文章编号：1674-2974（2025）04-0044-13 DOI：10.16339/j.cnki.hdxbzkb.2025265

顾及小尺度目标特征重建的全局语义分割模型

吴小所[1†]，乔煜栋[1]，贺成龙[1]，刘小明[2]，闫浩文[1]
（1. 兰州交通大学 电子与信息工程学院，甘肃 兰州 730070；
2. 青海理工学院 工学院，青海 西宁 810016）

摘 要：针对复杂背景下航空遥感图像中多类别小尺度目标特征的理解困难和特征边界分割不清晰的问题，本研究构建了一种新型的分割模型，该模型通过综合主干网络特征并进行特征分类与重构来提升分割效果.模型以Swin-Transformer作为基础编码结构，利用其强大的全局语义信息捕捉能力进行特征抽取.进一步，本研究创新性地提出了信息聚合重构模块（IGRM）和通道区分重构模块（CRRM），这两种结构能够依据信息量对抽取的特征进行分类和重构，以此细化了对小尺度目标特征的处理.模型结合了上采样与下采样的特征连接，并将重构特征与编码器特征融合，形成多尺度特征聚合块，进而输出精确的分割结果.在处理复杂背景下的多目标场景时，本模型能够对细小尺度目标特征进行精确重构，生成高分辨率的分割图像，显著提升了分割的准确度.在ISPRS Potsdam和ISPRS Vaihingen数据集上，本模型取得了平均交并比（mIoU）分别为87.15%和82.93%、整体准确率（OA）分别为91.53%和91.4%的优异表现.为评估模型对多类别小尺度目标特征提取的泛化性能，本文还进行了针对复杂背景下小车类别的对比实验，在UAVid数据集上的mIoU达到了67.86%.

关键词：航空遥感；语义分割；信息聚合重构模块；通道区分重构模块；整合上采样
中图分类号：TP751；TP391.41 **文献标志码**：A

Global Semantic Segmentation Model Considering Small-scale Target Feature Reconstruction

WU Xiaosuo[1†]，QIAO Yudong[1]，HE Chenglong[1]，LIU Xiaoming[2]，YAN Haowen[1]
（1. School of Electronic and Information Engineering，Lanzhou Jiaotong University，Lanzhou 730070，China；
2. Engineering School，Qinghai Institute of Technology，Xining 810016，China）

Abstract：To solve the problems of insufficient understanding and fuzzy feature boundary segmentation of multi-target categories small-scale feature semantic information in aerial remote sensing images in complex background，this paper designs a segmentation model that integrates the features of the backbone network information and classifies and reconstructs the features to improve the segmentation effect. The model takes Swin-Transformer as the coding structure and utilizes its ability to understand global semantic information for feature extraction. The

* **收稿日期**：2024-04-30
基金项目：国家重点研发计划资助项目（2022YFB3903604），National Key Research and Development Program of China（2022YFB3903604）；甘肃省自然科学基金资助项目（21JR7RA310），Gansu Provincial Natural Science Foundation（21JR7RA310）；兰州交通大学青年科学基金资助项目（2021029），Lanzhou Jiaotong University Youth Science Foundation（2021029）
作者简介：吴小所（1981—），男，甘肃兰州人，兰州交通大学副教授，博士后
† 通信联系人，Email：wuxs_laser@lzjtu.edu.cn

segmentation of small-scale target features is refined by the designed information grouping reconstruction convolution (IGRM) and channel classification reconstruction convolution (CRRM), which classify and reconstruct the extracted features by the amount of information. Finally, by integrating the up-sampling and down-sampling connections, the reconstructed features are fused with the features extracted by the encoder to form a multi-scale feature aggregation block to output the segmentation results. The refined reconstruction of small-scale target features is realized in multi-target scenarios with complex backgrounds, and high-quality segmentation maps are generated to improve the segmentation accuracy. Experimental results on the ISPRS Potsdam and ISPRS Vaihingen datasets show that the average intersection and merger ratio (mIoU) is 87.15% and 82.93%, respectively, and the overall accuracy (OA) is 91.53% and 91.4%, respectively. To verify the generalization ability of the model for small-scale target feature extraction in multi-target categories, this paper also designs a comparative experiment for the category of carts in complex backgrounds. The experimental results show that the mIoU on the UAVid dataset reaches 67.86%.

Key words: aerial remoteness; semantic segmentation; information grouping reconstruction module (IGRM); channel classification reconstruction module (CRRM); integrated up-sampling

在现代计算机视觉领域,图像分割是核心任务之一,旨在根据图像内的类别对像素进行精确分类.随着航空遥感技术[1]的迅猛进步,研究者发现,语义分割技术在处理背景复杂的遥感图像时,具有广泛的应用潜力.尽管如此,这类航空遥感图像往往要面对各类地物目标在尺度上的不一致性以及光谱特性的混淆问题,特别是在小尺度目标[2]的分割上存在挑战.这些小尺度目标如汽车、树木和部分建筑,往往只占据几个像素,且与背景的对比度较低,容易被忽视.现有的分割模型在提取复杂背景下的小尺度目标特征[3]方面仍存在上述问题,这限制了语义分割模型在航空遥感领域的潜力.

在分割模型技术演进的初期阶段,Long等科研人员开发出了全卷积网络[4](简称FCN),该网络在传统卷积神经网络(CNN)的基础上实现了从原始图像到分割结果的直接映射.这一创新允许对任意尺寸的图像进行处理并输出对应尺寸的分割图.FCN通过上采样技术扩大特征图的尺寸,从而为图像分割提供了语义层面的基础.尽管如此,FCN在处理较小尺度地物时,由于感受野的限制,往往难以获取细节信息.在FCN模型取得显著成效后,Chen等研究者对其进行了优化,提出了DeepLab[5]分割模型.DeepLab结合了FCN与扩张卷积技术,既扩展了感受野,也保障了分割的准确性,使其在处理较小尺度特征时更具优势.尽管如此,DeepLab在理解全局信息方面依旧存在一定的局限性.随着深度学习技术和硬件设备的进步,多种适用于航空遥感中小尺度地物语义分割的新模型与方法不断涌现,DeepLabV3+[6]便是基于DeepLab的进一步优化版本.DeepLabV3+通过融合扩张卷积与全局平均池化结构,以及采用不同膨胀率的扩张卷积,进一步提升了模型对小尺度特征的辨识能力.然而,网络加深可能导致特征细节的丢失.DANet[7]则将模型结构进行轻量化设计,通过编解码结构及自主注意力机制,提高了对小尺度特征的理解,但缺乏对全局语义的捕捉.BiSeNet[8]采用双分支的编解码器结构,通过全局和局部分支有效地联系上下文信息,提升了模型对细节的捕获能力.虽然这些方法在航空遥感图像的多种应用场景中表现出了较高的分割性能,但传统分割模型在处理复杂背景中的小尺度特征时,仍然面临全局语义理解能力不足和冗余特征提取等问题,进而影响了分割的精度和效率.近年来,Transformer被引入语义分割领域的研究引起了广泛关注.虽然Transformer在语义分割等任务中表现优异,但在小尺度地物的航空遥感数据集上,其分割精度仍有提升空间,尤其在细节恢复方面存在不足.这主要是因为在图像向序列转换的过程中,部分结构性细节信息的不完整性使得上采样环节难以精确地还原图像的细腻部分.尽管Transformer模型在捕捉全局特性方面表现出色,但在应对不同尺度的地物时,往往未能充分考虑到各类地物之间的重要联系.这些联系在精确辨别小尺度地物方面尤为关键.因此,有必要对现有模型进行优化,增强其对于多目标场景中小尺度信息的感知与运用,以保证图像分割的精确度

得到提升.

借鉴相关研究思路和问题分析,旨在充分利用整体信息,并显著提升在复杂背景中对于微小目标的提取准确性,本文设计的模型在编码器部分采用Swin-Transformer[9]的分层多头注意力机制与滑动窗口注意力机制,以捕获全局上下文并进行特征提取.Transformer Block作为骨干网络的核心组件,实现了全局特征的交流与整合.本文提出了一种重构策略,该策略将特征图细分为若干个子特征语义图,并对这些子特征语义图实施重构处理,旨在削减冗余属性并产出更加精炼的深层语义,这些信息精准反映了文本的核心内容和内在语义.最终,借助跨维度映射机制,加强各模块间的互动,弥补信息缺失.

