doi:10.12068/j.issn.1005-3026.2026.20240174

多尺度分层融合Swin Transformer模块辽河封育区土地覆盖分类

王 植[1]，王梦晴[1]，陈利娟[2]
（1.东北大学 资源与土木工程学院，辽宁 沈阳 110819；
2.辽宁省水利水电科学研究院有限责任公司沈阳分公司，辽宁 沈阳 110003）

摘 要：辽河封育区内地物边界缺乏规则、辨别难度大.为监测封育区内耕地和林地的侵占行为，提出一种多尺度分层融合Swin Transformer模块的土地覆盖分类方法，建立由Swin Transformer模块堆叠的逐级特征提取模块，提取全局空间信息和局部上下文信息；设计多尺度特征解码模块，能高效融合高低层级特征；增加额外的辅助解码头以达到更精准分类，并充分融合两个解码分支提取的特征.利用GF-2号卫星影像制作的封育区土地覆盖分类数据集和公开数据集GID-5进行验证.结果表明，主要指标平均交并比分别为82.83%和69.05%，在与其他主流土地覆盖分类方法对比中分类结果最优.

关 键 词：土地覆盖分类；Swin Transformer；高分辨率遥感影像；多尺度特征
中图分类号：P 237 **文献标志码**：A **文章编号**：1005-3026(2026)02-0108-09

Land Cover Classification of the Liaohe Conservation Areas Using Multi-scale Layered Fusion of Swin Transformer Module

WANG Zhi[1], *WANG Meng-qing*[1], *CHEN Li-juan*[2]
(1. School of Resources & Civil Engineering, Northeastern University, Shenyang 110819, China; 2. Shenyang Branch of Liaoning Institute of Water Resources and Hydropower Research Co., Ltd., Shenyang 110003, China. Corresponding author: WANG Zhi, E-mail: wangzhi@mail.neu.edu.cn)

Abstract: In the Liaohe conservation areas, the boundaries of land features are irregular and difficult to discern. To monitor encroachments into arable and forest lands within these areas, a multi-scale layered fusion Swin Transformer module method for land cover classification is proposed. This method includes a hierarchical feature extraction module with stacked Swin Transformer modules to capture both global spatial and local contextual information; a multi-scale feature decoding module for efficient fusion of high-level and low-level features; additional auxiliary decoding heads for more accurate classification; and comprehensive fusion of features from two decoding branches. Validation using a land cover classification dataset created from GF-2 satellite imagery and the publicly available GID-5 dataset shows that the method achieves average intersection-over-union ratio of 82.83% and 69.05%, demonstrating the best classification performance compared to other land cover classification mainstream methods.

Key words: land cover classification; Swin Transformer; high-resolution remote sensing imagery; multi-scale features

辽宁省辽河流域的自然生态系统敏感且脆弱，水资源供应面临日益加剧的压力，有效协调和管理辽河流域的生态和农业空间显得尤为重要.为此，辽宁省全面推进辽河流域的退田还河生态封育修复工作.近年来航天技术的快速发展，容易获取高分辨率遥感影像，其中GF-2号卫

收稿日期：2024-09-26
作者简介：王 植(1979—)，男，辽宁沈阳人，东北大学副教授.
通信作者：王 植，E-mail：wangzhi@mail.neu.edu.cn.

星能够提供具备高灵敏度、高动态范围和高遥感精度的卫星影像，可为辽河封育区监测耕地、林地的侵占行为提供可靠的数据支持.土地覆盖分类是高分辨率遥感影像的重要任务，且其在土地覆盖制图[1−3]、环境监测与保护[4−5]、土地利用规划[6−7]、农业生产管理[8]等方面具有广泛的应用.但是精细的空间细节信息会导致同一地物内部的异质性更强，而且高分辨率遥感影像的地物光谱可分性降低，类间相似性提高，增加了土地覆盖分类的难度[9−10].

利用高分辨率遥感影像识别土地覆盖特征的方法主要有传统机器学习方法和深度学习方法.传统机器学习方法有支持向量机[11]、决策树[12]、随机森林[13]、人工浅层神经网络[14]等.Huang等[15]结合3种机器学习方法实现城市级土地覆盖分类.上述机器学习算法在分类过程中不适用复杂的非线性关系，较难处理地物存在的丰富的空间结构信息和复杂的光谱纹理信息，而高分辨率遥感影像地物细节丰富，成像条件复杂并且对数据质量要求较高，容易受到噪声干扰，无法实现准确的分类结果.

