

基于改进 MCANet-CM 的多模态遥感图像分割算法^{*}

胡峻峰 杨泳

(东北林业大学计算机与控制工程学院 哈尔滨 150040)

摘要:合成孔径雷达图像与可见光图像通过不同的维度采集地表特征,为土地分类研究领域提供高度互补的信息,具有重要的应用价值。然而,现有 MCANet-CM 算法在跨模态特征交互过程中,难以有效捕捉多模态数据中目标轮廓,导致融合特征对复杂场景下目标边界的空间细节表征能力较弱,这使得如何有效结合两类模态数据以实现精准的像素级分类,仍然是目前的关键问题。针对这一问题,本文提出了基于改进 MCANet-CM 的多模态遥感图像语义分割算法。算法中提出了 DyCPCA 注意力机制,该机制通过动态校准通道间的依赖关系,自适应地增强与目标轮廓相关的特征响应,有效提升了模型对多模态数据中细节信息的捕捉能力;同时引入矩形自校准模块,该模块通过构建非对称的感受野结构,增强了模型对不同方向边缘信息的感知能力,显著提高了模型对前景对象的定位精度。通过这两个模块的协同作用,实现了光学数据与 SAR 数据的有效融合。在 WHU-OPT-SAR 数据集上进行实验,相较于基准模型 MCANet-CM,改进模型在平均交并比与平均 F1 分数上分别提高了 2.85% 与 2.81%。与 FTransUNet 等先进算法对比,改进模型同样实现了更好的分割效果。

关键词:合成孔径雷达;可见光图像;MCANet-CM;多模态;语义分割

中图分类号: TP751; TN958 文献标识码: A 国家标准学科分类代码: 520.2060

Multimodal remote sensing image segmentation algorithm based on improved MCANet-CM

Hu Junfeng Yang Yong

(College of Computer Science and Control Engineering, Northeast Forestry University, Harbin 150040, China)

Abstract: Synthetic Aperture Radar and optical images capture surface features through distinct dimensions, providing highly complementary information for land classification research with significant application value. However, existing MCANet-CM algorithms struggle to effectively capture target contours in multimodal data during cross-modal feature interaction, resulting in insufficient spatial detail representation of fused features for object boundaries in complex scenarios. This makes the effective integration of dual-modal data for precise pixel-level classification remain a critical challenge. To address this issue, this paper proposes an enhanced multimodal remote sensing image semantic segmentation algorithm based on improved MCANet-CM. The algorithm introduces the DyCPCA attention mechanism, which dynamically calibrates inter-channel dependencies to adaptively enhance feature responses related to target contours, thereby significantly improving the model's capability to capture fine-grained information from multimodal data. Simultaneously, a Rectangular Self-Calibration Module is incorporated, which constructs an asymmetric receptive field structure to strengthen the model's perception of edge information across different orientations, markedly enhancing localization accuracy for foreground objects. Through the synergistic operation of these two modules, effective fusion of optical and SAR data is achieved. Experiments on the WHU-OPT-SAR dataset demonstrate that compared with the baseline MCANet-CM model, the improved model achieves 2.85% and 2.81% enhancements in mean Intersection over Union and mean F1-score, respectively. When compared with state-of-the-art algorithms like FTransUNet, the proposed model also exhibits superior segmentation performance.

Keywords: SAR; optical image; MCANet-CM; multimodal; semantic segmentation

0 引言

遥感图像语义分割作为像素级分类技术,通过获取每

个像素的语义信息并分配标签的方式实现对地物的精准识别^[1-2]。在标签分配过程中,该技术对不同类别的像素采用不同颜色进行标注,从而形成分割图。目前,遥感图像语义

分割技术在城市规划、环境监测等领域具有重要应用价值^[3-5]。

随着传感技术的不断进步,遥感图像数据的获取方式更加便捷。其中,高分辨率遥感图像包含着丰富的语义信息,但传统方法往往难以充分利用这些信息,导致难以实现理想的分割效果。近些年,深度学习技术不断发展,这一技术的应用为图像分割领域提供了高效的分析手段,推动了遥感图像分割领域的发展。与传统的机器学习方法相比,卷积神经网络(convolutional neural networks, CNN)在遥感图像语义分割任务中表现优异,有效地提升了分割的准确率。其中,全卷积网络(fully convolutional networks, FCN)^[6]使用卷积层替代传统全连接层的方式进行结构改进,可以有效处理任意尺寸的输入图像,并实现了端到端的预测,有效提升了分割效率。同时,FCN 方法也存在缺陷,在 FCN 的池化与下采样操作中,随着特征图尺寸的缩小,网络无法兼顾图像细节,造成了细节的丢失,这使得最终的分割效果在图像细节等方面依然存在着提升的空间。针对这一问题,研究人员提出了以 U-Net^[7]为代表的网络,U-Net 网络在跳跃连接过程中可以实现高层次特征图与低层次特征图的融合,兼顾了图像细节信息与全局信息的提取。同时,随着深度学习的不断发展,应用于分割领域的网络层出不穷。PSPNet^[8]网络通过引用金字塔池化模块,充分提取上下文信息的同时,对不同尺度特征图进行特征融合,提升特征提取能力。提出的 DeepLab^[9]系列网络以 Deeplabv3+ 网络为代表,引入空间金字塔池化模块(atrous spatial pyramid pooling, ASPP),实现多尺度特征提取。HRNet^[10]网络采用并行结构进行多尺度特征融合,实现了全局信息与局部信息的结合,提升了分割精度。目前,单模态分割算法的研究不断推进,但受限于单模态数据的局限性^[11-12],分割效果的提升过程存在较大困难。

