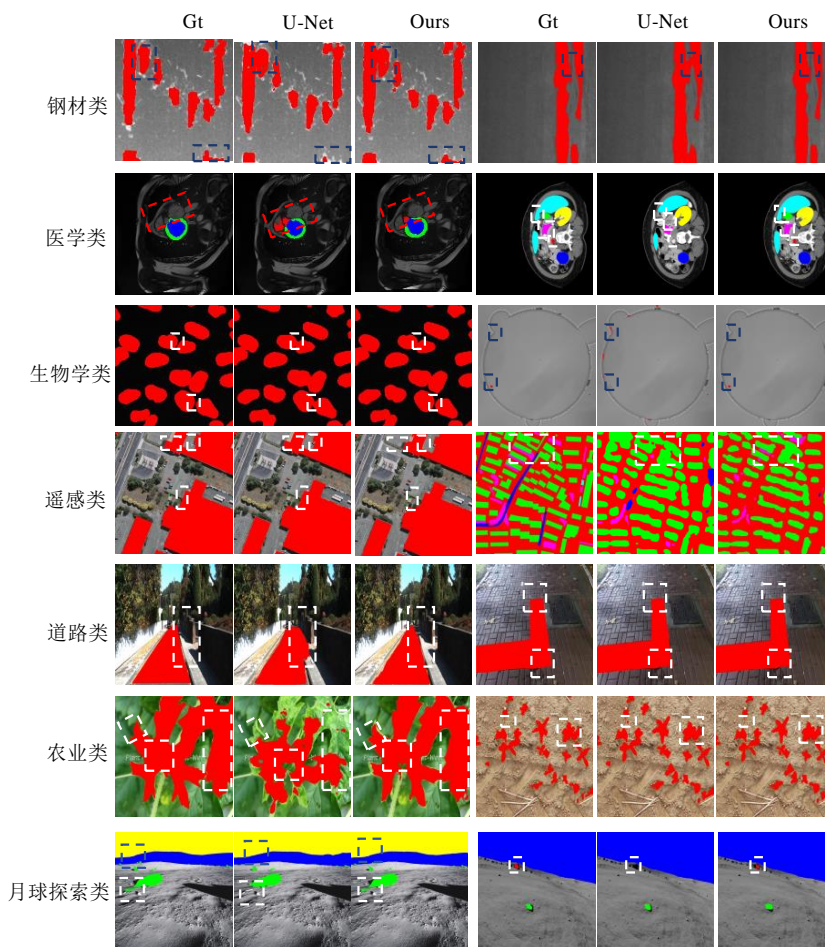


图6 与基线模型对比的可视化分割结果

相比之下,本文提出的方法在上述关键区域均表现出更准确且稳定的分割效果:目标边界更加清晰,结构保持良好,误分割与漏分割现象显著减少,尤其在小目标、边缘过渡和结构复杂区域上表现突出.然而,DFRNet所展现出的性能提升并非偶然,而是得益于其所引入的三大核心机制.

扩散机制在模型中主要承担语义信息的长距离传递任务,尤其是在浅层与高层特征交汇处,建立起跨尺度的连续传播路径.这种设计强化了浅层特征的语义一致性,避免了语义断裂与结构偏移(如图6中钢材类、医学类和农业类所示),使得模型在处理场景复杂度高、结构分布离散的图像时具备更强的全局感知能力.

聚焦机制则在关键区域的信息提炼方面发挥关键

作用.该机制通过对显著区域响应的增强,提升了模型在物体边缘、小目标等细粒度结构的建模能力(如图6中生物学类和月球探索类所示),进一步缓解了传统方法中存在的特征模糊与边界不连贯问题,特别在图像细节重建阶段形成有效补偿.

调节机制作为连接扩散与聚焦的“调控中枢”,根据特征图的局部变化动态调整扩散和聚焦.它有效防止了单一机制在特定区域的过度介入(如图6中遥感类和道路类所示),通过区域间自适应差异处理,提升了特征融合过程的稳定性与效率,也显著降低了上采样过程中因策略僵化导致的伪影与信息冗余现象.