由于强大的特征提取能力和泛化性，深度学习方法逐渐成为遥感影像土地覆盖分类的主流方法.目前主要有以下两类语义分割模型：①基于卷积神经网络的语义分割模型.如用于遥感影像土地覆盖分类的全卷积神经网络(FCN)[16]能进行端到端的训练，不需要手动设计特征提取器，具有压缩路径的编码器和扩展路径的解码器，但由于解码部分过于简单，导致边界细节丢失，难以处理稀疏特征.为了解决这个问题，Ronneberger等[17]提出了U−Net网络，其对称的U型架构和跳跃连接设计，能够高效捕捉和恢复多尺度特征，但其对地物尺度变化适应性较差.此后，一系列卷积神经网络模型被提出，如DeeplabV3+[18]，SegNet[19]，PSPNet[20]，DANet[21]等，为后续的研究提供了新思路.Aihichre等[22]使用连续的卷积层来扩大感受野，利用注意力机制将计算新的特征图作为这些原始特征图的加权平均值，提高了多尺度特征提取能力.He等[23]在DeeplabV3+基础上改进，将MobileNetV2作为骨干网络以减少参数数量，并引入了双注意力机制.Ma等[24]将模糊逻辑单元引入卷积神经网络中，以处理高分辨率遥感影像的模糊性和不确定性.这些卷积神经网络都依赖于卷积操作，无法有效提取全局特征.②基于变换器(Transformer)的语义分割模型.ViT(vision transformer)[25]首次应用于计算机视觉领域，只使用Transformer模块作为编码器提取特征，通过引入自注意力机制[26]更容易学习和捕捉全局位置信息.Xie等[21]提出一种简单高效且无需位置编码的SegFormer，并结合了局部注意力和全局注意力.Liu等[28]提出Swin Transformer，利用移动窗口将自注意力计算限制在不重叠的局部窗口，同时还允许跨窗口连接，这种分层架构能够灵活地在各种尺度上建模，并且对图像大小具有线性计算复杂度.Wang等[29]基于Transformer解码器构建类似U−Net结构的UNetFormer，开发了一种高效的全局−局部注意力机制来模拟解码器中的全局和局部信息.Ma等[30]采用双流结构来最大限度整合Transformer和卷积模块两个分支提取的信息.Hu等[31]提出SABNet，其自注意力双边网络主干由修改的ViT和堆叠的卷积层组成，用于提取全局空间信息和局部上下文信息.

为进一步提高高分辨率遥感影像土地覆盖分类精度，满足监测辽河封育区内耕地和林地侵占行为的需要，针对具有相似特征容易混淆、小尺度地物的边缘细节特征识别困难和难以充分提取全局信息并与局部特征融合等问题，本文提出多尺度分层融合Swin Transformer的模块分类方法.该方法将不同层级Swin Transformer模块提取的多尺度特征进一步处理并融合，在提取全局信息的同时也保留了不同层级的局部特征.

1　研究方法

1.1　网络结构

本文提出的网络模型主要包括特征逐级提取模块、多尺度特征解码模块和特征融合模块.算法网络总体框架如图1所示.

特征逐级提取模块包括4个阶段不同数量的Swin Transformer模块，利用其局部自注意力机制逐级提取特征；多尺度特征解码模块包含PSP(pyramid scene parsing)模块、FPN(feature pyramid network)模块和FCN模块.PSP模块选择第3层、第4层的特征图，通过不同尺度的池化操作，整合高层次特征图的信息.FPN模块包括侧边连接和卷积层，将更高层次的特征信息和PSP模块处理的多尺度特征传递给更低层次的特征图.为学习到多样化的特征添加FCN解码头，FCN模块处理Swin Transformer模块第3层级的特征信息.特征融合模块通过加权损失的方式将FPN模

块和FCN模块的分类结果融合.

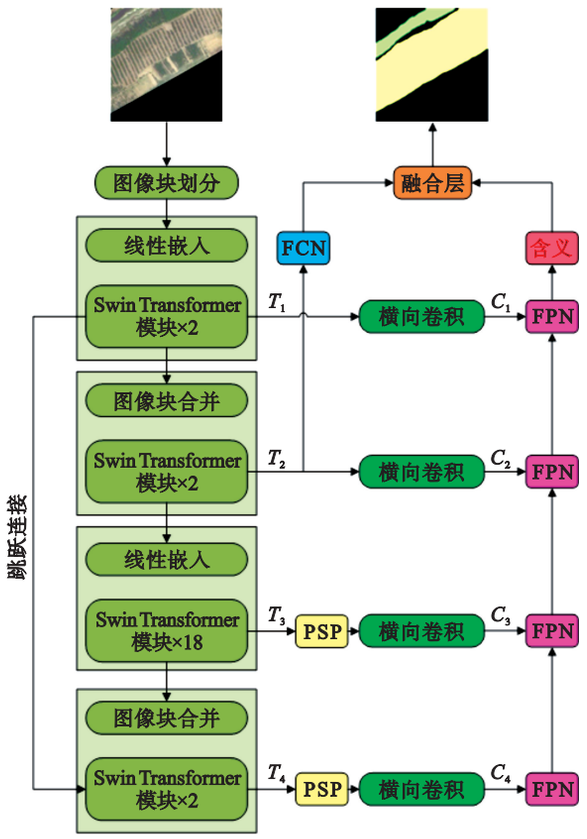

图1 算法网络总体框架

Fig. 1 Overall framework of the algorithm network

1.2 特征逐级提取模块

为更好地提取高分辨率遥感影像所包含的细节信息,Swin Transformer作为特征提取主干模块,通过引入局部窗口自注意力机制以及逐层的特征提取模式,使网络能更好地处理局部特征和全局特征.该主干网络由4个阶段组成,每个阶段堆叠了N个Swin Transformer模块,N值分别为2,2,18和2,每个Swin Transformer模块使用跳跃连接,每个阶段生成的特征图分别为T_1,T_2,T_3,T_4,如图1所示.Swin Transformer逐级提取特征的第1阶段,将输入图像划分为1组不重叠的图像块,使图像处理问题转化为序列建模问题,与第1阶段不同,后面每个阶段在输入模型前需要经过面片合并进行下采样,同时增加通道维度,产生分层表示.

每个Swin Transformer模块都由基于窗口的多头自注意力(windows multi-head self-attention,W_MSA)模块和基于移动窗口的多头自注意力(shifted windows multi-head self-attention,SW_MSA)模块串联组成,具体结构如图2所示.