随着遥感技术的不断发展,除传统光学图像外,SAR 图像、DSM 图像等单通道图像的获取也不再困难^[13-14]。传统光学图像能够提供高分辨率信息,但在雨雪等恶劣天气条件影响下,采集图像的质量往往难以保证。合成孔径雷达(synthetic aperture radar, SAR)传感器能够克服恶劣天气条件的影响,稳定获取图像数据,同时,SAR 图像也能为光学图像提供互补信息^[15]。因此,以光学图像和 SAR 图像作为输入的多模态遥感分割算法的研究,可以有效利用多模态信息,弥补单模态算法的不足,具有重要的研究意义^[16-18]。

当前,针对多模态遥感数据,主要采用决策级融合与特征级融合^[19-20]两种方法进行结合与分析。决策级融合是指在网络对多模态数据分别进行分类的基础上,对多个分类结果使用加权平均等方法进行处理,得到最后的分割结果。这种方法虽然易于实现,但分割结果往往不够理想。特征级融合以 PSCNN^[21]为例,相比决策级融合更进一步,采用特征堆叠等方式进行特征融合,但依然未能较好地利用多

模态数据信息,在融合方式上依然有较大的提升空间。MCANet-CM^[22]模型使用多模态交叉注意力模块与特征拼接方式,将光学图像与 SAR 图像特征进行结合,在特征提取过程中实现了多模态数据的融合,相比决策级融合与特征级融合,更好地融合了多模态数据特征,得到了更理想的分割结果。然而,MCANet-CM 在多模态数据融合方面虽取得一定成果,但仍存在局限性。该模型在目标轮廓提取上表现欠佳,在特征融合环节,其多模态交叉注意力模块对不同模态数据的特征权重分配较为固定,难以适配复杂的应用场景,致使融合效果受限,部分关键信息在融合过程中丢失;并且在全局上下文信息提取与前景目标定位方面,仅依赖传统的特征提取方式,对前景目标的定位精度不足,难以满足高分辨率遥感图像精确分割等场景的需求。为进一步提升分割效果,本文主要工作如下:

1) 提出基于 MCANet-CM 的改进网络架构,在编码器与解码器结构中分别进行改进,提升特征融合效果与分割精度。

2) 设计注意力模块 DyCPCA^[23]。针对 MCANet-CM 特征权重分配固定的问题,DyCPCA 以通道先验为特征,动态分配注意力权重,提升模块泛化性的同时,有效提升了模型对多模态数据中细节信息的捕捉能力,弥补了 MCANet-CM 在复杂环境下特征融合与轮廓提取的不足。

3) 引入矩形自校准模块(rectangular self-calibration module, RCM)^[24]。鉴于 MCANet-CM 在全局上下文信息提取与前景目标定位上的局限性,RCM 能够在提取全局上下文信息的同时,通过自校准方式灵活调整矩形关注区域,显著提高对前景分类目标的定位精度,从而解决 MCANet-CM 在高分辨率遥感图像分割中定位不准的问题。

1 相关工作

1.1 网络结构

本文使用 MCANet-CM 网络作为基准模型,模型采用了经典的编码器-解码器架构。以 MCANet-CM 网络为基础,本文提出了 MCANet-CMRC 网络。MCANet-CMRC 网络设计并引入了动态通道先验卷积注意力模块(DyCPCA)和矩形自校准模块(RCM),增强了特征提取与重建能力。模块间相互协同,确保了特征图在通道维度和空间维度上的高效表达,有效提升了模型的分割性能,网络结构如图 1 所示。

在编码器中,使用 ResNet101 网络进行特征提取,对输入数据进行四阶段特征提取与下采样。其中, Cov_{OPT}^{low} 与 Cov_{SAR}^{low} 由 ResNet101 网络第一阶段(Layer1)提取, Cov_{OPT}^{high} 与 Cov_{SAR}^{high} 由第 4 阶段提取(Layer4)提取。在 ResNet101 的残差结构中,输入数据经卷积、批归一化后,引入 CBAM 模块强化特征提取,再与原始输入进行残差连接,提升网络对关键特征的捕捉与表达能力。同时,编码器中采用了多模态交叉注意力模块(MCAM)。MCAM 模块能够高效捕捉