1 本文模型

本文构建了一个新型的语义分割网络架构—Swin-ERRM(通过重构模块提取各阶段特征并恢复分辨率),该模型通过融合Swin-Transformer在不同阶段捕获的语义信息,重新构建模块来增强特征图的清晰度.模型的总体布局如图1所示,依照编解码器框架搭建.在此框架中,编码器部分使用Swin-Transformer作为骨干网络,充分发挥其在全局语义信息捕获和特征提取方面的优势.在解码器部分,设计了信息聚合重构(IGRM)和通道区分重构(CRRM)模块,以细化局部小尺度目标的特征表达,并提升多目标边界的分割精度.这些模块在经典线性注意力机制[10,11]的基础上进行了优化,使模型能够更加精细地捕捉并恢复复杂场景中的多尺度特征.该设计旨在强化网络对细小目标及多目标边界的语义特征表达能力,同时优化多尺度复杂场景的分割准确性.

本文提出了信息聚合重构模块(IGRM)以及通道区分重构模块(CRRM).在此基础上,采用多尺度映射策略(包括合并上采样和下采样连接),有效融合了特征的多层级特性,从而优化了输出特征的多尺度表达能力.Swin-ERRM模型的设计初衷是为了应对语义分割任务中对小目标和复杂多目标场景的挑战.通过整合Swin-Transformer各阶段提取的全局语义信息,并利用重构模块恢复特征[10-11]的分辨率,细化局部小尺度目标的特征表达,提升多目标边界的分割精度.此外,通过融合上采样与下采样路径的多尺度投影,有效集成了不同尺度的特征信息,显著提升了分割结果的多维度表征效果.

Swin-ERRM的设计旨在增强模型对全局和局部较小目标特征信息的理解,实现高效且精确的语义分割.在编码器中,通过多个Transformer Block的堆叠处理输入特征,并运用其多头自注意力运算点积计算特征,以此提升整体语义结构的构建效率.该模型构建了四个不同层次、不同尺寸的特征图,每一层都先进行下采样处理.在解码阶段,在主干网络

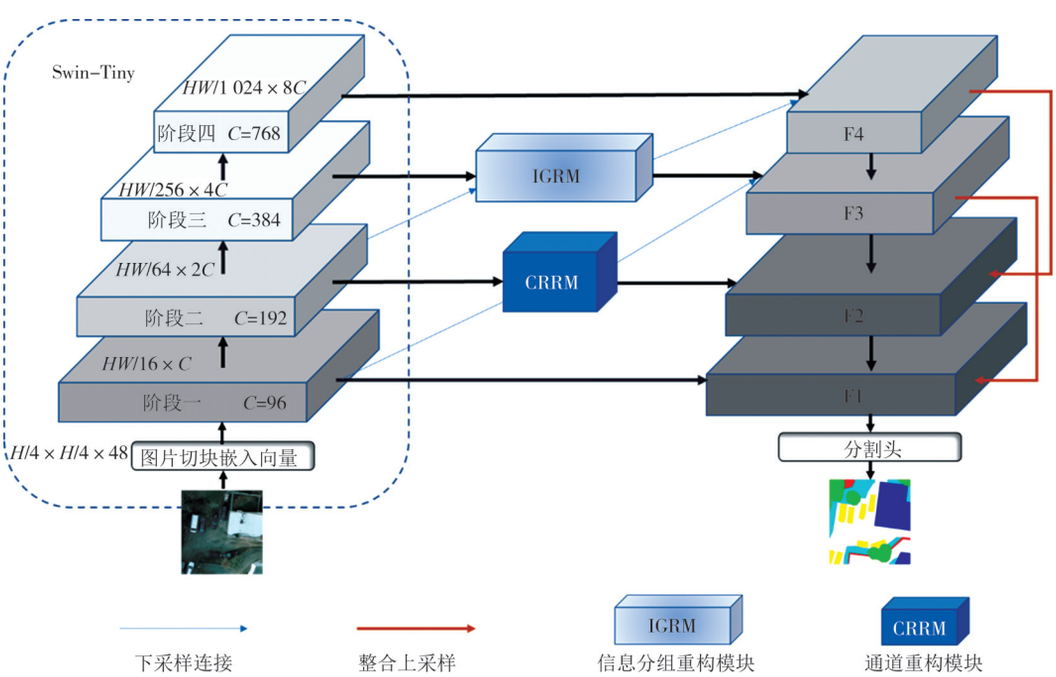


图1 整体模型结构图

Fig.1 Overall model structure diagram

里,对四个 Transformer 进行了处理,在 Block 阶段收集到的特征被转化为四个不同的特征区域,这些区域分别被标记为 F1、F2、F3 和 F4. 然后,利用设计的特征重构模块进行重建并补充细节信息,该单元对这些特征区域进行了细致的调节和增强,最后将其整合为一个综合特征,并通过头部分割生成最终的特征输出,以此强化模型对数据内在关联的处理能力,进而提高分割任务的性能.

1.1　信息聚合重构模块(IGRM)

信息聚合重构模块(information grouping reconstruction module, IGRM)的核心思想在于通过有效地分离和重构特征,以减少航空遥感图像分割中的冗余特征,从而提高分割结果的质量. 在航空遥感图像中,受多个波段光谱的影响,以及相邻像素间的相关性[12],常常会出现特征冗余并干扰不同尺度特征的分割,这严重影响了模型的性能. IGRM 模块通过分组归一化[13]的缩放因数实现权重的分离,将特征分成多个子集,采用 Sigmoid 函数对子特征集的权重进行转换,确保权重值落在(0,1)区间内. 随后,模型通过设定阈值进行筛选,并结合特征重构技术,有效去除空间维度的冗余信息,从而提升特征的表现力.

此外,IGRM 还引入了空洞空间金字塔池化模块(aisle spatial pyramid pooling, ASPP[14]),进一步提升了多尺度信息的融合效果,优化了特征表示. 通过这种设计,IGRM 能够有效分离信息丰富的特征与信息较少的特征,从而在保留重要信息的同时减少冗余,提升模型在复杂环境中的分割性能. IGRM 模型通过特征分离、权重映射与阈值筛选有效减少特征冗余,保留重要信息. 其特征重建机制增强了有效特征的表达能力,而 ASPP 模块的引入实现了多尺度信息的融合,进一步提升了模型对复杂环境的适应能力. 整体而言,IGRM 的结构设计确保了信息的有效捕捉与表达,从而显著提高了航空遥感图像的分割性能.

图 2 呈现了 IGRM 结构框架,在初始阶段,模块接收了来自主网络第二、三阶段输出的多维度特征 X. 紧接着,通过实施分组归一化(GN[15])来评估各个特征图的丰富度. 对多维度特征 X 通过 GN 层进行通道分类处理,其中 C 表示通道的总数. 利用可训练参数 γ 对每个特征通道的语义信息的方差进行度量,此操作可以反映特征图的语义信息分布密集程度. 在信息更为集中的区域,参数 γ 的值会相应增大,而对应的权重 w 也会与 γ 呈正比关系. 因此,信息含量较高的区域会被分配更大的 W_γ 权重,这样可以更精确地掌握空间像素的变化情况. 权重的归一化相关计算如式(1)所示,其中的 i,j 分别指代批次中的通道索引.

$$W_\gamma = \{w_i\} = \frac{\gamma_i}{\sum_{j=1}^{C}\gamma_j}, i,j = 1,2,\cdots,C \tag{1}$$