引入LN(LayerNorm)层对输入特征进行归一化处理,将特征输入到1个通过高斯误差线性单元(gaussian error linear units,GELU)非线性激活函数连接的2个全连接层的多层感知机(multilayer perceptron,MLP),对输入特征进行非线性变换和重新编码,更好地捕捉图像中的长距离关系,每个模块采用残差连接,计算过程如式(1)所示.

$$\left.\begin{aligned}\hat{x}_l&=\mathrm{W\_MSA}\left[\mathrm{LN}\left(x_{l-1}\right)\right]+x_{l-1},\\ x_l&=\mathrm{MLP}\left[\mathrm{LN}\left(\hat{x}_l\right)\right]+\hat{x}_l,\\ \hat{x}_{l+1}&=\mathrm{SW\_MSA}\left[\mathrm{LN}\left(x_l\right)\right]+x_l,\\ x_{l+1}&=\mathrm{MLP}\left[\mathrm{LN}\left(\hat{x}_{l+1}\right)\right]+\hat{x}_{l+1}.\end{aligned}\right\}\tag{1}$$

式中:$\hat{x}_l$和x_l分别为(S)W_MSA和MLP的输出特征;l为Swin Transformer模块的层序号.

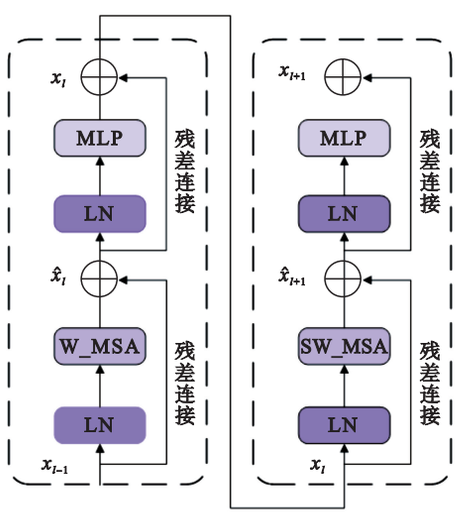

图2 Swin Transformer模块

Fig. 2 Swin Transformer block

1.3 多尺度特征解码模块

来自不同层级的特征图通常具有不同的通道数.例如,高分辨率有更多细节信息的浅层特征图的通道数较少,而低分辨率有更多抽象语义信息的深层特征图的通道数较多.为有效融合这些特征图,须将通道数统一.横向卷积使用1×1卷积将每层特征图调整到相同的通道数,再添加批量归一化(BN)和ReLU激活函数.为捕捉到整个图像的全局特征和不同尺度的上下文信息,将最后层级的深层特征图T_3,T_4通过PSP模块进行多尺度特征池化,利用1个3×3的卷积将池化后的特征图与原始特征图连接起来,形成更丰富的特征表示,具体结构如图3所示.PSP模块的优势在于其模块化设计、多尺度特征提取、高效的通道融合、灵活的配置管理,以及通过瓶颈层在减少计算量的同时提升了模型的表达能力.这些优势使该网络在语义分割任务中能具有更好的表现和更强的适应性,能从多尺度的信息中提取有效特征,进一步提高模型的准确率和效率.

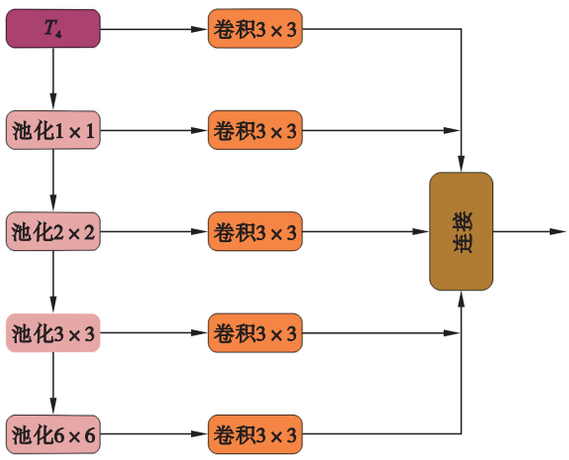

图3　PSP模块结构示意图

Fig. 3　Schematic diagram of the PSP block structure

每1层级经过卷积操作处理后的特征图为C_1,C_2,C_3,C_4,如图1所示.FPN模块先将4个特征图逐一通过双线性插值进行上采样,例如将C_4调整为与特征图C_3相同的大小并与C_3相加.FPN模块对每个层级的特征图进行3×3的卷积,在图像周围填充1层0值像素保持特征图的大小不变,再添加BN和ReLU,具体结构如图4所示.FPN的核心优势是通过金字塔结构和特征图的融合,提升不同尺度下物体的识别能力,尤其是在语义分割任务中,能够有效利用低层次的细节信息和高层次的语义信息,实现跨尺度特征融合和细化,提高模型在不同尺度下的表现能力.FBN是最后的卷积模块,包含3×3的卷积将所有特征图连接起来,并应用了相同的填充大小、归一化和激活函数配置.

添加辅助解码头FCN,能够端到端地处理输入图像.为了获得既具有较高的空间分辨率又能保持语义信息丰富度的同时,FCN模块处理特征图T_3.FCN模块包括2个3×3的卷积层,并对输入特征图T_3和处理后的特征图在通道维度上级联,再经过额外的3×3卷积层处理.