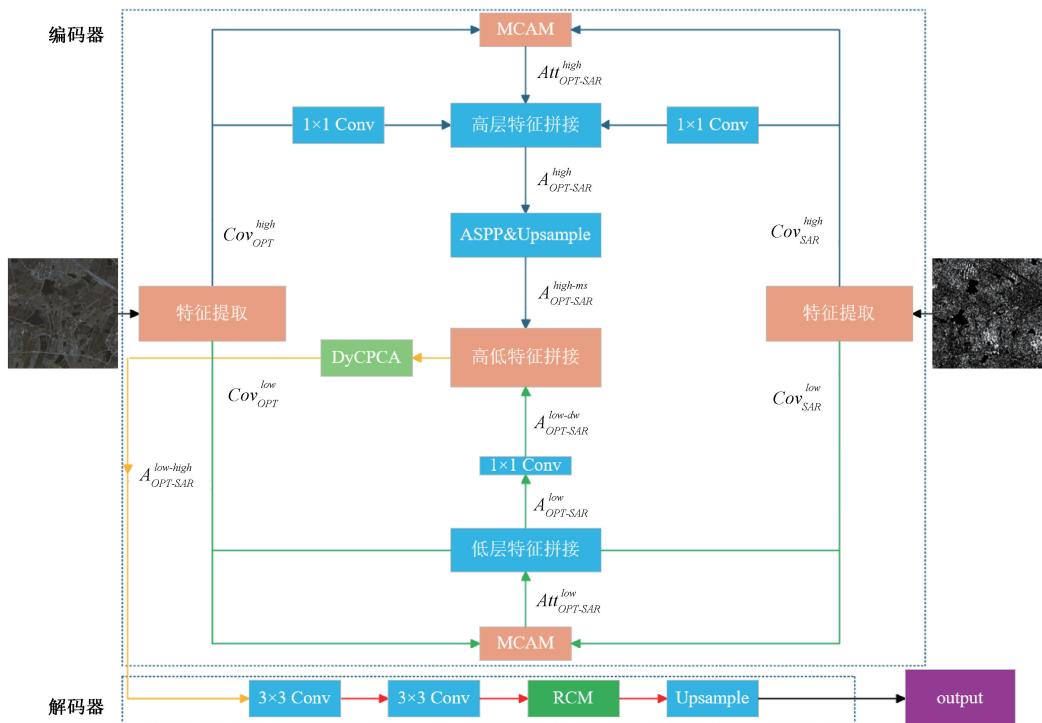


图 1 MCANet-CMRC 网络结构

单一数据源特征图中的位置信息,实现 SAR 图像与光学图像特征图之间的深度交互。MCAM 结构如图 2 所示,其中 Q 、 K 、 V 特征通过 1×1 卷积所提取。将高层特征 Cov_{OPT}^{high} 、 Cov_{SAR}^{high} 进行 1×1 卷积提取与经 MCAM 融合后特征直接拼接,再经 ASPP 模块与上采样获得高层融合特

征。将低层特征 Cov_{OPT}^{low} , Cov_{SAR}^{low} 与经 MCAM 融合后的特征直接拼接, 通过 1×1 卷积得到低层融合特征。高低层融合特征拼接后, 经 DyCPCA 模块处理, 作为编码器输出, 传输至解码器。解码器依次通过 3×3 卷积、RCM 模块与上采样操作, 生成最终分割结果。

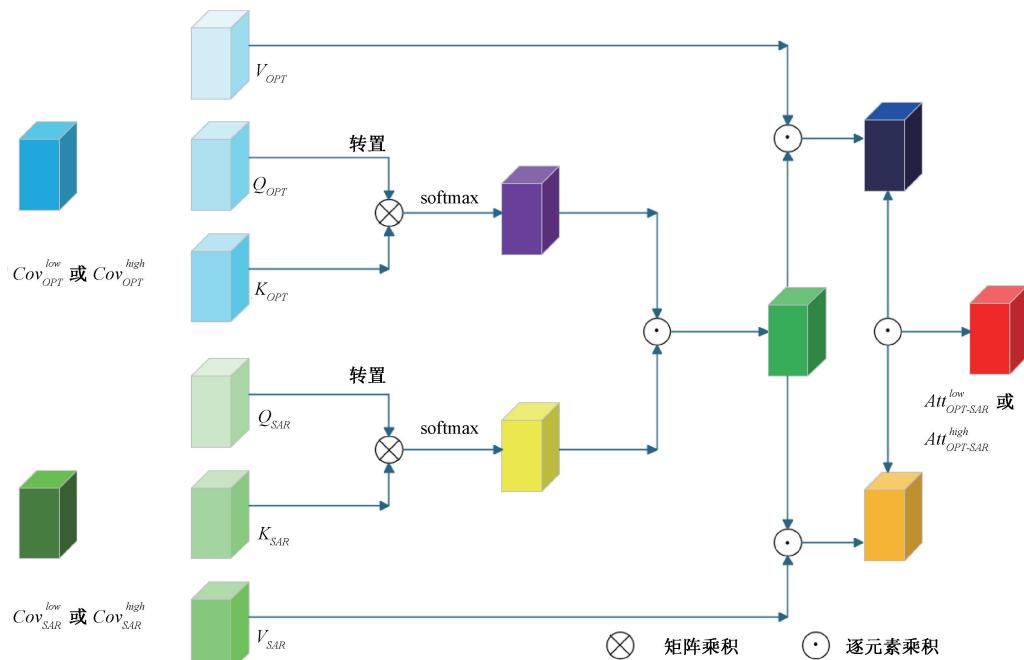


图 2 MCAM 结构
Fig. 2 MCAM structure

1.2 DyCPCA 模块

注意力机制从提出开始,在深度学习领域便发挥着重要的作用。在遥感领域,优秀的注意力机制可以使模型专注于地物目标的提取,在小目标类别的提取与大目标类别边界定位上,注意力机制都凸显出独特的优势。传统注意力机制在提升模型性能的同时,在计算成本与泛用性上存在着不小的提升空间,针对这一需求,本文设计了 DyCPCA 模块。

CPCA 模块对输入特征图并行进行最大、平均池化,池化结果经共享 MLP 处理,输出相加后经 Sigmoid 激活得通道注意力特征图,与输入特征图逐元素相乘获通道先验特征图。通道先验特征图经不同尺寸卷积核的深度