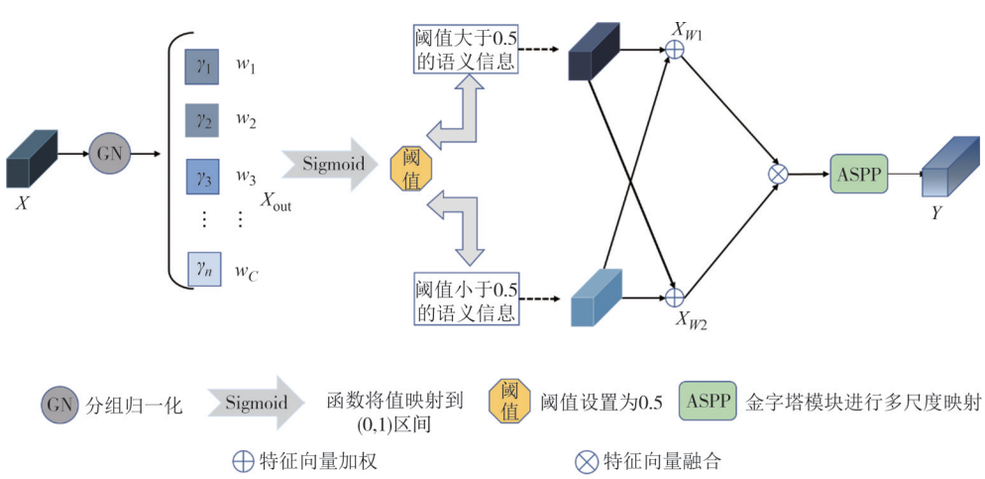


图 2　IGRM 模块结构图

Fig.2　IGRM module diagram

在实施分组归一化机制时,依据通道数量将输入特征 X 拆分成多个特征小组,经过减去每组的均值 μ 和除以每组的方差 σ 实现特征 X_{out} 的规范化. 这里,μ 和 σ 表示输入特征 X 的均值和方差,ε 是一个用于维持除法运算稳定的极小正数,γ 作为可训练的缩放因子,它们与 W_γ 权重共同参与加权. 特征 X_{out} 的计算如式(2)所示.

$$X_{\text{out}} = \text{GN}(X) = \gamma \frac{X - \mu}{\sqrt{\sigma^2 + \varepsilon}} + W_\gamma \tag{2}$$

经过归一化处理的X_{out}值将被约束在(0,1)内,并且依据预定阈值T(通常设定为0.5)来实施筛选机制.那些高于阈值T的特征值被认定具有较高信息含量,反之,低于阈值T的则被划为信息贫乏的特征.接下来,利用交叉重组算法,把经过权重调整的两组特征彻底结合,以此提升特征间的信息交互效率.将经过交叉重组处理的特征X_{W1}与X_{W2}相拼接,进而形成富含语义信息的特征集.在此环节,采用空间金字塔池化调整并聚合多尺度的特征信息,从而达到多尺度的特征表达效果.在IGRM模块中,输出特征图记为O,其中i表示不同扩张率下的卷积核尺寸,s为卷积操作的步长,d为扩张率,而$(d-1)$代表卷积操作所需的填充数量,k表示标准卷积的核大小,P代表经过交叉重组处理后的两组特征集合(集合1和集合2).经过ASPP模块的综合处理,最终得到的特征图O的计算方式如式(3)所示.IGRM模块的输出计算则遵循式(4),其中Y代表模块最终输出的特征.

$$O(X_{WP}) = \left\lfloor \frac{i + 2d - k - (k-1)(d-1)}{s} \right\rfloor + X_{WP}, \quad (P = 1, 2) \tag{3}$$

$$\text{IGRM}(\text{out}) = Y = O(X_{W1}) + O(X_{W2}) \tag{4}$$

1.2 通道区分重构模块(CRRM)

在深度网络处理特征提取过程中,随着网络层数的增加及外部干扰的介入,往往会出现语义信息的部分遗失.针对这个问题,本文提出了一种按通道进行分类与重构的模块,即通道区分重构模块(CRRM).该模块的核心策略在于通过通道分类和重构技术,提升特征表示的丰富性和准确性.

本模块采纳全局平均池化配合一维卷积的方式,对特征图的通道关联性进行模型构建,并通过残差将其融合到原始输入特征权重中,从而实现对特征更精细和丰富的表示.设计超参数a及全局平均池化操作被用于分类特征,而一维卷积操作则将特征映射到原始特征图的通道维度上,以实现通道的重建,从而达到去除冗余特征的目的,并同时精细化原始特征.这一结构有效补偿了因网络深度增加和外部因素导致的语义损失,从而提升了特征的表达精度与丰富度,使得CRRM模块能够起到补偿网络深度增加导致语义失真缺陷的作用,进而提高特征精细度的表达和丰富度.

如图3所示,CRRM模块的具体架构展现如下:起初,将核心网络初期及中期产生的多尺度特征集X_W,基于其通道数C进行分岔,形成两条并行路径,其中一条路径包含aC个通道,另一路则有$(1-a)C$个通道,这里的a作为超参数,其值介于0与1之间,在实验操作中赋值为0.5,以实现对X_W特征的有效分割.继而,借助1×1的卷积技巧减轻计算压力,并调整特征图的通道数,从而派生出X_{up}和X_{low}两种不同类型的特征.最终,借助残差连接技术,将原始特征与重构后的特征进行有效整合,从而生成新的特征集合.

在此环节中,X_{up}数据被引入"多元化特征通道".该通道运用高效能的卷积技术(例如GWC[16]和PWC[17])替代传统的卷积方式,旨在获得更为丰富的特征展现并减少计算负担.通过分组卷积(GWC)技术,降低了计算复杂度.但分组卷积作为简单的一维

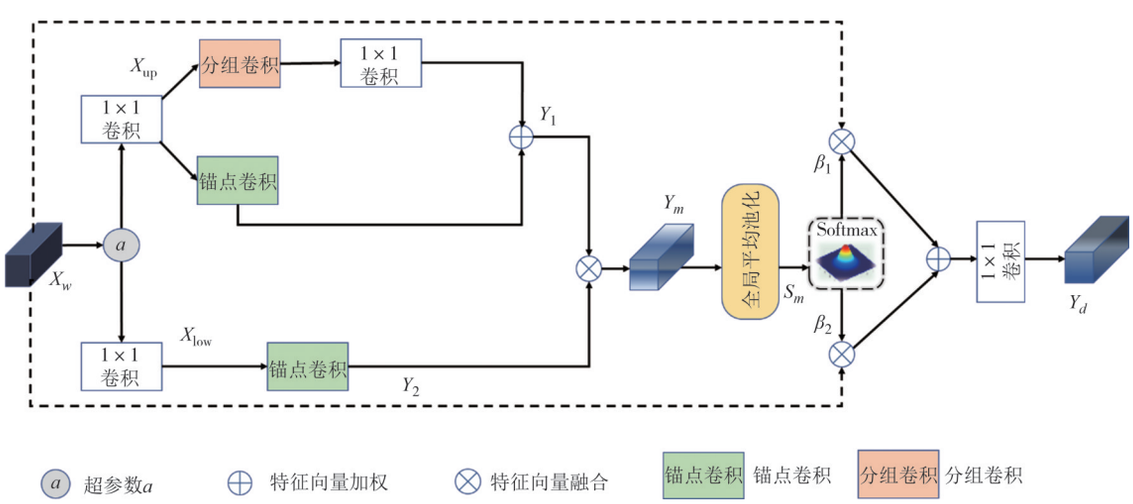


图3 CRRM模块结构图

Fig.3 CRRM module structure diagram

卷积,存在限制不同特征信息交流的问题.为了克服这个限制,为分组卷积搭配了锚点卷积(PWC)来增强通道间的信息交换.具体来说,X_{up}输入后首先通过GWC(在实验设定中分为两组进行),接着应用PWC进行处理,最后将这两个步骤产生的特征合并,生成特征映射Y_1.

在另一条轨迹上,X_{low}传递至次级转换通道中,借助1×1的卷积以及PWC卷积技术,形成了含有表层细节信息的特征映射,这作为增强特征提取路径的辅助环节.此举旨在制造出含有更丰富细节信息的特征映射,并将其映射为Y_2.