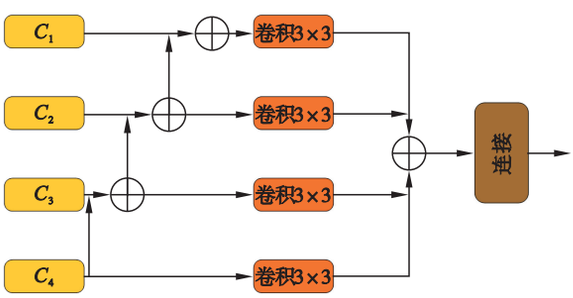

图4　FPN模块结构示意图

Fig. 4　Schematic diagram of the FPN block structure

1.4　特征融合模块

在训练过程中,主解码头FBN模块从每个特征层提取特征,进行多尺度池化、卷积、上采样,辅助解码头FCN模块从第3特征层提取特征进行卷积处理,生成辅助的分割结果.2个解码头通过交叉熵损失函数分别计算损失,如式(2)所示:

$$L_{\mathrm{CE}}=-\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{C}y_{i,j}\lg(p_{i,j}). \tag{2}$$

式中:N为样本总数;C为类别总数;$y_{i,j}$为样本i属于类别j的真实标签;$p_{i,j}$为模型对样本i属于类别j的预测概率.最终的损失是主解码头和辅助解码头的损失加权求和,达到最小化综合损失.

2　数据集和模型参数设置

2.1　构建辽河封育区数据集

辽宁省河流封育区范围为辽宁省内重要河流两侧约1 km的河岸区域,总面积约8.93万公顷,包含台安县、辽中区等30余个县区.基于GF-2号遥感卫星数据构建覆盖封育区的遥感影像数据集,为实现不同时相封育区的土地覆盖分类,数据集共包括2023年5月、6月、7月和10月全覆盖封育区的影像.遥感数据可能受到来自大气、云层以及地表表面的各种干扰,会影响分析的准确性,影像预处理是不可或缺的步骤,具体流程主要包括辐射定标、大气校正、正射校正以及图像融合.原始影像经过预处理后结合封育区的矢量图进行裁剪,为方便后续的处理与分析,将30个县区的矢量图裁剪成以县区为单位的小区域.同时受计算机性能以及网络输入影像大小限制,对影像作切片处理,切片大小为512像素×512像素.此外封育区的土地类型较少,主要包括耕地、林地和荒地,为实现监测封育区的土地覆盖类型并考虑到封育区覆盖范围较大、数据复杂度高、地物辨识难度较大(尤其是耕地、荒地和草地)等实际情况,本文只标注耕地和林地,其余全部定义为背景.

制作影像分类标签过程如下:利用ArcMap软件分别建立耕地和林地的矢量图层,在GF-2号影像上根据影像特征进行目视解译和手工标注,制作覆盖整个辽河封育区的耕地和林地标签,剩余区域全部划为背景,最后将矢量数据栅格化处理并赋予对应的灰度值,得到覆盖整个区域的灰度图,其中代表耕地的灰度值为255,表示林地的灰度值为155,背景灰度值全部为0.具体而言,数据集包括8 547张空间分辨率为1 m(512×512像素)的正射影像及其对应的标签,其中6 548张用于训练,1 000张用于验证,999张

用于测试.数据集部分影像和标签如图5所示.

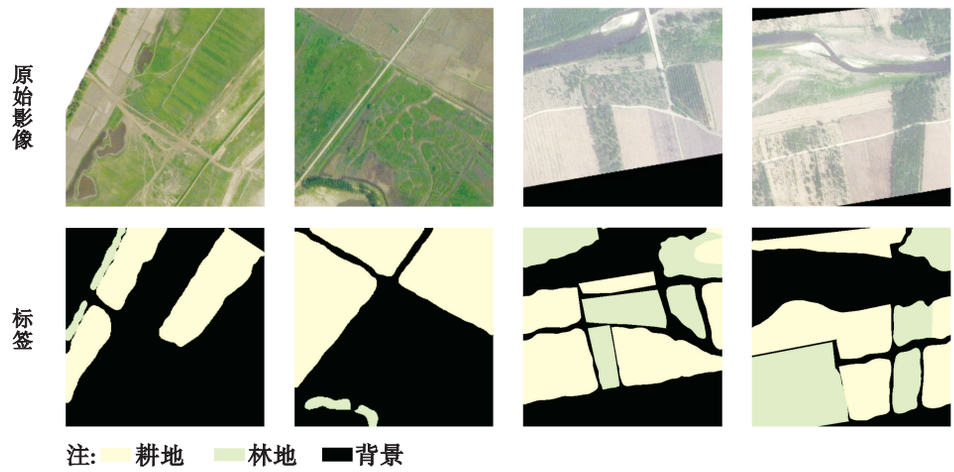

图5 辽河封育区数据集部分原始影像和标签示例

Fig. 5 Partial original images and annotated examples from the Liaohe ecological conservation zone dataset

2.2 公开数据集GID-5

利用公开的GID-5[32]大规模分类数据集横向验证提出模型的可迁移性和通用性.GID-5是基于GF-2号卫星数据构建的大规模高分辨率遥感图像土地覆盖数据集,具有覆盖广泛、注释精细和空间分辨率高等优势.数据集包含150景7 300像素×6 908像素的图像及其对应的标签,将每景影像划分为512像素×512像素的非重叠图像块,共获得31 500对对照影像,其中26 460张图片用于训练,2 520张用于验证,2 520张用于测试.部分原始影像及其标签如图6所示.

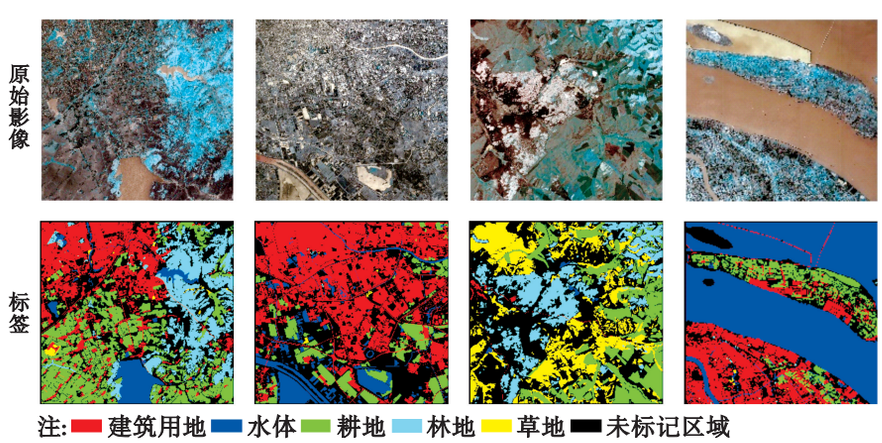

图6 GID-5数据集部分原始影像和标签示例

Fig. 6 Partial original images and annotated examples from the GID-5 dataset

2.3 实验环境与评价指标

本文使用的硬件配置和实验开发环境如表1所示.