卷积模块,卷积结果相加得空间注意力特征图,经 1×1 卷积处理后,与通道先验特征图逐元素相乘,输出处理后的特征。

DyCPCA 模块在空间注意力设计中使用动态卷积(dynamic convolution)^[25]替换 CPCA 模块中的深度卷积,动态卷积与通道先验同样具有动态调整能力并且动态卷积的低计算量特性,符合轻量化设计目标。多尺度动态卷积提取结构,保证特征细节完整性的同时提升了模块的特征提取能力与泛用性。DyCPCA 模块结构如图 3 所示,实现了模块的强泛用性与轻量化设计,在提升性能的同时避免了传统注意力机制中存在的大量的矩阵运算,有效减轻了计算负担。

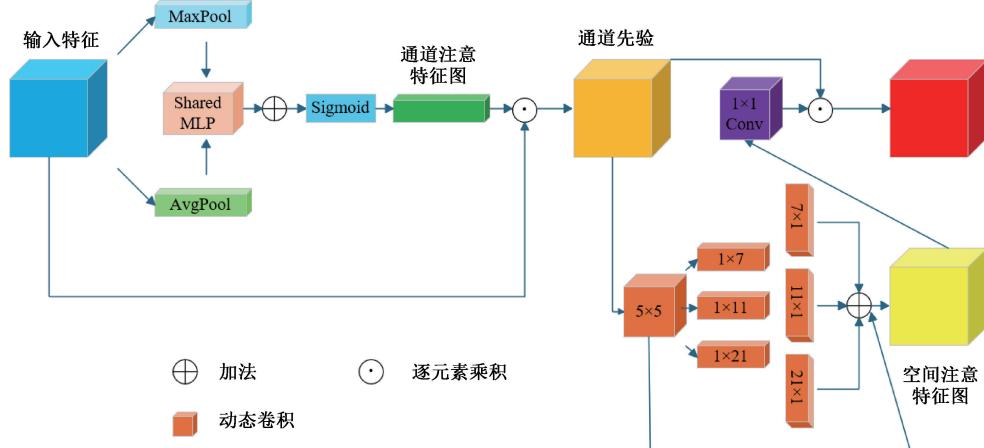


图 3 DyCPCA 结构
Fig. 3 DyCPCA structure

在 ParameterNet 网络中,提出了一种新的卷积方法—动态卷积。动态卷积可以根据输入特征动态调整卷积参数,在增加少量计算量(FLOPs)的基础上,成倍增加了模型的参数量。动态卷积的工作流程可以表示为:

$$Y = X * W' \quad (1)$$

$$W' = \sum_{i=1}^M \alpha_i W_i \quad (2)$$

其中, M 代表动态专家的数量,专家数量作为模块参数可自行设置, W_i 代表第 i 个卷积的权重参数, W' 代表动态卷积的权重参数, α_i 代表对应的动态系数。式(1)中,对输入特征 X ,使用动态卷积进行卷积运算。对于每个输入特征 X ,首先通过全局平均池化处理,然后通过以 softmax 为激活函数的 MLP 层生成动态系数 α_i ,过程如式(3)所示。

$$\alpha = \text{softmax}(\text{MLP}(\text{Pool}(X))) \quad (3)$$

动态卷积与标准卷积的参数量比值与浮点数计算量(FLOPs)比值的计算过程如式(4)和(5)所示。其中, K 代表卷积核的大小, C_{in} 和 C_{out} 对应输入特征图的通道数与输出特征图的通道数, H' 和 W' 分别代表输出特征图的高度和宽度。

$$R_{param} = \frac{C_{in}^2 + C_{in}M + MC_{out}C_{in}K^2}{C_{out}C_{in}K^2} = \frac{C_{in}}{C_{out}K^2} + \frac{M}{C_{out}K^2} +$$

$$M \approx \frac{1}{K^2} + M \quad (M \ll C_{out}K^2, C_{in} \approx C_{out}) \quad (4)$$

$$R_{flops} = \frac{C_{in}^2 + C_{in}M + MC_{out}C_{in}K^2 + H'W'C_{out}C_{in}K^2}{H'W'C_{out}C_{in}K^2} = \frac{C_{in}}{H'W'C_{out}K^2} + \frac{M}{H'W'C_{out}K^2} + \frac{M}{H'W'} + 1 \approx 1 \quad (1 < M \ll H'W', C_{in} \approx C_{out}) \quad (5)$$

动态卷积的参数大约是标准卷积的 M 倍,但浮点数计算量几乎一致,在不影响计算效率的前提下,有效地提升了性能。同时,由于动态卷积的卷积核参数是根据每个输入特征进行动态选择的,使得动态卷积可以更好地适应不同的输入特征,提高了模型的泛化能力。

1.3 RCM 模块

RCM 模块通过全局上下文信息提取和空间特征重建的方式提升语义分割的准确性。RCM 整体结构如图 4 所示,模块通过建立矩形关注区域并动态调整注意形状,增强了对前景对象的定位和识别能力。

