

DOI:10.19651/j.cnki.emt.2519064

基于动态三元注意力网络的火星地表图像分割^{*}

孟祥源 吴心悦 张雁皓 高润泽 单慧琳

(无锡学院电子信息工程学院 无锡 214105)

摘要: 作为地球在宇宙演化进程中具有同源关联的近邻类地行星,火星表面地貌特征的语义分割不仅能够构建行星尺度形成动力学与演化机制的认知框架,更可为行星科学领域奠定多维度研究范式,特别是在完善行星演化理论体系与验证天体物理模型方面具有关键性科学价值。然而,火星地表影像解析面临多维技术挑战,主要体现在光照条件复杂多变、地形特征结构化程度低以及目标尺度分布异质性显著等特性,这些特征共同构成行星表面智能解译的关键技术瓶颈。针对传统分割模型存在的细节特征丢失、复杂边界误分割等问题,本文提出一种基于动态三元注意力的火星地表图像分割算法,该算法通过自适应特征融合与动态三元注意力机制协同优化,提高分割精度。首先,构建动态三元注意力模块以自动调整不同注意力分支的重要性,可针对火星岩石、沙丘等典型地貌特征实现局部和全局信息的动态聚焦。其次,针对空间信息和语义信息设计了自适应双向特征融合模块,以缓解多尺度特征冲突。此外,提出通道注意力可分离卷积,在减少参数数量的同时也能够提高模型的泛化能力。实验结果表明,本文所提算法在 S⁵ Mars 数据集上的准确率和平均交并比分别达到了 89.06% 和 72.33%,证明本算法能够有效地提取语义特征,并将其有效融合,提高火星地表图像分割的准确率。

关键词: 火星地表图像;语义分割;动态三元注意力;自适应融合;通道注意力可分离卷积

中图分类号: TN911.73;P185.3 **文献标识码:** :A **国家标准学科分类代码:** 520.2070

Semantic segmentation of Martian surface images based on dynamic ternary attention network

Meng Xiangyuan Wu Xinyue Zhang Yanhao Gao Runze Shan Huilin

(School of Electronic and Information Engineering, Wuxi College, Wuxi 214105, China)

Abstract: As an adjacent planet with profound connections to the Earth in cosmic evolution, the semantic segmentation of Martian surface geomorphological features not only facilitates the construction of a cognitive framework for understanding the dynamic formation and evolutionary mechanisms at the planetary scale but also establishes a multifaceted research paradigm in the field of planetary science. This holds particular significance in refining the theoretical framework of planetary evolution and validating astrophysical models, thereby possessing critical scientific value. However, the analysis of Martian surface imagery encounters multifaceted technical challenges primarily characterized by complex and variable lighting conditions, low structural degree of topographical features, and pronounced heterogeneity in target scale distribution. Collectively, these characteristics form key technological bottlenecks in the intelligent interpretation of planetary surfaces. To address these issues, this paper proposes a Mars surface image segmentation algorithm based on dynamic ternary attention mechanism. Our approach synergistically optimizes adaptive feature fusion and dynamic attention mechanisms to enhance segmentation accuracy. First, we develop a dynamic ternary attention module that automatically adjusts branch significance weights, enabling dynamic focus on local and global features for typical Martian landforms like rocks and dunes. Second, an adaptive bidirectional feature fusion module is designed to reconcile spatial and semantic information conflicts across scales. Moreover, a channel-attentive separable convolution is proposed to reduce parameter complexity while enhancing model generalization capabilities. Experimental results demonstrate that the proposed algorithm achieves 89.06% accuracy and 72.33% mean intersection over union on the S⁵ Mars dataset, effectively extracting and integrating multi-scale features to significantly enhance segmentation precision for Martian surface imagery.

Keywords: Martian surface images; semantic segmentation; dynamic ternary attention; adaptive fusion; channel-attentive separable convolution

0 引言

地球上存在多种与地外环境相似的极端条件,既包括

深海热液、低氧、无氧、高压、蓝洞、冷泉及沉积物等海洋环境,也涵盖地表的高盐碱、干旱、强酸性和剧烈温差等极端环境^[1]。地球上的多种极端环境与现代或早期火星环境相

收稿日期:2025-06-10

* 基金项目:国家自然科学基金(62071240,62106111)、国家级大创项目(202413982010Z)资助

似^[2]。目前全球已建立30余个类地外天体环境研究点,我国柴达木盆地也被列为类火星基地^[3]。火星的早期演化阶段经历了诸多与地球类似的地质活动^[4-5]。深入开展火星地表地形研究有助于揭示火星上的地质构造、地貌、地质资源等信息。

然而火星地形的结构十分复杂,给高精度地表语义分析带来了巨大的挑战,传统地形分割方法要求操作者具备专业的领域知识以提取有效的图像特征,这类方法的适应范围非常有限,难以应对复杂多变的实际应用场景^[6]。