表1 实验开发环境

Table 1 Experimental development environment

平台组成	配置
系统	Windows10
CPU	Intel(R) Core(TM) i7-6900K CPU@ 3. 20GHz
GPU	NVIDIA GeForce GTX 1080
RAM	48. 0 GB
开发编程语言	Python 3. 8. 18
深度学习框架	PyTorch 2. 3. 1+cu118

在搭建网络的过程中使用Adam优化器和BN优化网络,加快收敛速度、提高收敛精度和提高模型的泛化能力.初始学习率为0.000 06,权重衰减系数设为0.01,控制模型参数更新时的正则化.在训练初期使用LinearLR调度器,随着训练的进行,使用PolyLR调度器逐渐降低学习率以细致调整模型参数,更精确地找到全局最优解,两种调度器配合可以有效平衡模型的训练速度和性能,提高整体训练效率和结果质量,更新方式如式(3)所示:

$$L = L_1 \times \left(1 - \frac{e}{M_e}\right)^p. \tag{3}$$

式中：L为当前迭代的学习率；L_1为初始学习率；e为当前已经训练的轮次；M_e为总训练轮次；p为衰减幂次.

此外，在训练开始前还需要进行一定的数据预处理，如式(4)所示：

$$I_c=\frac{I_c-\mu_c}{\sigma_c}. \tag{4}$$

式中：I_c为3个RGB不同通道的像素值；μ_c和σ_c分别为对应通道的均值和标准差.

本文使用常用的语义分割评价指标来评估所用方法的性能，包括总体准确率、平均交并比和平均准确率.

$$a=\frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{FP}+\mathrm{FN}}, \tag{5}$$

$$m_1=\frac{1}{M}\sum_{i=1}^{M}\frac{\mathrm{TP}_i}{\mathrm{TP}_i+\mathrm{FP}_i+\mathrm{FN}_i}, \tag{6}$$

$$m_A=\frac{1}{M}\sum_{i=1}^{M}\frac{\mathrm{TP}_i+\mathrm{TN}_i}{\mathrm{TP}_i+\mathrm{TN}_i+\mathrm{FP}_i+\mathrm{FN}_i}. \tag{7}$$

式中：TP，TN，FP，FN分别代表真正例数、真反例数、假正例数和假反例数；M为数据集中的分类类别数；i为特定的类别；a为总体准确率；m_1为分割结果与真实分割结果的平均交并比；m_A为网络模型所有分类类别的平均准确率.

3　实验结果与分析

3.1　辽河封育区数据集的结果与分析

为了证明提出方法的有效性，在辽河封育区数据集中使用2.3节中提到的评价指标定量分析使用方法的分类性能.选择与使用网络具有相似结构的其他主流的土地覆盖分类方法，包括DeepLabv3+，PSPNet和SegFormer，并对每个网络进行对比分析.其中DeepLabv3+代表编码器-解码器结构，PSPNet代表带有金字塔集成方法的网络，SegFormer代表使用Transformer结构的网络.定量评估结果见表2.

表2　不同模型在辽河封育区数据集中的分类性能比较

Table 2　Comparison of the classification performance of different models on the Liaohe ecological conservation area dataset　　%

方法	耕地		林地		背景		a	m_1	m_A
	交并比	准确率	交并比	准确率	交并比	准确率			
DeepLabv3+	56.12	65.08	65.80	81.75	88.29	95.55	89.76	70.07	80.79
PSPNet	64.86	74.15	68.98	79.77	90.38	96.25	91.77	74.74	83.39
SegFormer	66.14	80.74	67.01	82.92	89.78	94.15	91.38	74.31	85.94
本文方法	**78.11**	**87.26**	**76.38**	**85.62**	**94.01**	**97.31**	**95.56**	**82.83**	**90.06**

注：加粗表示最优，下同.

从表2中可以看出，本文提出的方法在3项总体评估中取得了最佳结果，a提高了4.13%~6.46%，m_A提高了4.47%~11.13%，主要指标m_1比代表基本编码器-解码器结构的DeepLabv3+提高了18.21%，比具有Transformer结构的SegFormer提高了11.47%，比多尺度金字塔结构的PSPNet提高了10.82%.表2统计了每个类别的分类精度，从中可以看出，提出方法在各个类别分类中均取得了最佳结果，且相较于其他主流的分类方法有了显著提升.但是仍然存在一定的错分问题，为详细分析分类的混淆情况，计算出分类结果的混淆矩阵如表3所示，总共的测试样本有261 881 856像素.从表中可以看出，模型在3个类别之间都存在一定的误分情况.不排除可能在制作数据集时存在标签误注问题，同时由于封育区地物边界不明显，尤其是耕地和荒地，林地与耕地无明显分界线，甚至交叉分布，在特征上具有相似性，使封育区分类识别难度增大.在后续的研究中，可进一步优化模型的特征提取方法、增加更多具有代表性的训练样本，以提高模型的分类准确性，减少各类别之间的混淆.