4.4 消融实验

为了全面评估本文所提出的三大核心机制模块在不同图像语义分割任务中的独立有效性与协同能力,

本文通过对模型中扩散机制、聚焦机制与调节机制的逐一去除与组合测试,开展消融实验.表9~11分别展示扩散、聚焦、调节独立模块或结合模块在7种类别14个数据集上的消融实验结果.表中“√”表示该模块

存在,“×”表示该模块不存在.

参考已有研究^[36-38,51-54]的做法,其消融实验普遍采用2种评价指标,因此本文在消融实验中亦选取 mIoU 和 Accuracy 作为主要评估指标.

表9 钢材类、医学类、生物学类的消融实验结果

单位:%

模块			NEU_Seg		FSSD12		Synapse		ACDC		science bowl		BF-C2DL-HSC	
扩散	聚焦	调节	mIoU	Acc	mIoU	Acc	mIoU	Acc	mIoU	Acc	mIoU	Acc	mIoU	Acc
√	√	√	82.168	90.790	78.569	87.292	80.855	95.118	83.416	91.259	86.108	92.384	90.822	96.466
×	√	√	82.068	90.301	77.781	85.405	79.085	94.122	82.393	90.125	86.028	91.728	90.631	96.361
√	×	√	81.935	90.226	77.843	85.353	77.091	91.572	79.131	90.692	85.851	92.186	90.389	96.397
√	√	×	82.244	90.213	77.998	86.092	77.605	93.315	79.698	87.604	85.810	91.982	90.446	96.376
×	×	√	81.746	90.106	77.438	85.148	77.049	91.238	76.327	87.204	85.683	91.601	90.342	96.359
×	×	×	81.554	90.100	76.637	84.491	68.332	91.200	75.175	86.934	85.391	91.113	90.295	96.263

表10 遥感类、道路类的消融实验结果

单位:%

模块			2021LoveDA		WHU		TP-dataset		KITTI	
扩散	聚焦	调节	mIoU	Acc	mIoU	Acc	mIoU	Acc	mIoU	Acc
√	√	√	54.668	74.737	86.593	92.829	76.872	83.936	86.109	92.384
×	√	√	53.918	72.654	85.527	91.668	72.576	83.801	83.971	91.728
√	×	√	53.146	70.800	85.797	92.171	69.120	79.210	82.546	92.186
√	√	×	54.009	72.823	86.336	92.297	72.284	82.172	84.045	91.982
×	×	√	53.143	70.790	85.265	91.455	67.544	76.287	80.461	91.601
×	×	×	52.812	70.105	84.799	91.143	65.614	75.430	79.899	91.113

表11 月球探索类、农业类的消融实验结果

单位:%

模块			LuSNAR		人工月球景观		PlantSeg v2		2016 Sugar Beets	
扩散	聚焦	调节	mIoU	Acc	mIoU	Acc	mIoU	Acc	mIoU	Acc
√	√	√	93.933	98.944	68.452	91.674	56.152	72.841	87.783	83.936
×	√	√	93.609	98.804	68.171	91.628	55.336	72.615	87.018	83.801
√	×	√	93.718	98.786	68.109	91.403	55.120	70.431	86.135	79.210
√	√	×	93.891	98.855	67.524	90.869	55.390	70.293	87.526	82.172
×	×	√	92.595	98.657	67.447	90.512	55.101	70.123	86.122	76.287
×	×	×	92.567	98.393	67.212	90.203	54.633	69.340	85.253	75.430

根据消融实验结果可知,扩散模块对各数据集的 mIoU 和 Accuracy 平均分别提升了 1.028% 和 0.832%, 聚焦模块对各数据集的 mIoU 和 Accuracy 平均分别提升了 1.898% 和 1.822%, 调节模块对各数据集的 mIoU 和 Accuracy 平均分别提升了 1.264% 和 1.183%. 当扩散与聚焦机制联合使用时,两项指标分别平均提升了 2.588% 和 2.538%;当三大机制同时引入时,性能进一步提升,对各数据集 mIoU 和 Accuracy 的平均提升幅度达到 3.738% 和 2.901%. 上述结果表明,各机制在不同任务背景下均发挥了显著作用,且在联合配置下展现出更强的协同增益效果.