当两个分支输出的特征Y_1和Y_2准备就绪后,将它们合并,进而采用全局均值汇聚技术提取融合特征Y_m的通道域空间属性S_m.假设输入特征图的尺寸标识为$H\times W\times C$,其中H指代图像的高度,W指代图像的宽度,C指代图像的通道维度,对每一个独立通道而言,对其包含的全部像素特征值执行平均运算.空间属性S_m的确定依据式(5)进行计算.

$$S_m=\text{pooling}(Y_m)=\frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W}Y_m(i,j),m=1,2 \tag{5}$$

全局平均池化计算出Y_m的通道空间信息S_1和S_2后,将这两个信息上下堆叠,运用Softmax激活机制计算得出特征关键系数β_1与β_2.随后,实施残差拼接策略,将基础特征与加权组合X_W的β_1、β_2系数相融合,以此提升特征信息的表达能力.随后,通过一个卷积层对通道数量进行优化调整,生成作为CRRM模块输出Y_d的优化特征.CRRM模块的输出计算过程如式(6)所示.在此公式中,原始特征权重的残差连接得到了体现.

$$\text{CRRM(out)}=Y_d=\beta_1\text{Res}(X_W)+\beta_2\text{Res}(X_W) \tag{6}$$

1.3　多尺度映射及模型整体的输出关系

多尺度连接在特征提取网络中起到了重要的作用,其主要由下采样连接和整合上采样连接两部分组成.为实现不同模块间的信息交互与融合,从而生成丰富的特征输出,设计了多尺度连接模块.其具体结构如图4所示.通过这一模块,网络能够有效整合不同尺度的特征信息,提升模型的表达能力和性能.

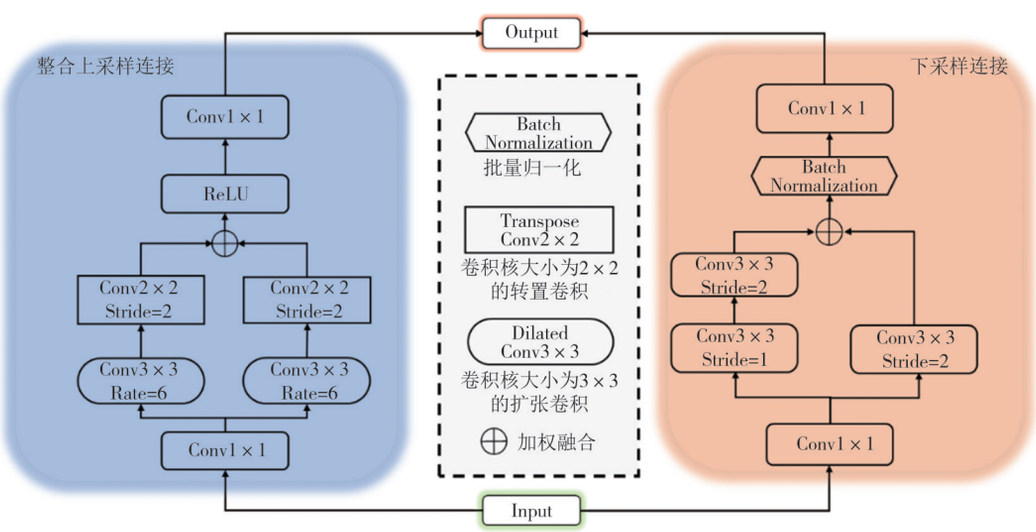


图4　多尺度映射结构图

Fig.4　Cross-scale structure diagram

利用混合的下采样与上采样连接,该网络能在不同级别上捕捉到丰饶的语义内容,并对这些数据进行高效的综合,从而增强模型的效能与抗干扰能力.

下采样连接(down-sample connection):起初,通过1×1卷积对输入特征进行通道调整,随后采用2倍步长的3×3卷积配合步长为1的3×3卷积依次进行,接着进行批量规范化处理,并利用1×1卷积实现通道尺寸的再次调整,最终生成下采样连接的输出$D(x)$.该下采样连接旨在结合低级与高级特征模块的优势,$D(x)$的相关表达式可参照公式(7).

$$D_j^i(x)=f_\sigma\left(f_\delta(x)+f_\mu\left(f_\varepsilon(x)\right)\right) \tag{7}$$

在构建下采样连接表达式时,输入变量x代表输入的向量.此处采用ReLU函数作为激活手段,并配置两个不同采样步长的3×3卷积层,其中前者的步长设定为2,后者则采用步长为1.每一组卷积层都融

入了批量标准化流程.而其中的i与j变量分别代表输入与输出的通道数目,这一设置与Transformer模块在确定输出特征通道数量时的机制相吻合.

整合上采样连接(integration up-sample connection):该过程启动于特征通过1×1的卷积操作来实现通道调整.接着,通过应用步长为2、内核尺寸为2×2的反卷积操作以及扩张率设定为6、内核尺寸为3×3的扩张卷积运算,并分别实施上采样操作完成特征整合.随后,借助ReLU激活函数进行处理,完成非线性变换,进而生成整合上采样后的输出向量LU(x).为了高效地获取多层次的空间上下文信息,将扩张卷积与步长为2的反卷积操作结合嵌入局部上采样连接中,整合上采样的输出向量LU(x)的计算方式如式(8)所示.

$$\mathrm{LU}_m^n(x)=f_\phi^{12}\left(f_\sigma\left(f_\phi^6(x)\right)\right) \tag{8}$$

其中,x代表输入的向量,通过复合函数将扩张率设置为12的扩张卷积层和一个转置率为6的转置卷积层进行嵌套;m和n分别表示输入和输出的通道数.

本文使用Swin-Tiny作为核心网络架构,其构成的多个Transformer模块依次设定的输出通道大小分别为96(首层)、192(次层)、384(第三层)和768(末层).在第1.1及1.2小节中,已经详细阐述了IGRM和CRRM两个功能单元的计算原理及其输出公式.而在1.3小节中,对多尺度特征融合的输出准则进行了阐述.整个模型从输入特征图像到最终输出的转换,是通过主干网络、IGRM模块、CRRM模块以及多尺度特征融合的连续处理实现的,具体的输出表达式可参照式(9)至式(12),其中各个符号代表了不同网络阶段的输出产物.

$$F_4=\mathrm{ST}_4+D_{384}^{768}\left(\mathrm{IGRM}\left(D_{192}^{384}\left(\mathrm{ST}_2\right)\right)\right) \tag{9}$$

$$F_3=\mathrm{IGRM}\left(\mathrm{ST}_3\right)+F_4+D_{192}^{384}\left(\mathrm{CRRM}\left(D_{96}^{192}\left(\mathrm{ST}_1\right)\right)\right) \tag{10}$$

$$F_2=\mathrm{CRRM}\left(\mathrm{ST}_2\right)+F_3+\mathrm{LU}_{768}^{192}\left(F_4\right) \tag{11}$$

$$F_1=\mathrm{ST}_1+F_2+\mathrm{LU}_{384}^{96}\left(F_3\right) \tag{12}$$

1.4 损失函数

切块损失函数能显著地标注并强化目标区域,显示出卓越的效能.但在进行模型训练时,其稳定性不足的问题变得尤为明显,尤其是在处理微小目标检测任务时,这一问题可能会引发梯度溢出,进而干扰模型的训练效果.为了解决这个问题,可以结合其他损失函数,如二元交叉熵损失,来增强模型的鲁棒性.因此,本文提出了一种复合损失函数,将切块损失与二元交叉熵损失进行融合,具体形式见式(13)至式(15).该组合旨在平衡检测精度与训练稳定性,特别是在小目标检测中,能够显著提升模型的整体性能.

$$l_{\mathrm{ce}}=-\frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K}y_k^{(n)}\log\hat{y}_k^{(n)} \tag{13}$$

$$l_{\mathrm{dice}}=1-\frac{2}{N}\sum_{n=1}^{N}\sum_{k=1}^{K}\frac{\hat{y}_k^{(n)}y_k^{(n)}}{\hat{y}_k^{(n)}+y_k^{(n)}} \tag{14}$$

$$l_{\mathrm{main}}=l_{\mathrm{ce}}+l_{\mathrm{dice}} \tag{15}$$

式中:N表示样本总数;K表示类别总数;$y^{(n)}$表示正确标签,经过Softmax函数处理后的预测图为$\hat{y}^{(n)}$.$\hat{y}_k^{(n)}$是n个样本的网络输出预测图,每个图有k个类别;$y_k^{(n)}$表示n个正确的语义分割标签,每个标签也有k个类别.