RCM 模块由矩形自校准注意力(residual cross-attention,

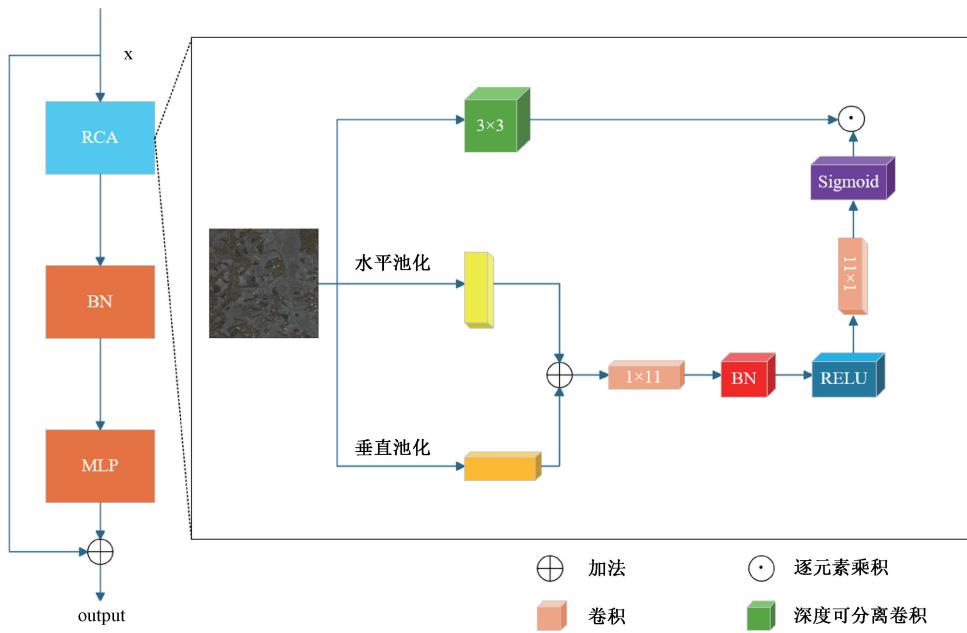


图 4 RCM 结构

Fig. 4 RCM structure

RCA), 归一化与多层感知机(multilayer perceptron, MLP)三部分组成。RCA 是模块的核心环节, 通过池化方法实现水平方向与垂直方向的双向上下文信息提取, 将提取信息相加, 建立矩形关注区域。RCA 进一步通过水平与垂直条带状卷积设计自校准函数, 生成注意力特征。由于条带状卷积的参数具有学习特性, 使关注区域更接近前景对象的同时可以有效适应不同尺度的输入特征。RCA 最后设计了特征融合功能, 使用深度可分离卷积提取输入特征, 将提取特征与注意力特征进行逐元素相乘, 得到最终的输出特征。将输出特征通过归一化与多层感知机对特征进一步处理, 强化特征的表达。RCM 模块的设计方式最终实现了全局上下文信息的提取与特征重建。

在语义分割任务中, RCM 模块凭借对前景对象的优异聚焦能力, 在遥感数据集上有着良好表现。针对遥感任务中难以解决的小目标定位与大目标边界划分问题, RCM 模块能够精确定位多种尺度的目标, 提升模型对前景对象的提取能力, 进而提升分割精度。

2 实验结果与分析

2.1 实验环境与评估指标

1) 实验环境

实验所用配置: 操作系统: Ubuntu 22.04; GPU: NVIDIA RTX 4090 24 GB; CPU: i7-13700k。模型训练环境: Python3.8, Pytorch2.4.1。

训练参数设定为: 训练轮次为 105, Batch-size=8, 学习率为 0.001, 输入图像为 256 pixel×256 pixel。

2) 评估指标

为评估模型的语义分割性能, 实验共使用了 3 个评估指标: 总体准确率 (overall accuracy, OA)、平均交并比 (mean intersection over union, mIoU)、平均 F1 分数 (mean F1-score, mF1)。

OA 表示正确分类的样本数占总样本数的比例, 计算公式如下:

$$OA = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

mIoU 和 mF1 是衡量分割性能最常用的两个指标, 计算公式如下:

$$IoU_k = \frac{TP}{TP + FP + FN} \quad (7)$$

$$F1_k = \frac{2TP}{2TP + FP + FN} \quad (8)$$

$$mIoU = \frac{1}{K} \sum_{k=1}^K IoU_k \quad (9)$$

$$mF1 = \frac{1}{K} \sum_{k=1}^K F1_k \quad (10)$$

其中, k 代表实验数据集的类别数量, TP 、 TN 、 FP 和 FN 分别代表真正例、真负例、假正例、假负例的元素数量。

2.2 实验数据集

本文实验数据采用了 WHU-OPT-SAR 数据集, 图像数据采集于湖北省, 包含 100 张高分辨率光学图像与 SAR 图像, 图像分辨率为 5 556×3 704。其中, 光学图像为四通道图像, 包含 RGB 通道与近红外通道, SAR 图像为单通道图像。数据集包含农田、城市、村庄在内的 8 个语义类别, 并配有详细的像素级标注, 适用于多模态遥感图像分割任务。

务的研究。实验中,将数据集按照 3:1:1 的比例划分为训练集、测试集与验证集。

2.3 实验结果与分析

1) 网络改进对比实验

MCANet-CMRC 网络融合了 DyCPCA 注意力机制和 RCM 模块,增强了特征提取的能力,有效地提升了多模态数据的融合效果。DyCPCA 模块在 CPCPA 注意力机制的基础上进行了改进,通过引入动态卷积,提升了对输入特征的感知能力,能够根据输入数据的特征进行动态的权重调整,使 DyCPCA 模块能够高效处理多模态输入数据,精准捕捉特征信息,优化分割效果。