进一步观察发现,扩散、聚焦和调节3个模块在图像分割任务中彼此相辅相成、相互依赖.每个模块的效能能在与其他模块的结合中得到了更充分的体现.除钢材类 NEU_Seg 数据集去掉调节模块比3个模块均存在的

mIoU 值稍高外,单独去除扩散或聚焦模块均导致性能下降,而当2个模块同时被移除时,性能损失更为显著.

为更直观地体现各模块的独立贡献与对特定视觉场景的适应性,本文进一步对总模型中移除单个模块的结果进行了可视化分析,对各类别数据集的典型分割样例进行了展示,如图7所示.图7中的“一扩散”表示在本文模型中移除扩散模块后的模型,“一聚焦”与“一调节”亦分别表示去除对应模块后的模型表现.

由图7可知,钢材类和医学类数据移除扩散模块后,部分目标类别未能完整识别或出现语义中断现象.尤其在复杂背景或远距离目标区域,模型未能形成稳定的全局感知能力.相较于移除扩散模块的分割结果,保留扩散模块显著增强了特征图的上下文表达能力,有效解决了语义结构中断与全局信息丢失问题.生物学类和道路类移除聚焦模块后,边缘轮廓呈现模糊、粘

连现象,目标类别之间过渡不清晰.这一现象表明聚焦机制在边缘感知中发挥着关键作用.聚焦机制通过引入显著性选择模块,强化了关键区域特征响应,使模型在边缘判别上具备更高精度,针对性地缓解了类别边界不明确与相邻区域融合问题.遥感类、农业类和月球探索类移除调节模块后,存在部分区域被过分激活或

部分忽略,导致出现冗余信息或漏分割问题,这表明调节机制在控制特征激活与抑制中发挥了核心作用.调节机制通过对扩散与聚焦路径的信息进行动态融合,平衡了浅层细节与深层语义之间的表达权重,使模型在保持语义完整性的同时减少了无效区域的激活,显著改善了信息表达失衡问题.