2 实验设计与结果分析

2.1 数据集

ISPRS Potsdam:标志性航空遥感数据集,以其丰富的城市景观著称.数据集中展示了丰富的地物类别,总共含有6个类别,涵盖不渗水表面、各类建筑物、低矮植物、林木、机动车以及其他琐碎元素.Potsdam数据集由38张尺寸为6 000×6 000像素的图片组成,每个像素代表实际距离的5 cm.此数据集中,有15张图片作为测试集使用,余下的23张图片则用作训练集.

ISPRS Vaihingen:数据集涵盖了33幅各式各样的航空遥感影像,其分辨率分布在2 000×4 000像素之间,对应的像素间距是9 cm.在这些图像上,可以观察到众多分散的微型建筑,具有丰富的地表特征和复杂的城市结构.与其他遥感数据集相比,Vaihingen在建筑物的多样性、道路网络的复杂性以及绿地和水体的分布方面展现了显著的特征.总共涵盖了6个图像类别.所有图像都被裁剪为1 024×1 024大小.

UAVid数据集:UAVid数据集是一个专注于城市场景理解的航空遥感语义分割数据集,由武汉大学、俄亥俄州立大学和特温特大学于2018年联合发布.该数据集中标注了8个类别,并选取了300张图像进行密集标记,为复杂背景下的语义分割任务提供了丰富的支持.该数据集能够评估算法在复杂背景下的性能.涵盖了街景中不同大小尺度的多种物体.

2.2 实验设备及评估指标

本实验基于Ubuntu 9.04系统和Pytorch框架,使

用 Tesla P40 显卡、32 GB 内存、512 GB 固态硬盘及 256 GB 虚拟内存进行模型训练.训练时采用 AdamW 优化器,学习率分别为 6×10^{-4} 和 6×10^{-5},权重衰减系数为 0.01,并结合预热和余弦退火策略调整学习率以优化模型性能.在测试阶段,采用综合准确率(OA)、平均 F_1 指标(MF_1)、平均交并比(mIoU)、精确度 P(precision)和召回率 R (recall)等指标全面评估模型性能,见式(16)~式(20).

$$OA = \frac{\sum_{k=1}^{K} TP_k}{\sum_{k=1}^{K} TP_k + FP_k + TN_k + FN_k} \tag{16}$$

$$mIoU = \frac{1}{K}\sum_{k=1}^{K}\frac{TP_k}{TP_k + FP_k + FN_k} \tag{17}$$

$$P_k = \frac{TP_k}{TP_k + FP_k} \tag{18}$$

$$R_k = \frac{TP_k}{TP_k + FN_k} \tag{19}$$

$$F_{1k} = 2 \times \frac{P_k \times R_k}{P_k + R_k} \tag{20}$$

式中:k 表示样本类别数;TP 为模型预测为正类的真实正样本;TN 为模型预测为负类的真实负样本;FP 为模型预测为正类的错误负样本;FN 为模型预测为负类的错误正样本.精确度高表明预测错误较少,召回率高表明检测遗漏较少.F_1 分数和 mIoU 指标共同反映了模型的综合性能,数值越高,模型表现越优.

2.3　消融实验

利用 ISPRS Potsdam 数据集进行独立模块效能评估.表 1 列举了各模块的名称及其缩写,在 ISPRS Potsdam 数据集上的消融实验的具体结果见表 2.

表 1　消融实验各部分名称及缩写

Tab.1　Abbreviations of relevant network names in ablation experiments

名称缩写	描述
Baseline	骨干基准
Baseline+IGRM	基准+信息分组重构卷积
Baseline+CRRM	基准+通道分类重构卷积
Baseline +IGRM+CRRM	基准+信息分组重构卷积+通道分类重构卷积

Baseline:在实验基础设置上,使用 Swin-Tiny 作为骨干网络,并通过四个核心 Transformer Block 单元执行特征的直接转换,从而构建出四个专门的特征融合模块,这些模块依次相互作用,最终汇聚成我们的基础模型框架.

表 2　ISPRS Potsdam 数据集上的消融实验结果

Tab.2　Ablation experiment results on ISPRS Potsdam dataset　　%

方法策略	OA	MF_1	mIoU
Baseline	87.90	89.51	81.50
Baseline+IGRM	88.71	90.50	83.19
Baseline+CRRM	90.33	91.25	85.16
Baseline +IGRM+CRRM	90.58	92.32	85.98

在原有基准模型之上,加入了信息重组功能模块(IGRM).根据在 ISPRS Potsdam 数据集上进行的训练成果分析,观察到全局精度(OA)提升了 0.81 个百分点,F_1 指标(MF_1)也有所提高,提升了 0.99 个百分点,同时 mIoU 指标显著上升了 1.69 个百分点.这一提升证明了 IGRM 模块在整合深层与浅层语义信息方面的有效性,进一步提升了多目标特征的语义刻画精度.正如图 5 第三行所展示的,相较于基准模型,IGRM 在处理诸如远离观测点和低分辨率下的语义信息(例如在红色背景和绿色植被中那些小型目标)时,表现出了更高的辨识精度,但仍需进一步优化对某些语义的识别.因此,IGRM 模块有效提升了高分辨率小目标的语义分割精度.

Baseline+CRRM:在原有基准模型之上融合了信道判别组件(CRRM),通过对 ISPRS Potsdam 数据集的深度训练,观察到全局精度(OA)实现了 2.43 个百分点的提升,F_1 指标(MF_1)增加了 1.74 个百分点,同时 mIoU 也得到了 3.66 个百分点的增强.该 CRRM 组件的设计宗旨是降低信道内信息的多余度,并通过调整参数来强化和优化多目标场景中细微目标的语义特性.正如图 5 中第五列展示的那样,CRRM 组件显著强化了低分辨率区域中细微目标(例如车辆、矮小植物及杂草)的语义表达能力,有效提高了特征表达的清晰度.

同时引入 CRRM 与 IGRM 后,基于基准模型的整体性能得到了显著提升.表 2 显示,模型的 OA 提升了 2.68 个百分点,MF_1 提升了 2.81 个百分点,而在 ISPRS Potsdam 数据集上的 mIoU 则提升了 4.48 个百分点.本研究证实了 IGRM 与 CRRM 组件在增强分割准确性上的显著作用,且无需引入跨尺度连接.如图 5 所示,通过引入这两个模块,在广袤的空中视角下,对繁杂场景中各式目标的分割处理得以达到更为细腻的切割效果,同时在细节呈现上更显丰富性与准确性.