为验证 DyCPCA 模块的有效性,以 MCANet-CM 网络为基准模型,以 WHU-OPT-SAR 为实验数据集,进行两组对比实验,对比不同注意力机制与改进方法的效果。第 1 组实验,将 DyCPCA 模块与 CPCPA 模块及其他先进注意力机制进行对比,结果如表 1 所示。在引用 DyCPCA 模块后,网络的分割指标 OA、mIoU、mF1 分别达到了 82.62%、54.93%、68.80%,与原 CPCPA 模块及 MCA^[26]、SCSA^[27] 等先进注意力相比,引用 DyCPCA 模块,可以使网络的分割性能得到更大的提升,验证了模块的有效性。第 2 组实验,采用不同类型的卷积对 CPCPA 模块改进,结果如表 2 所示。实验结果表明,与 PConv 等^[28-29] 先进卷积相比,使用动态卷积的改进方法,实现了最好的模型性能,验证了改进方法的有效性。

表 1 注意力改进对比实验

Table 1 Comparison experiment of attention improvement
%

模型	OA	mIoU	mF1
Base	81.75	53.50	67.34
+CPCA	82.59	54.60	68.25
+MCA	81.77	53.29	67.07
+SCSA	82.06	53.76	67.51
+MSCA	81.69	52.93	66.74
+PCAA	79.62	50.04	64.04
+DyCPCA	82.62	54.93	68.80

表 2 卷积改进对比实验

Table 2 Comparison experiment of convolution improvement
%

模型	OA	mIoU	mF1
Base	81.75	53.50	67.34
+CPCA	82.59	54.60	68.25
+CPCA(PConv)	81.68	52.88	66.56
+CPCA(ScConv)	81.62	52.14	65.57
+DyCPCA	82.62	54.93	68.80

2) 消融实验

为验证 MCANet-CMRC 网络各结构的有效性,以 MCANet-CM 为基准网络,针对 DyCPCA 和 RCM 模块进行消融实验,在 WHU-OPT-SAR 数据集上进行验证。消融实验结果如表 3 所示,每种改进方法均不同程度地提升了模型的分割性能。

表 3 消融实验

Table 3 Ablation experiment

%

DyCPCA	RCM	OA	mIoU	mF1
—	—	81.75	53.50	67.34
✓	—	82.62	54.93	68.80
—	✓	82.45	54.87	68.77
✓	✓	83.06	56.35	70.15

融合 DyCPCA 模块使模型在 OA、mIoU、mF1 三个分割指标上,相较于基准网络分别提升了 0.87%、1.43%、1.46%。DyCPCA 模块能够有效提升模型对不同类别像素的分类能力和语义信息的捕捉能力,从而增强整体分割性能。引用 RCM 模块,同样提升了网络的性能,分割指标在基准网络的基础上分别提升了 0.7%、1.37%、1.43%。RCM 模块通过优化前景物体的位置建模和上下文信息提取,提升了模型对复杂场景的处理能力。

MCANet-CMRC 网络在 OA、mIoU、mF1 指标上,相较于基准网络,分别提升了 1.31%、2.85%、2.81%。实验结果表明,融合 DyCPCA 模块和 RCM 模块的网络结构在综合性能上达到了更优水平。二者的协同作用实现了整体性能的优化,使模型在多模态任务场景下表现出良好的性能。

3) 对比实验与分割结果可视化

为验证 MCANet-CMRC 模型改进的有效性,本小节设计对比实验,选取先进的单模态与多模态算法进行对比,在 WHU-OPT-SAR 数据集上评估各网络的分割性能,对比实验结果如表 4 所示。

单模态算法如 PSPNet、DeepLabv3+ 在语义分割任务中表现良好的性能。PSPNet 通过空间金字塔池化能够有效捕捉多尺度上下文信息,DeepLabv3+ 则结合了空间金字塔模块和解码器-编码器结构,能够有效地关注目标的边界信息。然而,单模态算法无法充分利用多模态数据的互补性,在复杂场景中难以实现准确的分割效果。

多模态算法 CMFNet^[30] 与 FTransUNet^[31] 在网络结构设计中,均采用 U 型结构。CMFNet 网络针对跳跃连接结构进行改进,在跳跃连接中引入 Transformer 结构提升多尺度融合能力。FTransUNet 网络在编码器结构中结合了卷积网络与 Transformer 结构,同时提出以 SE 注意力为基础创新的特征融合模块,准确提取局部细节信息与全局语义信息。在 Vaihingen 与 Potsdam 的多模态数据集

表 4 对比实验
Table 4 Comparative experiment

模型	IoU							OA	mIoU	mF1	%
	农田	城市	村庄	水域	森林	道路	其他				
Base	67.88	55.75	48.04	67.41	80.48	37.26	17.67	81.75	53.50	67.34	
U-Net	66.10	55.24	46.19	66.51	79.19	34.08	19.26	80.77	52.37	66.48	
PSPNet	64.77	55.96	40.75	62.20	78.44	24.94	18.60	79.47	49.38	63.46	
DeepLabv3+	66.81	56.16	46.58	65.79	79.18	34.73	18.92	80.88	52.59	66.67	
HRNet	68.33	57.40	48.06	67.63	80.12	38.43	21.25	81.82	54.46	68.47	
CMFNET	69.23	59.55	49.62	68.91	80.93	38.38	20.35	82.58	55.28	69.02	
FTransunet	67.56	58.38	47.17	67.06	80.03	34.97	17.32	81.43	53.21	67.00	
MCANet-CMRC	69.90	58.91	50.08	70.12	81.31	40.82	23.30	83.06	56.35	70.15	

上,两种算法均取得了优异效果,但在 WHU-OPT-SAR 数据集上的分割效果仍需提升。

本文的分割结果如图 5 所示,通过四组实验样本的可视化展示,可以清晰观察到 MCANet-CMRC 网络的分割性能优势。在第 1 组与第 2 组分割图中,MCANet-CMRC 网络可以