表3　模型分类结果的混淆矩阵

Table 3　Confusion matrix of model classification results　　像素

类别	耕地	林地	背景
耕地	68 565 333	5 262 749	4 736 475
林地	11 091 133	66 591 329	882 095
背景	1 346 071	1 480 678	101 925 993

将不同模型的测试结果进行可视化比较，如图7所示.DeepLabv3+由于其局部与全局特征融合局限、特征提取灵活性不足等问题导致分类结果不理想；而SegFormer存在较大的错分问题是由于其分类时无法充分利用特征之间的互补信息并且提取特征侧重点不同；PSPNet模型以PSP

模块为主体相对取得了较好的分类效果.从图中可以直观地看出结合其他分类模型优点的本文模型具有最佳分类效果,在提取规则形状复杂、边界不清晰的耕地和林地时,比其他算法更具有优势,预测结果在纹理细节和整体上更接近真实标签.

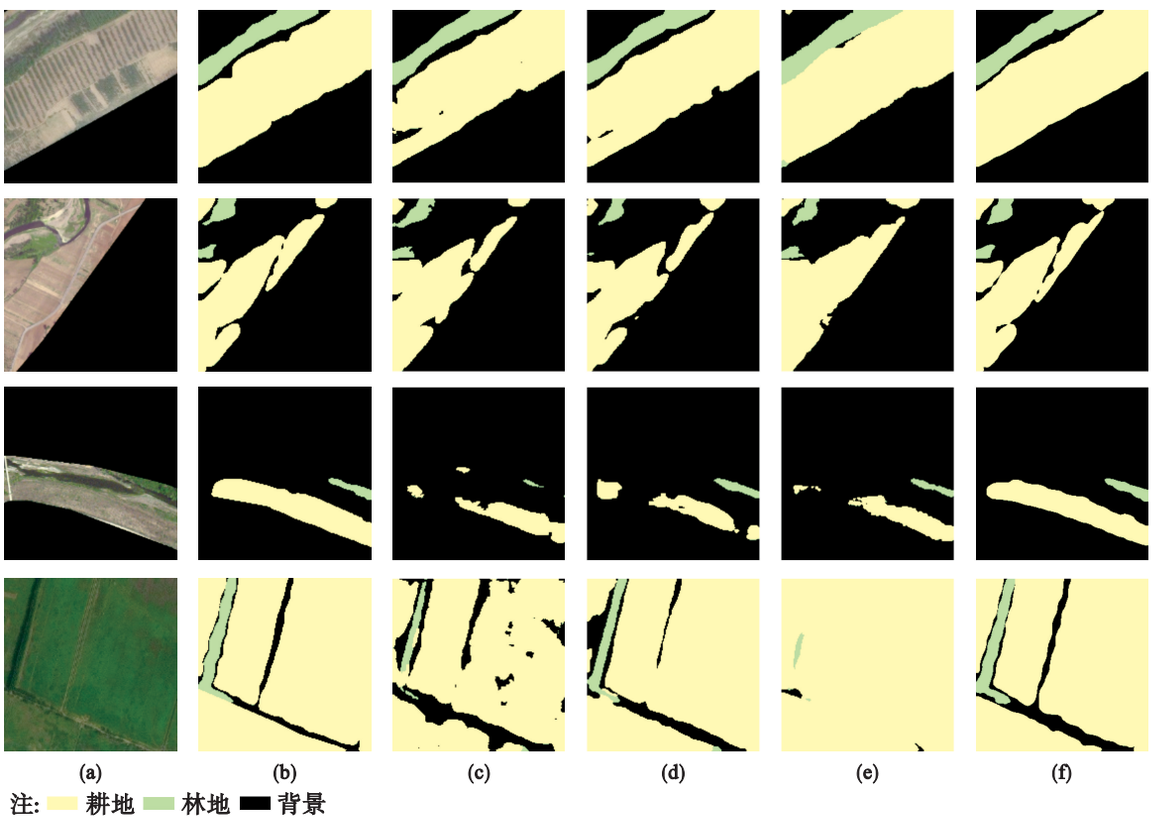

图7 不同模型在辽河封育区数据集中的可视化结果比较

Fig. 7 Comparison of the visualization results of different models on the Liaohe ecological conservation area dataset

(a)—原始影像;(b)—标签;(c)—DeepLabv3+;(d)—PSPNet;(e)—SegFormer;(f)—本文方法.

3.2 在GID-5数据集的结果与分析

为了进一步验证本文网络的泛化能力,在公开数据集GID-5中进行一系列对比实验,定量评估结果如表4所示.

表4 不同算法在GID-5中的分类性能比较

Table 4 Comparison of the classification performance of different algorithms on the GID-5 %

方法	建筑		耕地		林地		草地		水体		未标注		a	m_I	m_A
	交并比	准确率	交并比	准确率	交并比	准确率	交并比	准确率	交并比	准确率	交并比	准确率			
DeepLabv3+	53.84	66.33	72.18	84.25	70.15	79.58	66.33	82.46	34.08	85.60	80.62	92.95	76.83	62.87	81.86
PSPNet	56.05	68.47	71.85	85.36	69.44	78.05	65.27	83.43	39.01	83.27	70.46	90.79	77.06	62.01	81.56
SegFormer	58.12	73.01	70.69	85.39	70.54	77.99	69.40	85.69	48.34	71.18	70.53	92.42	78.67	64.60	80.94
本文方法	**60.29**	**78.53**	**73.63**	**85.46**	**73.41**	**82.47**	**73.59**	**90.27**	**51.41**	**86.73**	**81.98**	**94.56**	**82.27**	**69.05**	**86.34**

从表4中可以看出本文网络在GID-5数据集的3项总体评估中都达到了最优,其中a提高了4.58%~7.08%,m_A提高了5.47%~6.67%,主要指标m_I相比基础编码器-解码器结构的DeepLabv3+提高了9.83%,比代表Transformer的SegFormer模型提高了6.89%,充分说明了本文方法的高泛化能力.