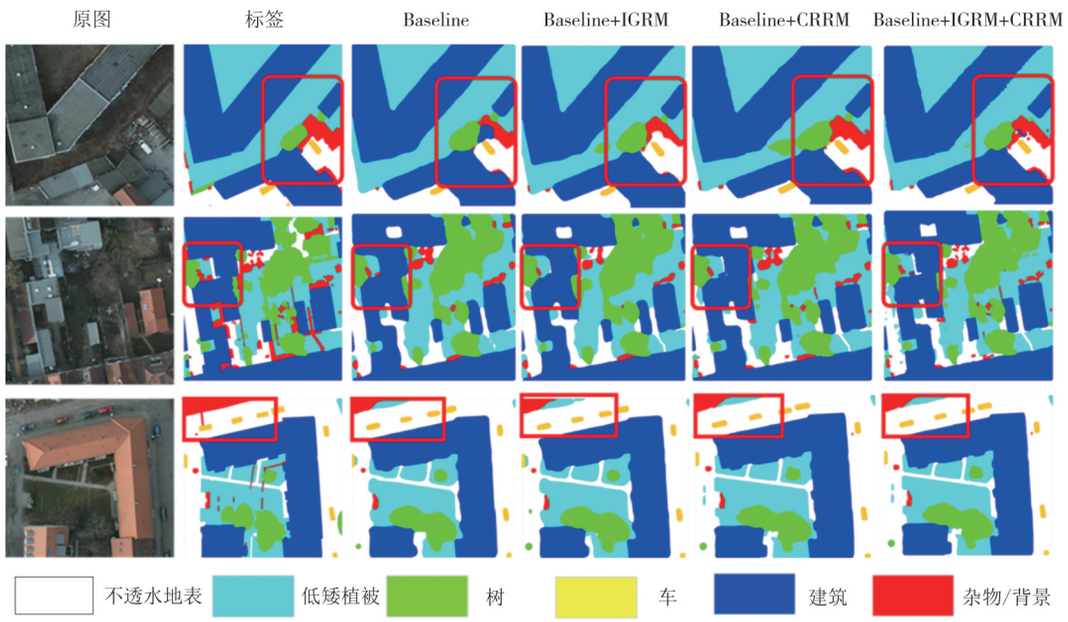


图 5 模型消融实验可视化结果对比图

Fig.5 Visualization comparison of model ablation experiments

2.4 对比实验

对比实验挑选了在 paperwithcode 平台上分割领域内较为先进的开源模型,例如,集成了位置和通道注意力的动态注意力网络 DANet(发布于 2019 年)、基于空间语义路径构建的 ABCNet[18](发布于 2021 年)、装备了复杂注意力机制的多元注意力网络 MANet[19](发布于 2022 年)、参考了 Unet 架构并结合 Transformer 的 Unetformer 网络[20](发布于 2022 年)、融合了多层级注意力与残差连接的 MAResUNet[21](发布于 2021 年)、通过 Transformer 构建全局信息的 Segmenter[22](发布于 2021 年),以及拥有双重编码器和单一解码器架构的 SSG2[23](发布于 2023 年). 表 3 对 Urswin 模型在 Potsdam 数据集中的内存使用、参数量级及运算复杂度进行了评估,其中,模型的计算量通过 10 亿次浮点运算(GFLOPs)来评估,单位 G 表示 10 亿次浮点运算;内存占用以兆字节(MB)为计量单位,用于衡量显存大小;模型参数以百万(M)为计量单位,用于评估网络的总参数数量. 并展示了其在 mF_1、OA、mIoU 等关键指标上的显著进步. 相较于 SSG2,Swin-ERRM 模型在降低参数数量和计算负担的同时,在 ISPRS Potsdam 数据集上依然保持了卓越的 mF_1、OA 及 mIoU 性能指标.

表 3 与其他语义分割模型结果进行参数比较

Tab.3 Parameter comparison with other semantic segmentation network results

网络	骨干	内存占用/MB	模型参数/M	计算量/G	mF_1/%	OA/%	mIoU/%
DANet	ResNet18	2 024.9	12.6	120.24	89.7	90.0	83.13
Segmenter	ViT-Tiny	2 665.9	23.6	67.30	90.4	90.1	83.8
MAResUNet	ResNet18	**658.51**	15.2	**41.29**	91.3	90.2	85.4
ABCNet	ResNet18	1 873.2	14.0	62.16	92.0	90.4	85.9
MANet	ResNet18	2 131.6	12.0	87.25	92.5	90.6	86.0
Unetformer	ResNet18	1 591.7	**11.9**	51.67	92.8	91.0	86.7
SSG2	ViT-Tiny	5 538.0	21.6	60.38	92.5	90.4	86.83
Swin-ERRM	Swin-Tiny	5 315.3	12.9	52.70	**93.2**	**91.53**	**87.15**

2.4.1 ISPRS Potsdam 数据集上的对比实验

对比分割网络采用了国际领先的算法设计. 具体参数对比情况见表 4,而图 6 则直观展示了相关可视化效果. 在 ISPRS Potsdam 数据集的测试中,相较于其他网络,本网络在各子项的 F_1 和 mIoU 指标上均实现了不同程度的提高. 在针对建筑、低矮植被及机动车辆的三项目标识别任务中,Swin-ERRM 模型分别实现了 95.8%、88.3% 及 96.3% 的 F_1 值峰值. 在分割图像中,语义数据被细致划分为六个不同类别,其中,建筑与不透水地表的语义信息较为显著,而其余

类别目标所携带的语义信息相对较少,共同构筑了这一复合多目标分割图像.与 paperwithcode 平台上的算法相比,本模型通过引入全局上下文信息的建模与重构模块,显著提高了识别精度,超越了大部分以 ResNet 作为主干网络的算法.最终,衡量指标 mF_1、OA 以及 mIoU 的分数依次攀升至 93.2%、91.53% 及 87.15%,凸显出卓越的表现,尤其在 mF_1 这一关键指标上显著优于其他对比网络成果.

表 4　与其他语义分割网络在 ISPRS Potsdam 数据集上进行的对比实验

Tab.4　Comparative experiments with other semantic segmentation networks on ISPRS Potsdam dataset　%

网络	骨干	F_1					mF_1	OA	mIoU
		不透水地表	建筑	低矮植被	树	车			
DANet	ResNet18	91.8	94.7	85.6	88.0	90.1	89.7	90.0	83.13
Segmenter	ViT-Tiny	92.1	95.3	86.3	88.2	95.0	90.4	90.1	83.8
MAResUNet	ResNet18	93.1	95.8	86.9	88.0	95.1	91.3	90.2	85.4
ABCNet	ResNet18	93.0	95.5	87.2	88.5	95.0	92.0	90.4	85.9
MANet	ResNet18	92.3	95.1	86.5	88.3	95.9	92.5	90.6	86.0
Unetformer	ResNet18	93.3	95.5	87.7	89.0	94.2	92.8	91.0	86.7
SSG2	ViT-Tiny	92.7	95.1	87.9	88.9	96.0	92.5	90.4	86.83
Swin-ERRM	Swin-Tiny	93.6	95.8	88.3	89.7	96.3	93.2	91.53	87.15