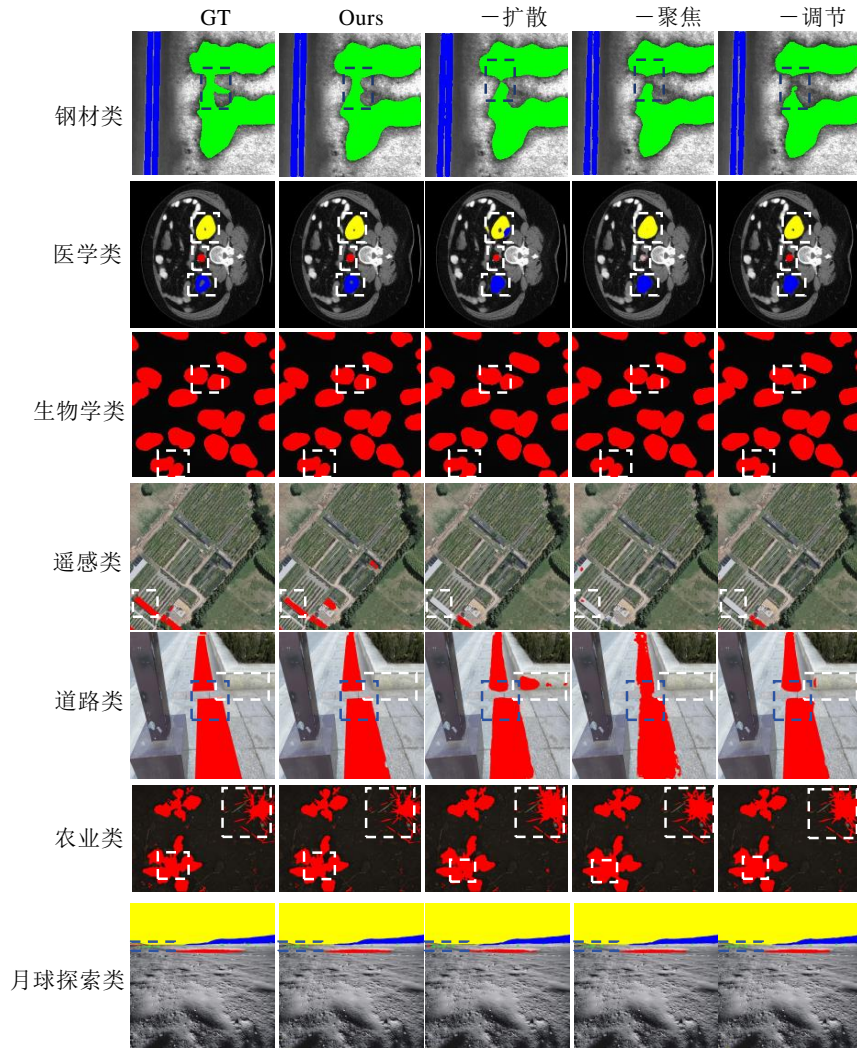


图7 各模块对比的可视化分割结果

4.5 实验结果分析

综合分析对比实验结果、消融实验结果以及其相对应的可视化结果,本文进一步验证了方法设计在理论上的合理性与在实践中的有效性,得出以下结论.

(1)扩散机制针对性地缓解了全局信息丢失问题.从表2~8及表9~11的整体性能指标来看,本文模型在mIoU指标上相较基线模型均实现了稳定提升.图6、7中的可视化结果进一步显示,多个类别的中间平滑区域得到了更完整的识别,目标边界完整性显著增强,充分表明本文设计的扩散机制通过类液体张力引导的上

下文扩散过程,有效增强了特征图中的长距离依赖建模能力,缓解了语义分割中常见的全局信息获取不足问题.

(2)聚焦机制针对性地缓解了边界模糊问题.由表2~8和表9~11数据结果可知,本文模型Accuracy评价指标值均得到了提升,聚焦模块对本文模型的贡献最大.图6与图7中本文模型显示的过分割区域减少、边缘轮廓更加清晰的现象,印证了聚焦机制通过对目标区域的强化与特征加权,在视觉显著区域与边界区域中展现出良好的分割能力,表现出对细节信息与空

间结构的强响应,进而有效缓解了由于过度平滑引起的细节信息丢失与边界模糊问题,提升了模型对小目标与边界结构的感知与定位能力。

(3)调节机制针对性地缓解了生成过程不稳定和信息冗余问题。由图7可视化结果可知,在去掉调节机制的情况下,模型在多个场景下表现出明显的不稳定特征,如边界不连续、细节表达杂乱等。调节机制通过引导网络根据语义复杂度自适应调节特征交互方式,显著提升了信息融合的结构稳定性与目标一致性,有效缓解了生成过程中因信息冲突引起的不稳定问题和信息冗余问题,在整体结构感知与细节保留之间实现更优的平衡。

5 结语

为了尽可能地恢复图像语义分割在下采样过程中丢失的细节信息,本文提出融合物理特性扩散-聚焦机制的语义分割模型DFRNet。该模型通过引入了扩散-聚焦机制的PIEPU和调节机制的DWS设计,在神经网络中特征图的重建过程中实现了动态调节的信息平滑扩展与关键区域聚焦,从而能够有效地提取图像表面的微小特征,提高分割精度。通过对比实验可知,该模型在7种类别14个数据集上均取得了较好的效果,所耗参数量显著下降到3.34 MB,在 $mIoU$ 、 F_1 分数和Accuracy指标上均领先于其他先进模型,在个别数据集上的Recall和Precision稍低于最佳数据,不超过0.716%。通过消融实验分析验证,所提方法能够很好地解决相应的问题。总的来说,本文方法在各场景下的语义分割任务中都展现出了更出色的性能。