不同模型的可视化结果如图8所示,更直观地显示本文方法与其他方法相比的优势,从图8中可以看出本文方法可以更加准确地提取不连续且没有规则的建设用地,水体及其边界也更加清晰,能够准确提取较难分辨的耕地和草地,证明了本文方法在小特征和容易混淆的特征类别中具有明显优势.

评估各个分类方法的模型性能和复杂度.评估指标有每秒浮点运算次数(floating point operations,FLOPs)和参数数量,如表5所示.从表中可以看出本文方法的计算量和学习到的参数数量均明显大于其他的分类方法,这种额外的计算量和参数数量带来了显著的性能改善,模型具有更好地学习复杂特征和进行更深层次的特征提取方面的能力,同时需要充足的计算资源来满足较高的计算成本.本文网络的主要骨干特征提取网络在处理不同分辨率图像时达到最优,特别是在高分辨率图像上,例如曾在ADE20K数据集中的取得突破性分类进展;同时在多个大型数据集如UCM数据集、AID数据集、ADE20K数据集等中取得优异的结果.模型通过骨干网络的分层架构和移位窗口,能够在计算复杂度和分类性能之间取得平衡,模型强大的特征提取能力和分类性能使其在处理大规模和复杂数据集时具有显著优势,在多种实际应用场景中具有广泛的应用前景.

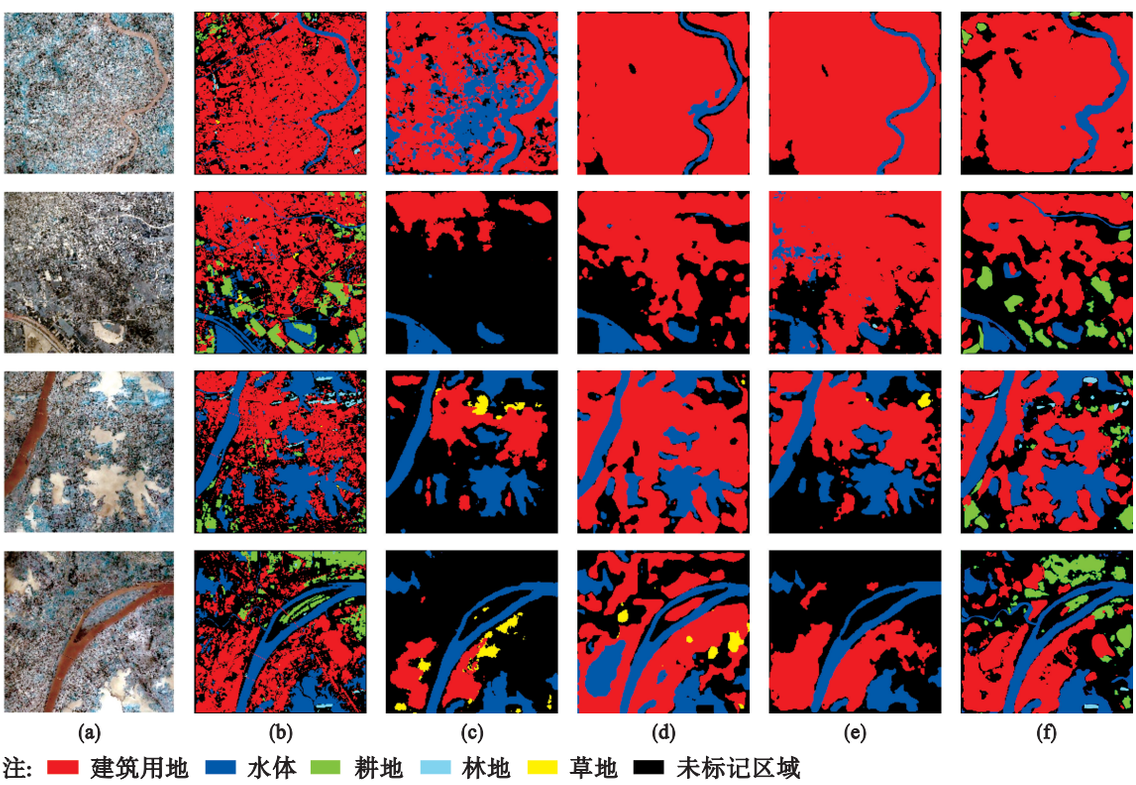

图8　不同模型在GID-5中的可视化结果比较

Fig. 8　Comparison of the visualization results of different models on the GID-5

(a)—原始影像;(b)—标签;(c)—DeepLabv3+;(d)—PSPNet;(e)—SegFormer;(f)—本文方法.

表5　计算量和参数量对比

Table 5　Comparison of computational complexity and number of parameters

方法	浮点运算/(次·s^{-1})	参数量×10^{-7}
DeepLabv3+	0.177	41.219
PSPNet	0.179	46.603
SegFormer	0.025	24.724
本文方法	0.298	122.88

4　结　　语

本文方法在复杂的辽河封育区数据集和公开数据集GID-5中取得了明显优于其他主流土地覆盖分类方法的分类结果.由显著提升的定量评估和视觉效果方面可以看出本文网络可以更加准确提取不连续且没有规则的地物,准确提取较难分辨的耕地、草地和荒地,能清晰描绘具有复杂边界的水体,在识别相似特征、复杂特征、纹理细节和全局特征中表现良好,同时导致了模型的计算量和参数量增大,需要更多的计算资源.