清晰勾勒出目标的轮廓,在道路这类小比例目标上,优势更为明显,展现出网络优秀的细节捕捉能力。在第 3 组与第 4 组分割图里,MCANet-CMRC 网络同样表现出色,有效降低了目标错误分类的概率,同时实现了对小目标的精准定位,进一步凸显了模型在图像分割任务中的优越性与稳定性。

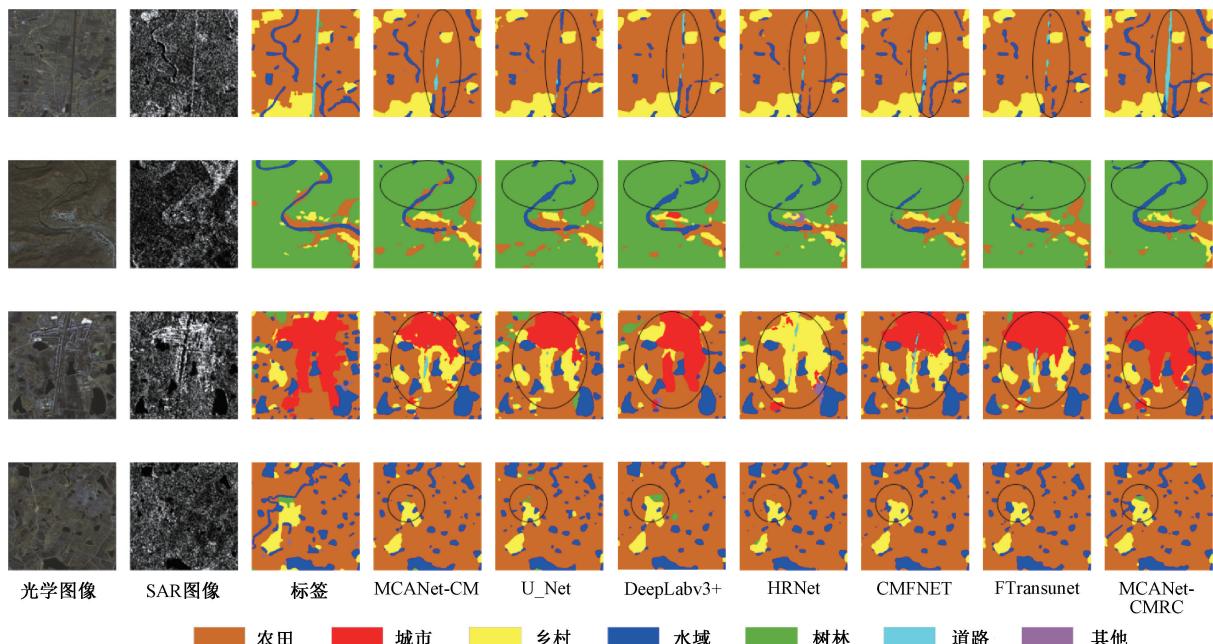


图 5 可视化结果图

Fig. 5 Visualization result plots

3 结 论

本研究从多模态遥感图像分割算法设计的角度出发,以提升模型分割性能为目标,提出了 MCANet-CMRC 网络。MCANet-CMRC 网络在编码结构中设计并引入了 DyCPCA 注意力机制,使用动态卷积对 CPCA 注意力进行改进,增强泛用性的同时提升了模型对重点区域的关注能力。网络在解码结构中引入矩形自校准模块(RCM),该模块在多种语义分割任务上表现出优异的性能,在多模态遥

感任务中,该模块的引入使模型更精准地捕捉上下文信息,提高了模型对前景对象的定位能力。

以 WHU-OPT-SAR 数据集作为实验数据集,检验模型效果,与基准模型相比,MCANet-CMRC 网络在 mIoU 和 mF1 指标上分别提高了 2.85% 和 2.81%。针对多模态数据间的差异问题,提出的改进方法协同作用,有效地提升了分割精度,实现了更精准的像素级分类。