参考文献

- [1] HE H Y, CAI J F, PAN Z Z, et al. Dynamic focus-aware positional queries for semantic segmentation[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2023: 11299-11308.
- [2] NITISH V, ACHARYA P G, KRISHNAPRASAD B S, et al. Design and evaluation of a real-time semantic segmentation system for autonomous driving[C]//2024 IEEE 3rd International Conference for Innovation in Technology(INOCON). Piscataway: IEEE, 2024: 1-6.
- [3] YAN J D, SHENG Y, PIAO M H. Semantic segmentation-based wafer map mixed-type defect pattern recognition[J]. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 2023, 42(11): 4065-4074.
- [4] LUCKE K, VAKANSKI A, XIAN M. A2DMN: Anatomy-aware dilated multiscale network for breast ultrasound semantic segmentation[C]//2024 IEEE International Symposium on Biomedical Imaging. Piscataway: IEEE, 2024: 1-5.
- [5] TARRY J, DONG X S, LI X F, et al. Unsupervised ensemble semantic segmentation for foreground-background separation on satellite image[C]//2024 IEEE 18th International Conference on Semantic Computing. Piscataway: IEEE, 2024: 212-217.
- [6] 杨潇, 陈伟, 任鹏, 等. 基于域适应的煤矿环境监测图像语义分割[J]. 煤炭学报, 2021, 46(10): 3386-3396.
YANG X, CHEN W, REN P, et al. Coal mine monitoring image semantic segmentation based on domain adaptation[J]. Journal of China Coal Society, 2021, 46(10): 3386-3396. (in Chinese)
- [7] 安喆, 徐熙平, 杨进华, 等. 结合图像语义分割的增强现实型平视显示系统设计与研究[J]. 光学学报, 2018, 38(7): 85-91.
AN Z, XU X P, YANG J H, et al. Design of augmented reality head-up display system based on image semantic segmentation[J]. Acta Optica Sinica, 2018, 38(7): 85-91. (in Chinese)
- [8] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2015: 3431-3440.
- [9] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional networks for biomedical image segmentation[C]//Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015. Cham: Springer, 2015: 234-241.
- [10] ZHOU Z W, SIDDIQUEE M M R, TAJBAKSH N, et al. UNet++: A nested U-Net architecture for medical image segmentation[EB/OL]. (2018-07-18) [2025-03-12]. <https://arXiv.org/abs/1807.10165>.
- [11] 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. Piscataway: IEEE, 2020: 1055-1059.
- [12] CHEN L C, PAPANDREOU G, SCHROFF F, et al. Rethinking atrous convolution for semantic image segmentation[EB/OL]. (2017-12-05) [2025-03-12]. <https://arXiv.org/abs/1706.05587>.
- [13] CHEN L C, ZHU Y K, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]//Computer Vision-ECCV 2018. Cham: Springer, 2018: 833-851.
- [14] LIU J Q, WANG Z L, CHENG K X. An improved algorithm for semantic segmentation of remote sensing imag-