近年来,深度学习技术推动图像语义分割领域取得了突破性的进展,卷积神经网络(convolutional neural networks, CNN)是深度学习中具有代表性的网络结构,在计算机视觉任务中展现出卓越的性能。在图像语义分割领域,基于卷积神经网络的模型大多采用相似的骨干网络架构,当前主流的骨干网络包括 AlexNet^[7]、VGGNet^[8]、ResNet^[9]、GoogLeNet^[10]以及 MobileNet^[11]等经典结构。Long 等^[12]提出的全卷积网络(fully convolutional networks, FCN)通过端到端的像素级预测能力,为语义分割奠定了重要基础。火星地表图像相较于自然图像,具有地形结构较为松散,且地形目标的尺寸差异显著等特点,亟需模型兼顾局部细节的同时融合多尺度上下文信息。Badrinarayanan 等^[13]提出的 SegNet 采用了编码器-解码器架构,但编码过程会导致分辨率损失。Ronneberger 等^[14]提出的 U-Net 凭借跳跃连接实现浅层细节与深层语义的跨层融合,成为医学领域的重要基准模型。然而,这些模型在特征复用过程中存在低级特征语义增强缺乏、多尺度信息交互不足等瓶颈。为提升特征的判别能力,研究者们相继提出多种多尺度的网络模型, Lin 等^[15]提出了 FPN 网络,它由自下而上的路径和横向路径组成,构成特征金字塔,用以融合低分辨率和高分辨率的特征。Zhao 等^[16]提出的 PSPNet 通过金字塔池化模块(pyramid pooling module, PPM)聚合多级上下文特征, Chen 等^[17]提出的 DeepLabV3+ 则引入空洞空间金字塔池化(atrous spatial pyramid pooling, ASPP),在扩大感受野的同时保持特征分辨率,但缺乏深层特征交互,对大尺度目标分割效果不足。面向实时性需求, Yu 等^[18]提出的 BiSeNet 创新性地构建空间细节与语义上下文双路径网络,兼顾了精度与效率,但通过简单拼接融合双路径特征,导致低层细节与高层语义的融合不够充分。此外,还有的研究者尝试通过注意力机制引导特征选择, Chen 等^[19]提出了一种新的注意力机制,该机制能够为每个像素位置学习相应的权重,并对多尺度特征进行软加权。与此同时, Li 等^[20]设计了一种金字塔注意力网络,该网络将注意力机制与空间金字塔结合在一起。Fu 等^[21]提出了一种双重注意力网络,专门用于场景分割,该网络利用注意力机制来捕获丰富的上下文依赖关系。尽管这些模型展现出优越的性能,然而在上下文融合的效率上仍有待进一步改善。

由于火星地表存在光照条件复杂多变、地形特征结构化程度低以及目标尺度分布异质性显著等特性,传统的图像处理方法无法对火星地表图像进行准确的分割,因此,本文提出一种基于动态三元注意力与自适应融合网络(dynamic ternary attention adaptive fusion network, DTAFNet)用于火星地表图像分割。主要创新点包括:

1) 动态多元特征提取。针对火星光照条件复杂多变的问题,在主干网络中提出动态三元注意力机制(dynamic ternary attention, DTA),通过多维度注意力协同与参数化残差连接机制,过滤光照干扰,提取关键特征,有效处理复杂空间布局的图像数据。

2) 自适应双向特征融合。针对火星地表地形特征结构化程度低,岩石、沙丘、空洞等形态随机破碎,边界模糊的问题,在融合模块中,提出一种自适应双向特征融合模块(adaptive bidirectional fusion module, ABFM),通过自适应学习权重,调整不同特征图的贡献,充分融合语义信息与空间信息,避免固定规则导致细节信息丢失,导致边缘分割模糊。

3) 火星地表图像中目标尺度分布异质性显著,同时存在巨型地貌特征与微小目标特征,针对以上问题,本文提出通道注意力可分离卷积(channel-attentive separable convolution, CASConv),将通道注意力与深度可分离卷积相结合,利用通道注意力对空间特征进行动态筛选,避免了逐点卷积前的特征冗余,增强对小目标敏感的通道信息,动态提升关键通道的响应,使单层卷积同时适配不同尺度目标。

1 数据集及评价指标

本文使用 S⁵ mars 数据集^[22]对 DTAFNet 进行火星地表图像语义分割性能评估。S⁵ mars 数据集包含 6 000 张由好奇号探测车的彩色桅杆相机拍摄的火星地形图像,分辨率为 1 200 pixel×1 200 pixel。该数据集的图像可分为两类:3 000 张简单图像和 3 000 张复杂图像。简单图像中的地形结构单一,多为平坦区域,单一地貌类型占比较高,且目标边界清晰、背景干扰少、特征明显且易于识别,非常适合用于模型的初步训练;而复杂图像则包含较为复杂的地形特征,多类地貌并存,遮挡或重叠明显,且纹理复杂度高。数据集包含 10 种类别标签,分别是天空(Sky)、山脊(Ridge)、土壤(Soil)、沙地(Sand)、基岩(Bedrock)、岩石(Rock)、漫游车(Rover)、车迹(Trace)、空洞(Hole)以及其他(Background)。数据集按 8:1:1 的大小比例被随机分为训练集、验证集、测试集。

本文采用的评估指标包括准确率(accuracy, Acc)、单类交并比(IoU)、所有类别的平均交并比(mIoU)、F1 指数(F1-score, F1)以及均值平均精度(mAP)。

2 DTAFNet 网络结构

本文所提出的 DTAFNet 网络的整体结构如图 1 所

示,通过双分支协同架构实现了高效特征提取与融合。DTAFNet 主要由空间分支、语义分支和自适应特征融合模块三大部分构成。空间分支采用了三层次渐进式特征提取结构,通过多层常规卷积构建了基础特征表达。针对卷积操作中关键特征易丢失的问题,设计了一个新的注意力机制 DTA,DTA 通过动态权重分配,同步增强了空间、通道及位置 3 个维度的特征响应,在空间分支的第二层与最后一层加入 DTA,显著增强了边缘细节与纹理信息等关键空间特征的表达能力。语义分支的设计采用层次化特征精炼策略,首先通过常规卷积进行初级语义特征建模,随后引入 CASConv,在保证分割精度的同时,将注意力引导的通道特征与深度可分离卷积相结合,在保证模型特征提取能力的同时减少参数量,并且能充分挖掘通道信息,以提高分

割精度。此外,采用了残差结构,以尽可能保留有效的语义特征,确保语义特征在卷积层数增加时不会减少,经过多次常规卷积与 CASConv 模块的级联处理后,采用一次三元注意力机制对高层语义特征进行全局关系建模,以增强对目标整体结构的表征能力,最后通过一次常规卷积和 CASConv 组合操作生成语义分支的输出特征。

本文采用的自适应双向特征融合模块 ABFM 通过双向跨尺度特征交互机制实现空间细节与语义信息的深度耦合。该模块采用自适应权重学习策略,调节不同层次特征的贡献度,通过特征重加权与跨维度交互有效解决融合后特征丢失的问题。融合后的多尺度特征图经过一次卷积调整通道数,再使用经双线性插值对特征图进行上采样,最后通过一次卷积得到最终分割结果。