本文提出的改进网络在实现分割效果提升的同时,在模型结构与泛用性上仍有优化空间。今后的研究将致力

于模型分割性能的进一步提升,同时对模型在多模态遥感任务中的泛用性进行优化。

参考文献

- [1] 胡宇翔,余长宏,高明.多模态融合的遥感图像语义分割网络[J].计算机工程与应用,2024,60(15):234-242.
HU Y X, YU CH H, GAO M. Remote sensing image semantic segmentation network with multimodal fusion [J]. Computer Engineering and Applications, 2024, 60(15): 234-242.
- [2] 李智杰,程鑫,李昌华,等.跨模态多层特征融合的遥感影像语义分割[J].计算机科学与探索,2025,19(4):989-1000.
LI ZH J, CHENG X, LI CH H, et al. Semantic segmentation of remote sensing images with cross-modal multi-layer feature fusion [J]. Journal of Frontiers of Computer Science and Technology, 2025, 19(4): 989-1000.
- [3] 李忠伟,王晓政,罗偲,等.基于门控金字塔融合的遥感影像海陆分割方法[J].电子测量技术,2023,46(15):111-117.
LI ZH W, WANG X ZH, LUO S, et al. Remote sensing image land-sea segmentation method based on gated pyramid fusion [J]. Electronic Measurement Technology, 2023, 46(15): 111-117.
- [4] KOTARIDIS I, LAZARIDOU M. Remote sensing image segmentation advances: A meta-analysis [J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2021, 173: 309-322.
- [5] LIU B H, DING J L, ZOU J, et al. LDANet: A lightweight dynamic addition network for rural road extraction from remote sensing images[J]. Remote Sensing, 2023, 15(7): 1829.
- [6] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2015: 3431-3440.
- [7] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation [C]. Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015: 18th International Conference, Springer International Publishing, 2015: 234-241.
- [8] 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.
- [9] CHEN L C, ZHU Y K, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]. European Conference on Computer Vision(ECCV), 2018: 801-818.
- [10] SUN K, XIAO B, LIU D, et al. Deep high-resolution representation learning for human pose estimation[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 5693-5703.
- [11] 张银胜,单梦姣,陈昕,等.基于多模态特征提取与层级感知的遥感图像分割[J].地球信息科学学报,2024,26(12):2741-2758.
ZHANG Y SH, SHAN M J, CHEN X, et al. Remote sensing image segmentation based on multimodal feature extraction and hierarchical perception [J]. Journal of Geo-information Science, 2024, 26 (12): 2741-2758.
- [12] 孙汉淇,潘晨,何灵敏,等.多模态特征融合的遥感图像语义分割网络[J].计算机工程与应用,2022,58(24):256-264.
SUN H Q, PAN C H, HE L M, et al. Semantic segmentation network for remote sensing images with multimodal feature fusion[J]. Computer Engineering and Applications, 2022, 58(24): 256-264.
- [13] HU W, WANG X H, ZHAN F, et al. Opt-sar-ms2net: A multi-source multi-scale siamese network for land object classification using remote sensing images[J]. Remote Sensing, 2024, 16(11): 1850.
- [14] LI W B, SUN K M, LI W ZH, et al. Aligning semantic distribution in fusing optical and SAR images for land use classification [J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2023, 199: 272-288.
- [15] LIU J, GONG M G, QIN K, et al. A deep convolutional coupling network for change detection based on heterogeneous optical and radar images[J]. IEEE Transactions on Neural Networks and Learning Systems, 2016, 29(3): 545-559.
- [16] WANG Y Y, ZHANG W G, CHEN W D, et al. MFFnet: Multimodal feature fusion network for synthetic aperture radar and optical image land cover classification[J]. Remote Sensing, 2024, 16(13): 2459.
- [17] XU D D, LI ZH, FENG H, et al. Multi-scale feature fusion network with symmetric attention for land cover classification using SAR and optical images [J]. Remote Sensing, 2024, 16(6): 957.
- [18] WANG Y Y, ZHANG W G, CHEN W D, et al. Bsdnet: Dual-stream feature extraction network based on segment anything model for synthetic aperture radar land cover classification[J]. Remote Sensing, 2024, 16(7): 1150.

- [19] ZHU X X, TUIA D, MOU L CH, et al. Deep learning in remote sensing: A comprehensive review and list of resources[J]. *IEEE Geoscience and Remote Sensing Magazine*, 2017, 5(4): 8-36.
- [20] XU X D, LI W, RAN Q, et al. Multisource remote sensing data classification based on convolutional neural network[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2017, 56(2): 937-949.
- [21] MOU L CH, SCHMITT M, WANG Y Y, et al. Identifying corresponding patches in SAR and optical imagery with a convolutional neural network[C]. 2017 IEEE International Geoscience and Remote Sensing Symposium(IGARSS). IEEE, 2017: 5482-5485.
- [22] LI X, ZHANG G, CUI H, et al. MCANet: A joint semantic segmentation framework of optical and SAR images for land use classification [J]. *International Journal of Applied Earth Observation and Geoinformation*, 2022, 106: 102638.
- [23] HUANG H J, CHEN Z G, ZOU Y, et al. Channel prior convolutional attention for medical image segmentation [J]. ArXiv preprint arXiv: 2306.05196, 2023.
- [24] NI ZH L, CHEN X H, ZHAI Y J, et al. Context-guided spatial feature reconstruction for efficient semantic segmentation [C]. European Conference on Computer Vision, Cham: Springer Nature Switzerland, 2024: 239-255.
- [25] HAN K, WANG Y H, GUO J Y, et al. ParameterNet: Parameters are all you need for large-scale visual pretraining of mobile networks[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 15751-15761.
- [26] YU Y, ZHANG Y, CHENG Z Y, et al. MCA: Multidimensional collaborative attention in deep convolutional neural networks for image recognition[J]. *Engineering Applications of Artificial Intelligence*, 2023, 126: 107079.
- [27] SI Y ZH, XU H Y, ZHU X ZH, et al. SCSA: Exploring the synergistic effects between spatial and channel attention[J]. *Neurocomputing*, 2025: 129866.
- [28] CHEN J R, KAO S H, HE H, et al. Run, don't walk: Chasing higher FLOPS for faster neural networks[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 12021-12031.
- [29] LI J F, WEN Y, HE L H. Scconv: Spatial and channel reconstruction convolution for feature redundancy[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 6153-6162.
- [30] MA X P, ZHANG X K, PUN M O. A crossmodal multiscale fusion network for semantic segmentation of remote sensing data[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2022, 15: 3463-3474.
- [31] MA X P, ZHANG X K, PUN M O, et al. A multilevel multimodal fusion transformer for remote sensing semantic segmentation[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2024, 62: 5403215.

作者简介

胡峻峰(通信作者),副教授,硕士生导师,主要研究方向为机器视觉、图形处理、模式识别与智能控制。

E-mail:nefuhjf@126.com

杨泳,硕士研究生,主要研究方向为计算机视觉与图像处理。

E-mail:yu258377@163.com