- es based on DeepLabv3+[C]//Proceedings of the 5th International Conference on Communication and Information Processing. New York: ACM, 2019: 124-128.
- [15] BADRINARAYANAN V, KENDALL A, CIPOLLA R. SegNet: A deep convolutional encoder-decoder architecture for image segmentation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(12): 2481-2495.
- [16] ISOBE S, ARAI S. Deep convolutional encoder-decoder network with model uncertainty for semantic segmentation[C]//2017 IEEE International Conference on Innovations in Intelligent Systems and Applications (INISTA). Piscataway: IEEE, 2017: 365-370.
- [17] ZHAO H S, SHI J P, QI X J, et al. Pyramid scene parsing network[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 6230-6239.
- [18] YANG S, LI J B, LI Y, et al. Imbalanced segmentation for abnormal cotton fiber based on GAN and multiscale residual U-Net[J]. *Alexandria Engineering Journal*, 2024, 106: 25-41.
- [19] 许国良, 毛骄. 基于协同注意力的小样本的手机屏幕缺陷分割[J]. *电子与信息学报*, 2022, 44(4): 1476-1483.
XU G L, MAO J. Few-shot segmentation on mobile phone screen defect based on co-attention[J]. *Journal of Electronics & Information Technology*, 2022, 44(4): 1476-1483. (in Chinese)
- [20] LI Z H, WEI C. Dual dense upsampling convolution for road scene semantic segmentation[C]//2024 5th International Conference on Computer Engineering and Application. Piscataway: IEEE, 2024: 721-726.
- [21] ZHOU R X, ZHANG R Q, CHEN Y. Enhanced semantic segmentation with hierarchical upsampling and CBAM attention mechanism[C]//2024 5th International Conference on Computer Vision, Image and Deep Learning. Piscataway: IEEE, 2024: 830-834.
- [22] XUN S Y, ZHANG Y, DUAN S X, et al. ARG-UNet: Advanced U-Net segmentation model using residual grouped convolution and attention mechanism for brain tumor MRI image segmentation[J]. *Virtual Reality & Intelligent Hardware*, 2024, 6(3): 203-216.
- [23] 徐亮亮, 马开森, 王霞, 等. LA-UNet网络模型在城市绿地遥感分类中的应用[J]. *应用生态学报*, 2024, 35(4): 1101-1111.
XU L L, MA K S, WANG X, et al. Application of LA-UNet network model in remote sensing classification of urban green space[J]. *Chinese Journal of Applied Ecology*, 2024, 35(4): 1101-1111. (in Chinese)
- [24] MAGNUSSEN HELGESEN S E, NAKASHIMA K, TØRRESEN J, et al. Fast LiDAR upsampling using conditional diffusion models[C]//2024 33rd IEEE International Conference on Robot and Human Interactive Communication. Piscataway: IEEE, 2024: 272-277.
- [25] XIA B, ZHAN B, SHEN M, et al. Explicit implicit priori knowledge-based diffusion model for generative medical image segmentation[J]. *Knowledge-Based Systems*, 2024, 11(303): 234-241.
- [26] CHEN J L, LI G Y, ZHANG Z J, et al. EFDCNet: Encoding fusion and decoding correction network for RGB-D indoor semantic segmentation[J]. *Image and Vision Computing*, 2024, 142. DOI: 10.1016/j.imavis.2023.104892.
- [27] 张银胜, 吉茹, 童俊毅, 等. 基于双模态高效特征学习的高分辨率遥感图像分割[J]. *遥感学报*, 2024, 28(2): 481-493.
ZHANG Y S, JI R, TONG J Y, et al. High resolution remote sensing image segmentation based on dual-modal efficient feature learning[J]. *National Remote Sensing Bulletin*, 2024, 28(2): 481-493. (in Chinese)
- [28] 张大锦, 刘辉, 陈甫刚, 等. 频域多方向 C-UNet 及动态损失的工业烟尘图像分割[J]. *控制理论与应用*, 2024, 41(3): 543-554.
ZHANG D J, LIU H, CHEN F G, et al. Industrial smoke image segmentation based on frequency domain multi-directional C-UNet and dynamic loss[J]. *Control Theory & Applications*, 2024, 41(3): 543-554. (in Chinese)
- [29] CHANG J, HE X H, SONG D J, et al. A Multi-Scale attention network for building extraction from high-resolution remote sensing images[J]. *Scientific Reports*, 2025, 15. DOI: 10.1038/s41598-025-09086-9.
- [30] WANG H Y, XIE S, LIN L F, et al. Mixed transformer U-Net for medical image segmentation[C]//ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing. Piscataway: IEEE, 2022: 2390-2394.
- [31] LEI L, YANG Q L, YANG L, et al. Deep learning implementation of image segmentation in agricultural applications: A comprehensive review[J]. *Artificial Intelligence Review*, 2024, 57(6). DOI: 10.1007/s10462-024-10775-6.
- [32] LIN A L, CHEN B Z, XU J Y, et al. DS-TransUNet: Dual swin transformer U-Net for medical image segmentation[J]. *IEEE Transactions on Instrumentation and Measurement*, 2022, 71. DOI: 10.1109/TIM.2022.3178991.