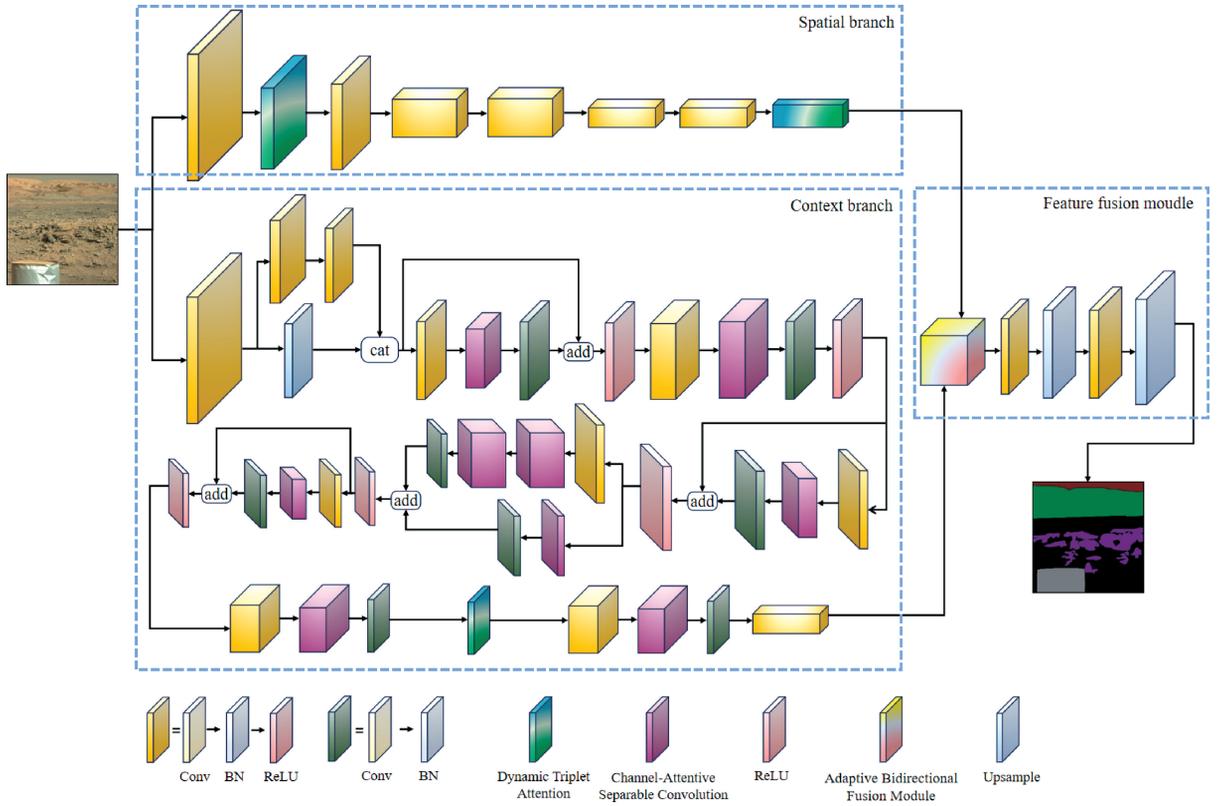


图 1 DTAFNet 的整体结构

Fig. 1 The overall framework of DTAFNet

2.1 动态三元注意力机制

现有的注意力机制,如卷积注意力模块(convolutional block attention module,CBAM)^[23],在处理跨维度特征交互时,空间与通道注意力通常采用串行堆叠或简单相加,需要更多的计算资源,导致计算复杂度显著增加,高效通道注意力机制(efficient channel attention,ECA)^[24]未能充分挖掘多维特征的协同效应,难以适应不同层级的动态需求。本文创建了一种动态三元注意力机制 DTA,DTA 的结构如图 2 所示。

DTA 能够通过低成本的操作融合多个方面的特征

图,同时引入可学习的动态权重分配策略,以实现网络对各维度注意力贡献度的自主调节,此外,利用残差结构缓解梯度消失的问题。DTA 的整体过程如式(1)所示。

$$O = \mathcal{X} + \gamma \cdot \sum_{i=1}^3 \omega_i \mathcal{B}_i(\mathcal{X}) \tag{1}$$

其中, O 即为 DTA 的最终输出,保留了原始特征与注意力特征的加权融合结果。 \mathcal{X} 表示输入特征图, γ 表示可学习的残差缩放因子,初始值为 0.2,控制注意力特征的贡献程度, ω_i 表示归一化后的分支权重,其计算方式由式(2)表示, $\mathcal{B}_i(\mathcal{X})$ 表示第 i 个分支注意力的输出, $\mathcal{B}_i(\mathcal{X})$ 的

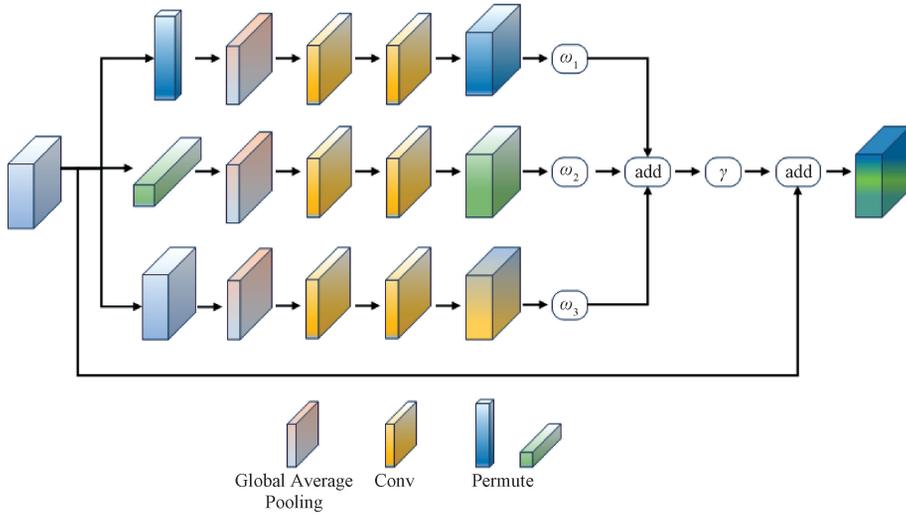


图 2 DTA 结构

Fig. 2 DTA structure

计算过程如式(3)~(5)所示。

$$\omega_i = \frac{\exp(\alpha_i)}{\sum_{j=1}^N \exp(\alpha_j)}, \alpha_i \in \mathbb{R}, \sum_{i=1}^N \omega_i = 1 \quad (2)$$