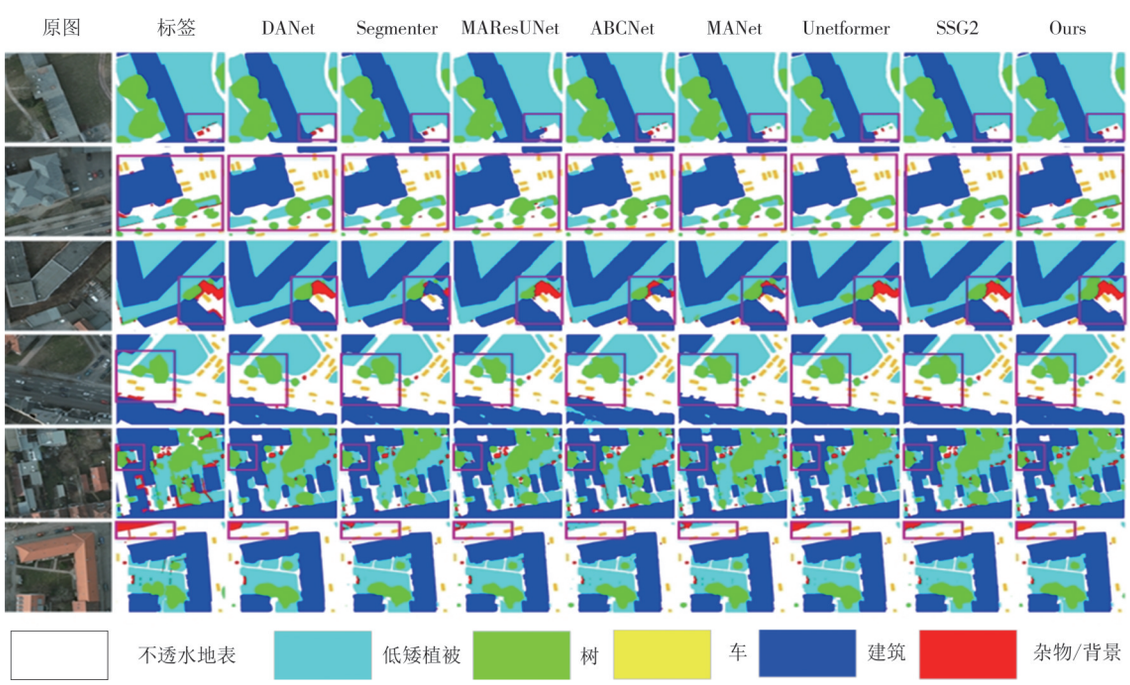


图 6　在 ISPRS Potsdam 数据集上实验结果可视化对比图

Fig.6　Visualization comparison of experimental results on ISPRS Potsdam dataset

2.4.2　ISPRS Vaihingen 数据集上类别对比实验

表 5 列出了采用 ISPRS Vaihingen 数据集的实验分割结果对比,而图 7 则对这些结果进行了可视化展示.特别地,在图 7 的每一行中,我们可以观察到六个不同类别的语义划分图像.模型预测[24]的边界与标签非常接近,特别是在处理不规则建筑群落、汽车和杂物时,表现出良好的分割能力,表明 Swin-ERRM 在多目标特征分割和细小目标特征的精细化处理上都很有效.在不透水表面的分类精度方面,Swin-ERRM 网络略低于 Unetformer 网络,差距仅为 0.2 个百分点.然而,在 F_1 分数和 mIoU 指标的多个子项上,Swin-ERRM 网络均优于其他网络.通过对 paperwithcode 平台上近年来全球发布的分割算法进行比较,Swin-ERRM 网络在 Vaihingen 数据集上的表现较为优异.

2.4.3　复杂背景下较小目标的对比实验分析

为了更有效地展示本文提出的 Swin-ERRM 模型在复杂背景下对小尺度目标特征的分割能力[25],设计了针对复杂背景下小车类别的对比实验.选用 UAVid 数据集,以充分体现本文模型在多目标类别

表5 与其他语义分割网络在ISPRS Vaihingen数据集上进行的对比实验

Tab.5 Comparative experiments with other semantic segmentation networks on ISPRS Vaihingen dataset %

方法	骨干	F_1					mF_1	OA	mIoU
		不透水地表	建筑	低矮植被	树	车			
DANet	ResNet18	90.7	93.6	81.5	88.3	80.8	85.2	87.8	77.2
Segmenter	ViT-Tiny	91.0	93.8	82.5	89.3	83.9	86.5	88.5	78.0
MAResUNet	ResNet18	90.3	93.2	81.7	89.6	85.6	87.2	88.9	78.3
ABCNet	ResNet18	92.1	94.4	83.2	89.4	88.3	88.5	89.7	80.9
MANet	ResNet18	92.5	95.0	84.0	90.0	88.8	89.7	89.6	81.3
Unetformer	Resnet18	**92.9**	95.3	84.3	90.1	89.9	89.5	90.3	81.7
SSG2	ViT-Tiny	92.1	95.0	83.8	89.8	90.7	90.3	90.8	82.2
Swin-ERRM	Swin-Tiny	92.7	**95.3**	**84.5**	**90.3**	**96.3**	**91.7**	**91.4**	**82.93**

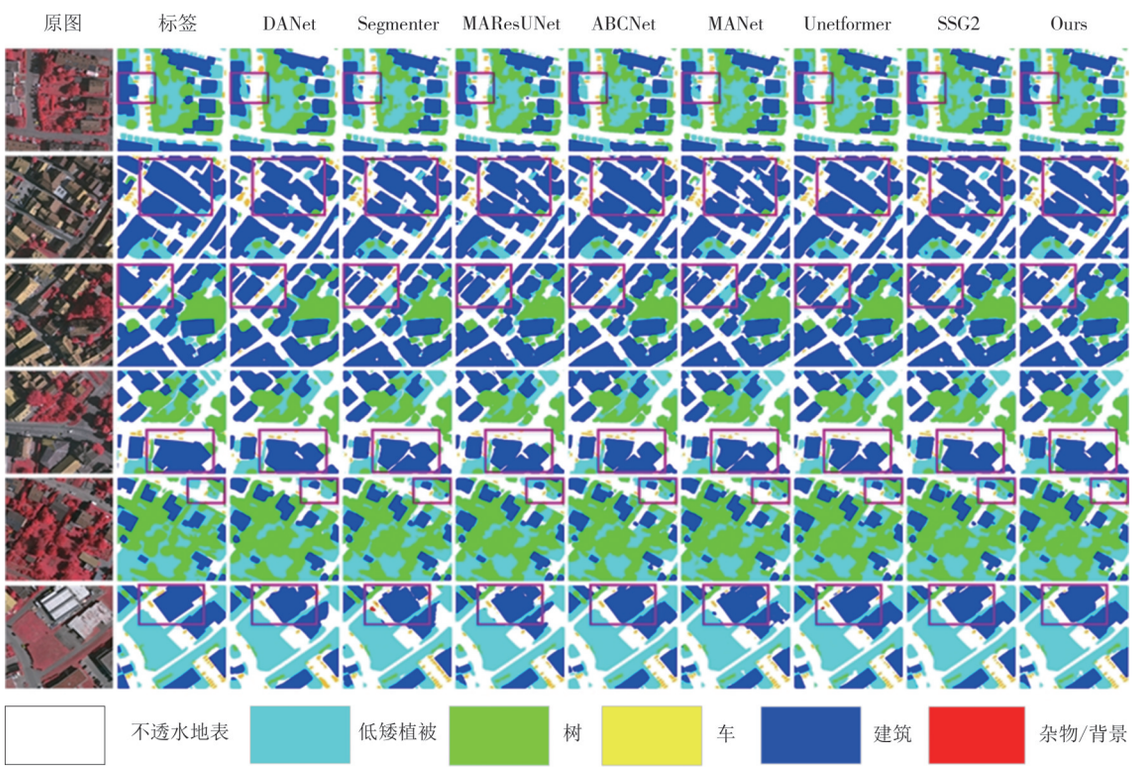


图7 在ISPRS Vaihingen数据集上实验结果可视化对比图

Fig.7 Visualization comparison of experimental results on ISPRS Vaihingen dataset

中特征提取的优势.表6展示了模型在复杂航空背景中针对小车这一小尺度目标类别[26]的分割结果对比,而图8则对这些结果进行了可视化展示.

由表6可知:在UAVid数据集中针对复杂背景下小车这一小尺度目标类别的分割精度($P_{车}$)进行比较,Swin-ERRM模型相比于其他两个分割模型分别提高了13.3个百分点和4.9个百分点;同时,mIoU和OA也分别提高了6.46个百分点和1.96个百分点,以及2.6个百分点和1.6个百分点.

在可视化结果图8中,比较标签图可发现,在复杂背景下进行多目标分割,DANet模型的分割结果存在邻近像素间的像素沉没和部分语义失真的问题.虽然Swin Transformer能够在一定程度上完成小车目标的分割,但其对小尺度目标边界的清晰度仍存在不足.在复杂场景中,邻近像素的模糊边界可能导致模型在区分目标与背景时遇到困难,特别是当

表6 与其他语义分割网络在UAVid数据集上的对比实验

Tab.6 Comparative experiments with other semantic segmentation networks on UAVid dataset %

模型	骨干	$P_{车}$	mIoU	OA
DANet	ResNet18	47.2	61.4	83.8
Swin Transformer	Transformer	55.6	65.9	84.8
Swin-ERRM	Swin-Tiny	**60.5**	**67.86**	**86.4**