- [33] NASIM M A AL, MUNEM A AL, ISLAM M, et al. Brain tumor segmentation using enhanced U-Net model with empirical analysis[C]//2022 25th International Conference on Computer and Information Technology. Piscataway: IEEE, 2023: 1027-1032.
- [34] LIU L Z, CHEN X H, ZHU S Y, et al. CondLaneNet: A top-to-down lane detection framework based on conditional convolution[C]//2021 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2022: 3753-3762.
- [35] WILM F, AMMELING J, ÖTTL M, et al. Rethinking U-Net skip connections for biomedical image segmentation [EB/OL]. (2024-12-13) [2025-03-12]. <https://arXiv.org/abs/2402.08276>.
- [36] GUO B Y, WANG Y T, ZHEN S, et al. SPEED: Semantic prior and extremely efficient dilated convolution network for real-time metal surface defects detection[J]. IEEE Transactions on Industrial Informatics, 2023, 19(12): 11380-11390.
- [37] XU Q, MA Z C, DUAN W. DCSAU-Net: A deeper and more compact split-attention U-Net for medical image segmentation[J]. Computers-in Biology and Medicine, 2023, 154. DOI: 10.48550/arXiv.2202.00972.
- [38] ZHANG Y H, LU H Y, MA G Y, et al. MU-Net: Embedding MixFormer into unet to extract water bodies from remote sensing images[J]. Remote Sensing, 2023, 15(14). DOI: 10.3390/rs15143559.
- [39] FENG H, SONG K C, CUI W Q, et al. Cross position aggregation network for few-shot strip steel surface defect segmentation[J]. IEEE Transactions on Instrumentation and Measurement, 2023, 72. DOI: 10.1109/TIM.2023.3246519.
- [40] SCHERR T, LÖFFLER K, BÖHLAND M, et al. Cell segmentation and tracking using CNN-based distance predictions and a graph-based matching strategy[J]. PLoS One, 2020, 15(12). DOI: 10.1371/journal.pone.0243219.
- [41] 张卓尔, 潘俊, 舒奇迪. 基于双路细节关注网络的遥感影像建筑物提取[J]. 武汉大学学报(信息科学版), 2024, 49(3): 376-388.
ZHANG Z E, PAN J, SHU Q D. Building extraction based on dual-stream detail-concerned network[J]. Geomatics and Information Science of Wuhan University, 2024, 49(3): 376-388. (in Chinese)
- [42] ZHANG X L, LIANG L, ZHAO S L, et al. GRFB-UNet: A new multi-scale attention network with group receptive field block for tactile paving segmentation[J]. Expert Systems with Applications, 2024, 238. DOI: 10.1016/j.eswa.2023.122109.
- [43] PAL K, YADAV P, KATAL N. RoadSegNet: A deep learning framework for autonomous urban road detection[J]. Journal of Engineering and Applied Science, 2022, 69(1). DOI: 10.1186/s44147-022-00162-9.
- [44] LIU J Y, ZHANG Q Y, WAN X, et al. LuSNAR: A lunar segmentation, navigation and reconstruction dataset based on Muti-sensor for autonomous exploration[EB/OL]. (2024-10-26) [2025-03-12]. <https://arXiv.org/abs/2407.06512>.
- [45] WEI T Q, CHEN Z, YU X, et al. PlantSeg: A large-scale in-the-wild dataset for plant disease segmentation[EB/OL]. (2024-09-06) [2025-03-12]. <https://arXiv.org/abs/2409.04038>.
- [46] WANG Y M, HA T, ALDRIDGE K, et al. Weed mapping with convolutional neural networks on high resolution whole-field images[C]//2023 IEEE/CVF International Conference on Computer Vision Workshops. Piscataway: IEEE, 2023: 505-514.
- [47] RUAN J C, XIE M Y, XIANG S C, et al. MEW-UNet: Multi-axis representation learning in frequency domain for medical image segmentation[EB/OL]. (2022-10-25) [2025-03-12]. <https://arXiv.org/abs/2210.14007>.
- [48] XU J C, XIONG Z X, BHATTACHARYYA S P. PID-Net: A real-time semantic segmentation network inspired by PID controllers[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2023: 19529-19539.
- [49] HONG Y, PAN H, SUN W, et al. Deep dualresolution networks for real-time and accurate semantic segmentation of road scenes[EB/OL]. (2021-09-01) [2025-03-12]. <https://arxiv.org/abs/2101.06085>.
- [50] IBTEHAZ N, KIHARA D. ACC-UNet: A completely convolutional UNet model for the 2020s[M]//Medical Image Computing and Computer Assisted Intervention - MICCAI 2023. Cham: Springer Nature Switzerland, 2023: 692-702.
- [51] AZAD R, ASADI-AGHBOLAGHI M, FATHY M, et al. Bi-directional ConvLSTM U-Net with densley connected convolutions[C]//2019 IEEE/CVF International Conference on Computer Vision Workshop. Piscataway: IEEE, 2019: 406-415.
- [52] LIU T H, HE Z S, LIN Z J, et al. An adaptive image segmentation network for surface defect detection[J]. IEEE Transactions on Neural Networks and Learning Systems, 2024, 35(6): 8510-8523.