其中, α_i 表示可学习权重参数向量, 控制各分支的原始重要性程度。N 表示注意力分支的数量, 此处取值为 $N = 3$ 。

$$\mathcal{B}_1(\mathcal{X}) = \sigma(\text{Conv}_{k \times k}^{(2)}(\text{Conv}_{k \times k}^{(1)}(\text{GAP}(\text{PERH}(\mathcal{X})))) \quad (3)$$

$$\mathcal{B}_2(\mathcal{X}) = \sigma(\text{Conv}_{k \times k}^{(2)}(\text{Conv}_{k \times k}^{(1)}(\text{GAP}(\text{PERW}(\mathcal{X})))) \quad (4)$$

$$\mathcal{B}_3(\mathcal{X}) = \sigma(\text{Conv}_{k \times k}^{(2)}(\text{Conv}_{k \times k}^{(1)}(\text{GAP}(\mathcal{X})))) \quad (5)$$

其中, σ 表示 Sigmoid 激活函数, $\text{Conv}_{k \times k}^{(l)}$ 表示第 l 层卷积核大小为 $k \times k$, 此处取 $k = 7$, 第一层卷积操作包含了批量归一化 BN(batch normalizaion) 以及 ReLU 激活, GAP 表示全局平均池化操作, PERH 表示在高度方向上进行维度转置操作, PERW 表示在宽度方向上进行维度转置

操作, 通过对输入特征图进行高度和宽度不同方向上的转置, 以充分提取不同维度上的空间特征, 第 3 条注意力分支用于提取全局注意力, 最终经过各分支权重的重新加权, 实现特征的动态选择。

2.2 自适应双向特征融合模块

在火星地表图像中, 目标尺度差异大、远景小物体与近景大物体并存, 图像分割需同时理解空间特征与全局语义特征之间的关系。传统的静态融合方式难以兼顾不同场景的空间与语义需求。边界区域的细节和语义信息可能存在冲突, 因此需要动态调整融合权重。本文提出了一种自适应双向特征融合模块 ABFM, 根据输入内容自动调整空间与语义特征的融合权重, 缓解特征冲突。ABFM 的结构如图 3 所示。ABFM 模块的融合过程可以由式(6)表示。

$$\mathbf{F}_{\text{out}} = \sigma(\text{Conv}(\text{Conv}([\mathbf{F}_c, \mathbf{F}_s]))) \odot \mathbf{F}_c + (1 - \sigma(\text{Conv}(\text{Conv}([\mathbf{F}_c, \mathbf{F}_s]))) \odot \mathbf{F}_s + \mathbf{F}_c \quad (6)$$

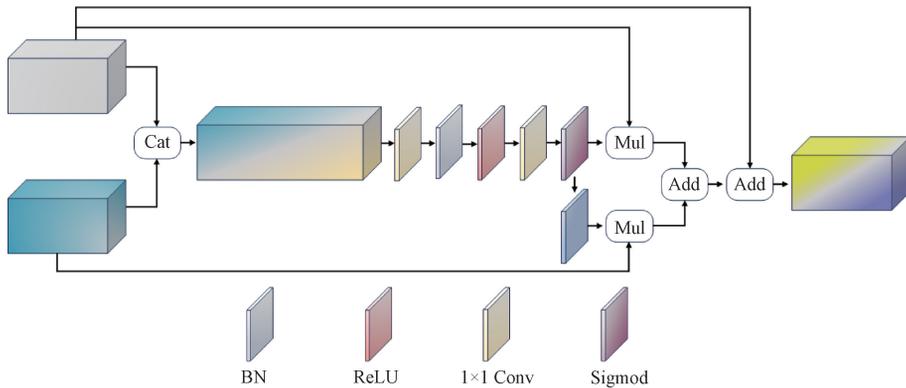


图 3 自适应双向特征融合模块结构

Fig. 3 Adaptive bidirectional fusion module structure

其中,为 F_{out} 为最终输出的融合特征, σ 表示 Sigmoid 激活函数, F_s 与 F_l 分别表示空间分支与语义分支的输出结果, \odot 表示逐元素乘法 (Hadamard 乘积)。首先针对空间分支与语义分支输出的特征图进行拼接,并使用一次常规卷积,结合 ReLU 激活函数提取非线性特征,通过一次卷积调整通道数,再经 Sigmoid 函数归一化,生成权重注意力图;其次将两个特征按注意力权重逐元素相乘并进行相加;最后将融合结果与原始空间分支的输出特征 F_s 进行残差连接得到最终输出的融合特征 F_{out} 。

2.3 通道注意力可分离卷积

传统的深度可分离卷积 (depthwise separable convolution)^[25]通过解耦空间卷积与通道卷积来降低计算量,但其静态权重分配机制难以适配语义分割任务中复杂多变的上下文场景。后续研究尝试引入通道注意力增强特征选择能力,但直接将标准通道注意力模块堆叠于深度卷积后会导致空间信息损失,本文提出通道注意力可分离卷积 (CASConv),其核心思想是将通道注意力机制嵌入深度可分离卷积的中间阶段,在深度卷积后应用 SE 模块 (squeeze-and-excitation block)^[26],利用通道注意力对空间特征进行动态筛选,避免了逐点卷积前的特征冗余,实现效率和精度的平衡,CASConv 的结构如图 4 所示。

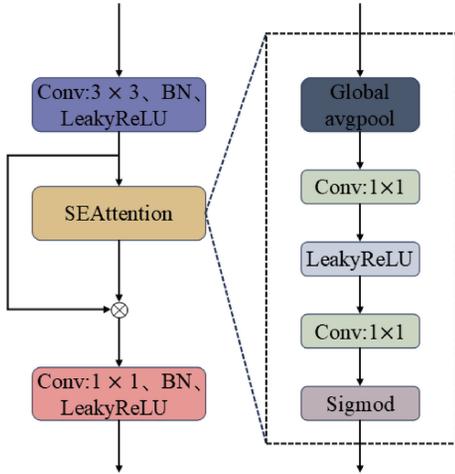


图 4 CASConv 结构
Fig. 4 CASConv structure

通道注意力可分离卷积首先通过深度卷积提取局部空间信息,然后采用改进型 SE 模块,以激活函数 LeakyReLU 代替原来的 ReLU 函数。基于全局平均池化后的特征生成通道注意力,随后通过 1×1 逐点卷积重组语义信息,将输出维度映射至目标通道数。