参考文献:

[1] Li R, Zheng S Y, Duan C X, et al. Land cover classification from remote sensing images based on multi-scale fully convolutional network [J]. *Geo-spatial Information Science*, 2022, 25(2): 278-294.

[2] Yin H, Pflugmacher D, Li A, et al. Land use and land cover change in Inner Mongolia-understanding the effects of China's re-vegetation programs [J]. *Remote Sensing of Environment*, 2018, 204: 918-930.

[3] Tu Y, Chen B, Zhang T, et al. Regional mapping of essential urban land use categories in China: a segmentation-based approach[J]. *Remote Sensing*, 2020, 12(7): 1058.
[4] Cao Y X, Huang X. A coarse-to-fine weakly supervised learning method for green plastic cover segmentation using high-resolution remote sensing images[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2022, 188: 157-176.
[5] Johnson B A, Ma L. Image segmentation and object-based image analysis for environmental monitoring: recent areas of interest, researchers' views on the future priorities[J]. *Remote Sensing*, 2020, 12(11): 1772.
[6] Enoguanbhor E C, Gollnow F, Nielsen J O, et al. Land cover change in the Abuja city-region, Nigeria: integrating GIS and remotely sensed data to support land use planning [J]. *Sustainability*, 2019, 11(5): 1313.
[7] Shi H, Chen L, Bi F K, et al. Accurate urban area detection in remote sensing images[J]. *IEEE Geoscience and Remote Sensing Letters*, 2015, 12(9): 1948-1952.
[8] Sassu A, Gambella F, Ghiani L, et al. Advances in unmanned aerial system remote sensing for precision viticulture[J]. *Sensors*, 2021, 21(3): 956.
[9] Lu T, Li S T, Fang L Y, et al. From subpixel to superpixel: a novel fusion framework for hyperspectral image classification[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2017, 55(8): 4398-4411.
[10] Sun B, Kang X D, Li S T, et al. Random-walker-based collaborative learning for hyperspectral image classification [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2016, 55(1): 212-222.
[11] Sargent I, Zhang C, Atkinson P M. Joint deep learning for land cover and land use classification: US10984532 [P]. 2021-04-20.
[12] Papoutsis I, Bountos N I, Zavras A, et al. Benchmarking and scaling of deep learning models for land cover image classification [J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2023, 195: 250-268.
[13] Manzanarez S, Manian V, Santos M, et al. Land use land cover labeling of GLOBE images using a deep learning fusion model[J]. *Sensors*, 2022, 22(18): 6895.
[14] Kroupi E, Kesa M, Navarro-Sánchez V D, et al. Deep convolutional neural networks for land-cover classification with Sentinel-2 images [J]. *Journal of Applied Remote Sensing*, 2019, 13(2): 024525.
[15] Huang C Q, He C, Wu Q, et al. Classification of the land cover of a megacity in ASEAN using two band combinations and three machine learning algorithms: a case study in Ho Chi Minh City[J]. *Sustainability*, 2023, 15 (8): 6798.
[16] Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation[J] *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2017,39(4): 640-651.
[17] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation[C]//Medical Image Computing and Computer-Assisted Intervention 2015. Cham: Springer, 2015: 234-241.
[18] Chen L C, Papandreou G, Schroff F, et al. Rethinking atrous convolution for semantic image segmentation[EB/OL]. (2017-12-05)[2024-07-23]. https://arxiv. org/abs/1706.05587.
[19] 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.
[20] 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 (CVPR). Honolulu: IEEE, 2017: 6230-6239.
[21] 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: IEEE, 2020: 3141-3149.
[22] Alhichri H, Alswayed A S, Bazi Y, et al. Classification of remote sensing images using EfficientNet-B3 CNN model with attention[J]. *IEEE Access*, 2021, 9: 14078-14094.
[23] He C, Liu Y L, Wang D C, et al. Automatic extraction of bare soil land from high-resolution remote sensing images based on semantic segmentation with deep learning [J]. *Remote Sensing*, 2023, 15(6): 1646.
[24] Ma X Y, Xu J D, Chong Q P, et al. FCUnet: refined remote sensing image segmentation method based on a fuzzy deep learning conditional random field network [J]. *IET Image Processing*, 2023, 17(12): 3616-3629.
[25] Khan S, Naseer M, Hayat M, et al. Transformers in vision: a survey [EB/OL]. (2022-01-19)[2024-07-23]. https://arxiv.org/pdf/2101.01169.
[26] Ashish V. Attention is all you need[EB/OL]. (2023-08-02) [2024-07-23]. https://arxiv. org/abs/1706.03762? ref= andrealarosa.org.
[27] Xie E Z, Wang W H, Yu Z D, et al. SegFormer: simple and efficient design for semantic segmentation with transformers [EB/OL]. (2021-10-28) [2024-05-21]. https://arxiv.org/abs/2105.15203.
[28] 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: IEEE. 2022: 9992-10002.
[29] Wang L, Li R B, 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.
[30] Ma J G, Shen H R, Cai Y X, et al. UCTNet with dual-flow architecture: snow coverage mapping with Sentinel-2 satellite imagery[J]. *Remote Sensing*, 2023, 15(17): 4213.
[31] Hu Z H, Qian Y R, Xiao Z Q, et al. SABNet: self-attention bilateral network for land cover classification[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2024, 17: 8559-8569.
[32] Tong X Y, Xia G S, Lu Q K, et al. Land-cover classification with high-resolution remote sensing images using transferable deep models [J]. *Remote Sensing of Environment*, 2020, 237: 111322.