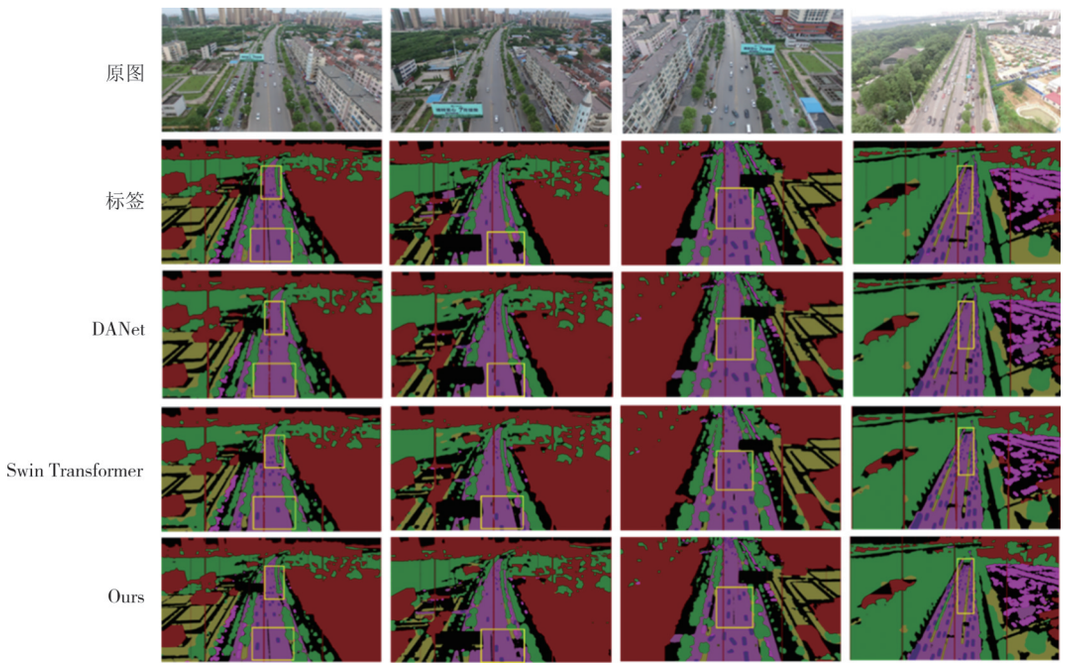


图 8　在 UAVid 数据集上实验结果可视化对比图

Fig.8　Visualization comparison of experimental results on UAVid dataset

邻近像素在颜色、纹理等特征上高度相似时,这会导致小目标被错误分类或与背景混淆.

相比之下,Swin-ERRM 模型在小尺度目标类别及其边界的分割效果上表现优于 DANet 和 Swin Transformer. 这表明 Swin-ERRM 模型在处理小尺度目标方面具有更强的能力,能够更准确地提取和区分小尺度目标.

3　结 论

基于 Swin Transformer 网络,本文构建了一种新型航空遥感图像分割模型,命名为 Swin-ERRM. 该架构的设计宗旨是在含有众多目标的复杂场景中,识别并提取出细微尺度目标,同时针对 Swin Transformer 在特征提取阶段可能遗漏的语义细节进行补偿,有效应对航空遥感图像处理中关于小尺度目标识别精度不足的挑战. 该架构融合了信息分组重构模块(IGRM)、通道分类重构模块(CRRM)以及复合上采样结构. IGRM 对信息熵进行分级处理,从而达成对不同复杂度特征的融合与再塑,同时借助多层次的金字塔架构完成信息的跨级别传递. CRRM 单元通过融合参数自适应调整以及锚点与分组形式的卷积操作,对特征维度进行了细致的归类与筛选,并利用全局均值池化与基础特征实现了残差整合,打造出成熟特征表示. Swin-ERRM 架构大幅增强了在航空遥感图像中多目标分割的精确性,显示出其在处理复杂背景下小尺度目标特征方面的强大能力.

参考文献

[1] 吕杰,沈琦,吕敏,等. 基于深度学习的遥感影像语义分割研究进展[J]. 生态与进化前沿, 2023, 11: 1201125.
LÜ J, SHEN Q, LÜ M, et al. Research progress on semantic segmentation of remote sensing images based on deep learning[J]. Frontiers in Ecology and Evolution, 2023, 11: 1201125. (in Chinese)

[2] 刘光宇, 曹禹, 曾志勇, 等. 结合多特征赋权的谱聚类水下多目标分割技术[J]. 湖南大学学报(自然科学版), 2022, 49(10): 51-60.
LIU G Y, CAO Y, ZENG Z Y, et al. Underwater multi-object segmentation technology based on spectral clustering with multi-feature weighting [J]. Journal of Hunan University (Natural Sciences), 2022, 49(10): 51-60. (in Chinese)

[3] KUMAR D, KUMAR D. Hyperspectral image classification using deep learning models: a review [J]. Journal of Physics: Conference Series, 2021, 1950(1): 012087.

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

[5] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. IEEE Transactions on

Pattern Analysis and Machine Intelligence, 2018, 40(4): 834-848.

[6] CHEN L C, ZHU Y K, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[M]//Computer Vision-ECCV 2018. Cham: Springer International Publishing, 2018: 833-851.

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

[8] YU C Q, WANG J B, PENG C, et al. BiSeNet: bilateral segmentation network for real-time semantic segmentation [M]// Computer Vision-ECCV 2018. Cham: Springer International Publishing, 2018: 334-349.

[9] LIU Z, LIN Y T, CAO Y, et al. Swin transformer: hierarchical vision transformer using shifted windows [C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal, QC, Canada. IEEE, 2021: 9992-10002.

[10] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module [M]//Computer Vision-ECCV 2018. Cham: Springer International Publishing, 2018: 3-19.

[11] HUANG Z L, WANG X G, HUANG L C, et al. CCNet: criss-cross attention for semantic segmentation [C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, Korea (South). IEEE, 2019: 603-612.

[12] CHEN Y P, FAN H Q, XU B, et al. Drop an octave: reducing spatial redundancy in convolutional neural networks with octave convolution [C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, Korea (South). IEEE, 2019: 3435-3444.

[13] CHOLLET F. Xception: deep learning with depthwise separable convolutions[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA. IEEE, 2017: 1800-1807.

[14] CHEN L C, PAPANDREOU G, SCHROFF F, et al. Rethinking atrous convolution for semantic image segmentation [EB/OL]. [2024-04-30]. https://arxiv.org/abs/1706.05587v3.

[15] WU Y X, HE K M. Group normalization[C]// Computer Vision-ECCV 2018.Cham: Springer International Publishing, 2018.

[16] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [J]. Communications of the ACM, 2017, 60(6): 84-90.

[17] HUA B S, TRAN M K, YEUNG S K. Pointwise convolutional neural networks [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA. IEEE, 2018: 984-993.

[18] LI R, ZHENG S Y, ZHANG C, et al. ABCNet: attentive bilateral contextual network for efficient semantic segmentation of fine-resolution remotely sensed imagery [J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2021, 181: 84-98.

[19] JIANG K X, LIU J, ZHANG W H, et al. MANet: an efficient multidimensional attention-aggregated network for remote sensing image change detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 3328334.

[20] WANG L B, LI R, ZHANG C, et al. UNetFormer: a UNet-like transformer for efficient semantic segmentation of remote sensing urban scene imagery [J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2022, 190: 196-214.

[21] LI R, ZHENG S Y, DUAN C X, et al. Multistage attention ResU-net for semantic segmentation of fine-resolution remote sensing images[J]. IEEE Geoscience and Remote Sensing Letters, 2021, 19: 8009205.

[22] STRUDEL R, GARCIA R, LAPTEV I, et al. Segmenter: transformer for semantic segmentation [C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal, QC, Canada. IEEE, 2021: 7242-7252.

[23] DIAKOGIANNIS F I, FURBY S, CACCETTA P, et al. SSG2: a new modelling paradigm for semantic segmentation [EB/OL]. 2023: 2310.08671. https://arxiv.org/abs/2310.08671v1.

[24] ELHAJ K, ALSHAMSI D, ALDAHAN A. GeoZ: a region-based visualization of clustering algorithms [J]. Journal of Geovisualization and Spatial Analysis, 2023, 7(1): 15.

[25] HONG X, ROOSEVELT C H. Orthorectification of large datasets of multi-scale archival aerial imagery: a case study from türkiye [J]. Journal of Geovisualization and Spatial Analysis, 2023, 7(2): 23.

[26] 吴嘉昕, 王小鹏, 刘扬洋. 子空间与KL信息结合的FCM多光谱遥感图像分割[J]. 湖南大学学报(自然科学版), 2024, 51(8): 23-33.
WU J X, WANG X P, LIU Y Y. Fuzzy C-mean multi-spectral remote sensing image segmentation with combined subspace and KL information [J]. Journal of Hunan University (Natural Sciences), 2024, 51(8): 23-33. (in Chinese)