3 数据结果处理与分析

本文使用 SGD 优化器对模型进行训练,初始学习率设为 0.05,动量为 0.9,权重衰减为 0.000 4 批量大小为 8,迭代次数设为 100,为了进一步的缩减模型参数的梯度,从而

减少计算量,提高训练速度,使用余弦退火学习率调整策略,通过将学习率在每个周期内按照余弦函数的形式逐渐减小来优化训练过程,使得模型在训练的早期阶段可以快速收敛,而在后期则降低学习率以便更精细地调整权重。实验环境如表 1 所示。

表 1 实验环境

Table 1 Experimental environment

配置	型号
CPU	Intel i5-12490F
GPU	NVIDIA GeForce RTX 4060Ti
操作系统	Windows 10
运行内存	16 G
显存	8 G
语言及框架	Python3. 8+Pytorch1. 13. 1

3.1 DTAFNet 与其他模型对比实验分析

为证明本文提出的 DTAFNet 的分割性能,首先将其与已有的图像分割网络 U-Net、PSPNet、FCN、DeepLabV3 +、SegNet、BiSeNet 在 S^5 mars 数据集的简单部分上进行对比。各类别的单类交并比、准确率、均值平均精度及平均交并比的实验结果如表 2 所示。

由表 2 可知,本文所提的 DTAFNet 比已有的图像分割网络具有更好的分割性能。DTAFNet 在天空、山脊、土壤、沙地、岩石等类别达到了最好的分割性能,mAP 与平均交并比均为最高,与采用了编码器-解码器结构的经典网络 U-Net 相比,DTAFNet 的 mIoU 比 U-Net 提高了 6.61%,mAP 比 U-Net 提高了 6.07%,准确率比 U-Net 提高了 2.5%。与一些优秀的传统网络相比,如 FCN、PSPNet、DeepLabV3 + 和 SegNet,DTAFNet 在 mIoU、mAP、准确率 3 个指标上均表现出色,处于领先地位。而 BiSeNet 虽然将语义信息和空间信息相结合,但采用固定的融合策略,导致细节信息的丢失,进而影响了最终的分割效果。DTAFNet 则通过自适应融合的设计,避免了简单拼接带来的信息丢失问题,更有效地捕捉了火星地表图像中的多尺度、多维度特征。因此,DTAFNet 的 mIoU 相较于 BiSeNet 提高了 6.33%,mAP 提高了 7.32%,准确率提高了 3.21%。

同时,本文在 S^5 mars 数据集的复杂部分进行了相同的对比实验,实验结果如表 3 所示。由于 S^5 mars 数据集的复杂部分中,空洞类别的训练样本数量极少,以至于无法进行有效的评估,为清晰起见,该类别在表格中用“—”表示。从表 3 中数据可知,DTAFNet 的各项指标均领先于其他模型,说明模型即使在面对较为复杂的图像时,仍能够准确地识别目标,具有较好的鲁棒性。

最后,本文在 S^5 mars 全量数据集上进行对比实验,全量数据集包括了简单图像部分与复杂图像部分,实验结果

如表4所示,从表中数据可知,DTAFNet在全量数据集上的mIoU、mAP、准确率3个指标均优于其他模型,其平均交并比达到72.33%,相较于次优模型DeepLabV3+提升3.65%。尤其是mAP指标,DTAFNet达到了82.12%,是众多模型中唯一超过80%的模型。

为了进一步验证所提出的DTAFNet分割的优越性,将该模型与U-Net、PSPNet、FCN、DeepLabV3+、SegNet、BiSeNet在S⁵mars数据集上进行了泛化对比实验。图5和6对S⁵mars数据集的简单图像部分和复杂图像部分分别进行分析,主要分为基岩、岩石、山脊、沙地、空洞等。

表2 S⁵mars数据集简单部分中的不同网络模型的验证结果

Table 2 Verification results under different network models on the easy part of the S⁵ mars dataset

模型	背景	天空	山脊	土壤	沙地	基岩	岩石	漫游车	车迹	空洞	Acc	mAP	mIoU
U-Net	84.57	96.98	97.20	49.09	49.74	81.19	23.87	94.00	58.01	52.32	89.12	78.84	69.70
PSPNet	75.82	89.67	90.21	34.30	43.44	70.92	15.77	72.54	37.29	10.86	82.03	68.21	54.08
FCN	82.41	96.97	96.75	44.79	44.37	78.34	16.87	86.34	51.29	21.93	87.65	70.58	62.01
DeepLabV3+	86.41	96.59	96.26	59.39	66.55	79.30	27.49	91.14	64.08	55.49	89.86	83.48	72.27
SegNet	82.82	97.47	97.47	45.14	51.08	79.18	19.97	85.50	50.89	25.33	88.28	71.82	63.48
BiSeNet	83.45	97.16	97.19	45.77	46.59	79.90	21.63	86.38	61.45	50.31	88.41	77.59	66.98
DTAFNet	87.34	98.09	97.88	60.10	75.80	82.85	40.73	79.93	81.26	59.12	91.62	84.91	76.31

表3 S⁵mars数据集复杂部分中的不同网络模型的验证结果

Table 3 Verification results under different network models on the hard part of the S⁵ mars dataset