[53] XU R G, HAO R Y, HUANG B Q. Efficient surface defect detection using self-supervised learning strategy and segmentation network[J]. Advanced Engineering Informatics, 2022, 52. DOI: 10.1016/j.aei.2022.10156.

[54] YU H, CHO Y, KANG B, et al. Embedding-free transformer with inference spatial reduction for efficient semantic segmentation[M]//Computer Vision - ECCV 2024. Cham: Springer Nature Switzerland, 2024: 92-110.

作者简介



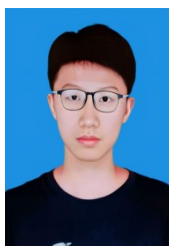
黄依莎 女, 2004年2月出生于江西省萍乡市. 现为湖南工商大学人工智能与先进计算学院(湘江书院)本科生. 主要研究方向为语义分割.
E-mail: 2142827479@qq.com



梁欣 女, 2005年5月出生于湖南省郴州市. 现为湖南工商大学人工智能与先进计算学院(湘江书院)本科生. 主要研究方向为语义分割.
E-mail: 3278028915@qq.com



姜林 男, 1977年11月出生于湖南省常德市. 现为湖南工商大学人工智能与先进计算学院(湘江书院)教授. 主要研究方向为智能语音处理、机器视觉、机器人应用.
E-mail: jlcdf@163.com



曾伟豪 男, 2004年4月出生于湖南省益阳市. 现为湖南工商大学人工智能与先进计算学院(湘江书院)本科生. 主要研究方向为语义分割.
E-mail: 2307225478@qq.com



管亚菲 女, 2005年8月出生于湖南省永州市. 现为湖南工商大学人工智能与先进计算学院(湘江书院)本科生. 主要研究方向为语义分割.
E-mail: 2645740476@qq.com



方晓萍 女, 1984年12月出生于湖南省娄底市. 现为湖南工商大学数学与统计学院教授. 主要研究方向为机器学习、大数据统计分析方法.
E-mail: fxp1222@163.com



张亚莎 女, 2004年5月出生于湖南省浏阳市. 现为湖南工商大学智能工程与制造学院本科生. 主要研究方向为语义分割.
E-mail: 2980083632@qq.com