模型	背景	天空	山脊	土壤	沙地	基岩	岩石	漫游车	车迹	空洞	Acc	mAP	mIoU
U-Net	79.56	91.32	89.96	30.45	63.15	70.01	20.88	37.98	60.87	—	85.22	63.16	54.42
PSPNet	76.77	93.46	86.96	35.61	65.48	65.91	19.96	38.47	50.63	—	83.55	62.55	53.33
FCN	81.63	88.31	90.09	36.84	60.73	65.24	15.86	15.15	61.16	—	85.04	57.77	51.50
DeepLabV3+	79.78	91.16	89.16	35.66	64.75	69.68	24.75	34.70	73.30	—	85.42	64.78	56.29
SegNet	77.88	92.42	88.86	30.64	66.09	69.27	18.25	81.20	67.22	—	84.50	66.69	59.18
BiSeNet	79.73	90.98	88.07	32.59	61.53	69.30	22.00	29.70	69.03	—	85.12	61.95	54.29
DTAFNet	80.21	94.13	92.12	36.88	72.35	70.89	24.64	83.84	68.00	—	86.30	69.17	62.31

表4 S⁵mars数据集中不同网络模型的验证结果

Table 4 Verification results under different network models on the entire S⁵ mars dataset

模型	背景	天空	山脊	土壤	沙地	基岩	岩石	漫游车	车迹	空洞	Acc	mAP	mIoU
U-Net	82.42	94.65	93.90	41.76	62.85	76.42	27.13	81.81	59.12	28.14	87.60	73.82	64.84
PSPNet	78.47	92.48	90.71	44.44	61.89	70.64	26.30	70.85	69.97	52.17	85.11	78.98	65.79
FCN	80.38	93.74	92.60	39.62	57.01	74.67	15.93	77.80	57.61	11.92	86.07	67.99	60.13
DeepLabV3+	81.87	95.43	94.55	49.27	65.51	74.86	29.51	85.80	64.42	45.59	87.51	79.58	68.68
SegNet	79.80	91.14	80.63	56.38	43.77	68.95	28.40	63.91	59.74	34.95	84.67	79.57	60.77
BiSeNet	81.63	95.13	93.65	44.09	60.94	74.24	23.44	77.93	58.45	58.01	87.02	75.34	66.75
DTAFNet	83.65	97.26	96.11	55.16	71.35	76.20	36.60	80.22	69.63	27.12	89.06	82.12	72.33

S⁵mars数据集简单部分的语义分割结果对比图如图5所示,第1行图片以山脊与土壤为主,PSPNet、FCN、SegNet与BiSeNet左边区域和右边区域存在多处将背景分割为土壤的情况,且分割边缘较为模糊;而U-Net与DeepLabV3+分割错误区域较少,但仍存在部分土壤的区域未正确分割的情况。DTAFNet分割准确度最高,且左边区域错误分割的部分最少。第2行与第5行图片以空洞为主,本文模型分割空洞边缘最清晰,与标签图吻合度最

佳。第3行图片中基岩占比较多,且与岩石相似度较高,人眼难以准确地分辨出两者的区别。诸如U-Net、PSPNet、FCN、SegNet与BiSeNet等网络,在处理这些颜色相似区域时,通常依赖于图像的颜色特征进行分割。然而,这些网络在分割时并未充分考虑空间特征与全局语义特征之间的关系,导致部分岩石区域被误判为基岩。DeepLabV3+在网络架构上采用了空洞卷积的方法,以获得更大感受野,能够显著改善误分割的问题,但其边缘分

割仍较为粗糙,DTAFNet的分割效果最佳,不仅能正确分割出岩石,而且岩石与土壤交界处的边缘分割最为清晰。第 4 行图片以车迹为主,U-Net、FCN、SegNet 与 BiSeNet 将大部分车迹区域误分割为背景,U-Net 与 DeepLabV3+ 则将部分背景区域误分割为车迹,只有 DTAFNet 能够精确识别两者的边界,并保持两者的分割结果清晰且精确,展示了其在复杂环境下的优越性能。第 6 行图片以漫游车为主,边缘分割较为粗糙,尽管背景较为复杂且存在部分阴影遮挡,DTAFNet 依然能够保持较为精准的分割效果,且对边缘的处理十分细腻。

S⁵ mars 数据集复杂部分的语义分割结果对比图如图 6 所示,以第 6 行图片为例,图中含有较大比例的沙地。U-Net、FCN 存在将沙地误判为土壤的问题,PSPNet 将大面积沙地误判为了岩石,DeepLabV3+ 和 SegNet 在图片的下方区域出现了误分割现象,且沙地边缘分割较为粗糙。这些网络在处理沙地边缘时,由于只关注局部的颜色信息,难以准确提取沙地和土壤之间的细微差异,导致边缘分割精度不足。BiSeNet 在图片的上方区域没有将沙地正确分割出。DTAFNet 的分割结果在边缘清晰度和类别识别准确性上均表现最佳。

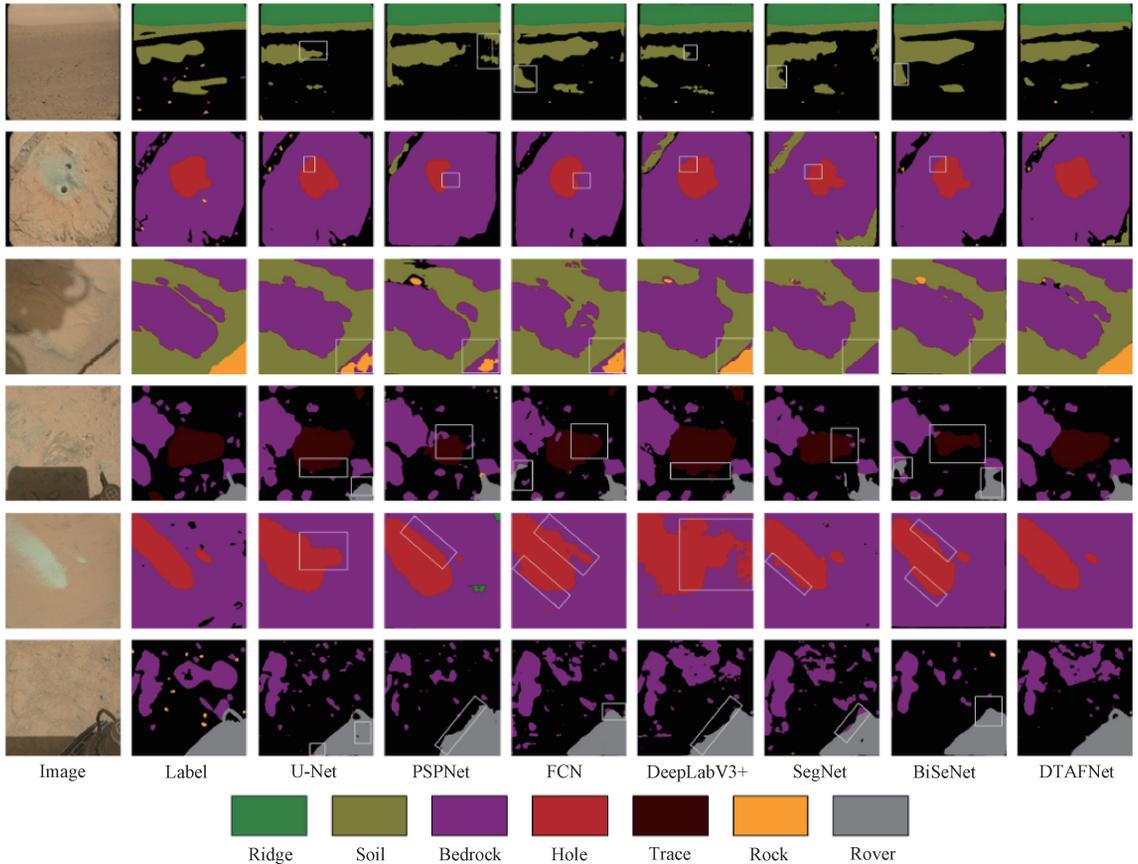


图 5 S⁵ mars 数据集中简单部分的分割图

Fig. 5 Segmentation of easy parts in the S⁵ mars dataset

3.2 消融实验结果对比

为了评估提出的各个模块的性能,在 S⁵ mars 上分别对本文所提出的模块进行实验。消融实验中基准网络(Baseline)的主干网络采用去除本文所提出的 DTA 以及 CASConv 后的结构,融合特征的方式采用元素相加的方式,上采样采用 unpool 的操作。

如表 5 所示,在进行特征融合时,使用 ABFM 替换基准网络中元素相加的方式,使得火星地表图像分割的指标大幅提高,与基准网络相比,准确率(Acc)提升了 1.77%,

F1 指数提升了 6.14%,均值平均精度(mAP)提升了 4.51%,平均交并比(mIoU)提升了 6.97%,因此,ABFM 是 DTAFNet 中提升性能最大的模块,DTA 次之。与基准网络相比 DTAFNet 的指标均有所提升,准确率达到了 89.06%,F1 指数达到了 79.21%,均值平均精度达到了 82.12%,平均交并比达到了 72.33%。因此,DTAFNet 能够显著增强火星地表图像语义特征与空间特征的提取与融合的能力,从而提高火星地表图像语义分割的性能。

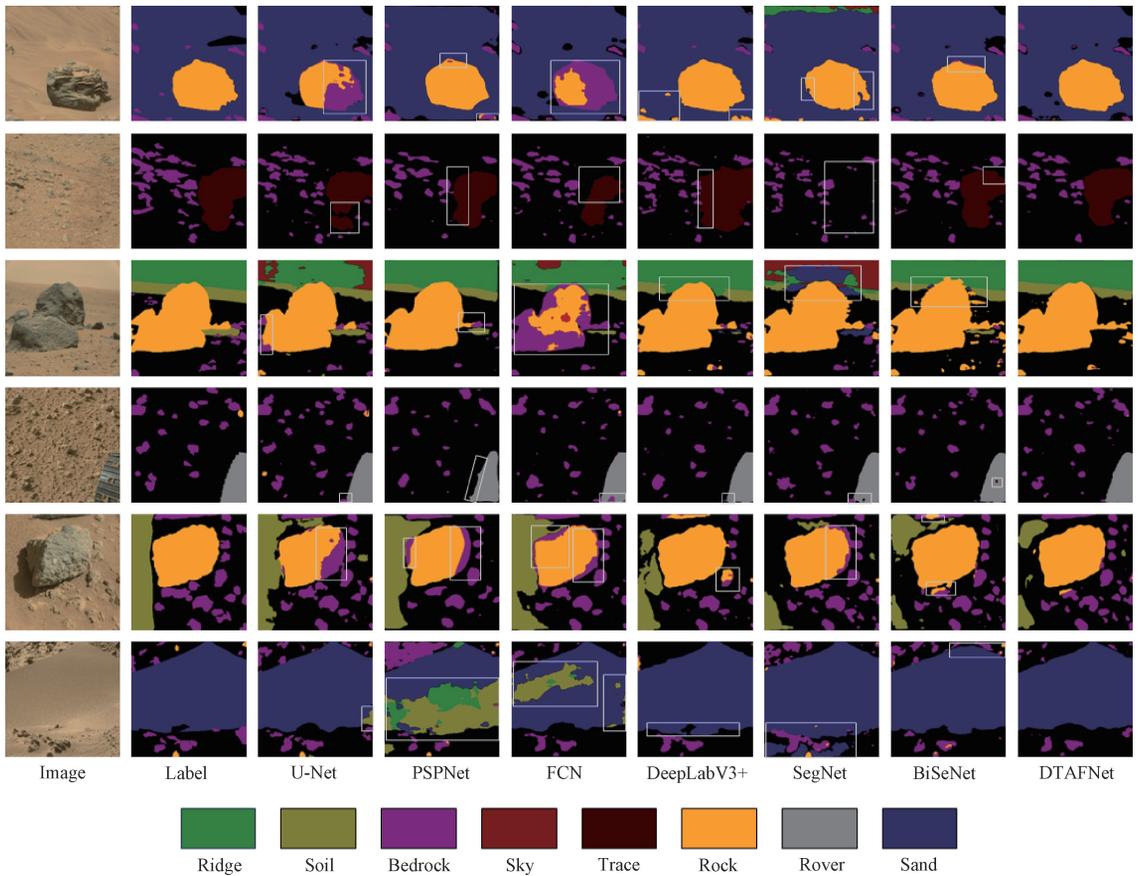


图 6 S⁵ mars 数据集中复杂部分的分割图

Fig. 6 Segmentation of hard parts in the S⁵ mars dataset

表 5 各个模块的消融实验

Table 5 Ablation experiments of each module

方法	Acc	F1	mAP	mIoU
Baseline	86.30	71.23	69.17	62.31
Baseline+DTA	88.90	74.86	75.08	65.19
Baseline+ABFM	88.07	77.37	73.68	69.28
Baseline+CASConv	88.23	74.05	74.37	64.46
DTAFNet	89.06	79.21	82.12	72.33

4 结 论

针对火星地表图像分割中存在的光照条件多变、地形结构复杂、目标尺度差异大等问题,本文提出了一种基于动态三元注意力与自适应融合的火星地表图像分割算法。首先,设计了动态三元注意力模块,通过调整不同注意力分支的权重,有效增强模型对关键地形特征的感知能力。其次,针对多尺度特征冲突问题,构建自适应双向特征融合模块对空间信息与语义信息的交互进行优化,实现了特征的有效融合并避免了无效特征的干扰。最后,提出通道注意力可分离卷积,在减少模型参数数量的同时,通过引入通道注意力机制增强特征表达能力,提高了模型的泛化性

能。本文在 S⁵ mars 数据集上进行了实验,实验结果表明,与其他分割模型相比,DTAFNet 在火星图像分割任务中表现出了优秀的性能,它更适用于分割颜色非常相似的基岩和岩石、沙地与土壤。

本文提出的 DTAFNet 提升了分割准确度,但对于极小尺度地形目标的分割仍有提升空间。在未来的研究中,可以考虑跨模态数据融合,获取更多特征信息,提高模型对高复杂度地形的理解能力,以实现更为精准的火星地表图像分割效果。

参考文献

[1] 国家自然科学基金委员会,中国科学院. 极端地质环境微生物学[M]. 北京: 科学出版社, 2022. National Natural Science Foundation of China, Chinese Academy of Sciences. Microbiology in extreme geological environments[M]. Beijing: Science Press, 2022.

[2] 史语桐,赵健楠,张明杰,等. 火星地质环境演变与生命宜居性探索[J]. 深空探测学报, 2025, 12(1): 86-96. SHI Y T, ZHAO J N, ZHANG M J, et al. Evolution of the martian geological environment and exploration of habitability of Mars[J]. Journal of Deep Space

- Exploration, 2025, 12(1): 86-96.
- [3] MARTINS Z, COTTIN H, KOTLER J M, et al. Earth as a tool for astrobiology: A European perspective [J]. *Space Science Reviews*, 2017, 209(1): 43-81.
- [4] CARR M H. The fluvial history of Mars [J]. *Philosophical Transactions of The Royal Society A Mathematical Physical and Engineering Sciences*, 2012, 370(1966): 2193-2215.
- [5] CARR M H, HEAD J W. Geologic history of Mars [J]. *Earth and Planetary Science Letters*, 2010, 294(3-4): 185-203.
- [6] 王嫣然, 陈清亮, 吴俊君. 面向复杂环境的图像语义分割方法综述 [J]. *计算机科学*, 2019, 46(9): 36-46.
WANG Y R, CHEN Q L, WU J J. Research on image semantic segmentation for complex environments [J]. *Computer Science*, 2019, 46(9): 36-46.
- [7] KRIZHEVSKY A, SUTSKEVER I, HINTON G. ImageNet classification with deep convolutional neural networks [C]. *Communications of the ACM*, 2017, 60(6): 84-90.
- [8] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [J]. *ArXiv preprint arXiv: 1409.1556*, 2014.
- [9] HE K M, ZHANG X Y, REN SH Q, et al. Deep residual learning for image recognition [C]. *IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 770-778.
- [10] SZEGEDY C, LIU W, JIA Y P, et al. Going deeper with convolutions [C]. *IEEE Conference on Computer Vision and Pattern Recognition*, 2015: 1-9.
- [11] HOWARD A G, ZHU M L, CHEN B, et al. MobileNets: Efficient convolutional neural networks for mobile vision applications [J]. *ArXiv preprint arXiv: 1704.04861*, 2017.
- [12] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation [C]. *IEEE Conference on Computer Vision and Pattern Recognition*, 2015: 3431-3440.
- [13] BADRINARAYANAN V, KENDALI 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.
- [14] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation [M]. *Lecture Notes in Computer Science*. Cham: Springer International Publishing, 2015: 234-241.
- [15] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection [C]. *IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 2117-2125.
- [16] ZHAO H SH, SHI J P, QI X J, et al. Pyramid scene parsing network [C]. *IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 2881-2890.
- [17] CHEN L CH, ZHU Y K, PAPANDEIOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation [C]. *European Conference on Computer Vision (ECCV)*, 2018: 801-818.
- [18] YU CH Q, WANG J B, PENG CH, et al. Bisenet: Bilateral segmentation network for real-time semantic segmentation [C]. *European Conference on Computer Vision (ECCV)*, 2018: 325-341.
- [19] CHEN L CH, YANG Y, WANG J, et al. Attention to scale: Scale-aware semantic image segmentation [C]. *IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 3640-3649.
- [20] LI H CH, XIONG P F, AN J, et al. Pyramid attention network for semantic segmentation [J]. *ArXiv preprint arXiv: 1805.10180*, 2018.
- [21] FU J, LIU J, TIAN H J, et al. Dual attention network for scene segmentation [C]. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019: 3146-3154.
- [22] ZHANG J H, LIN L L, FAN Z J, et al. S³Mars: Semi-supervised learning for mars semantic segmentation [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2024, 62: 1-15.
- [23] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional block attention module [C]. *European Conference on Computer Vision (ECCV)*, 2018: 3-19.
- [24] WANG Q, WU B, ZHU P, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks [C]. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020: 11534-11542.
- [25] CHOLLET F. Xception: Deep learning with depthwise separable convolutions [C]. *IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 1251-1258.
- [26] HU J, SHEN L, SUN G, et al. Squeeze-and-excitation networks [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 42(8): 2011-2023.

作者简介

孟祥源, 本科, 主要研究方向为语义分割、深度学习。

E-mail: 2318343324@qq.com

单慧琳(通信作者), 博士, 副教授, 硕士生导师, 主要研究方向为遥感图像处理、深度学习。

E-mail: shanhuilin@nuist.edu.cn