

文章编号 1004-924X(2025)19-3121-14

全向矩形校准的高分辨遥感影像细节分割

陈思睿, 胡方敏, 王洪亮, 谢 涛*

(昆明理工大学 民航与航空学院, 云南 昆明 650500)

摘要:针对高分辨率遥感影像分割中特征提取不全、边缘模糊及小目标漏检等难题,提出一种基于全向矩形校准的高分辨遥感影像细节分割方法,构建全向矩形校准网络(Omnidirectional Rectangular Calibration Network, ORCNet)。首先,设计全向矩形校准状态空间模块(Omnidirectional Rectangular Calibration State Space Module, ORSM),通过八向扫描与几何敏感权重校准,提升特征几何适应性与目标保留能力。随后,构建互补滤波混合特征融合模块(Complementary Filtering Hybrid Attention Fusion Module, CFHAF)融合通道、空间与像素级注意力机制,实现多尺度特征的自适应融合,增强语义感知。最后,融合动态点上采样技术(Dynamic Point Upsampling, DySample)优化边界细节恢复能力,结合 Focal-Dice 混合损失函数进行训练优化。实验结果表明,ORCNet 在 Massachusetts buildings 数据集上 F1score 达到 84.64%, mIoU 达到 77.07%, 在 deepglobe-road-dataset 上, F1score 达到 85.32%, 较 RSMamba 提升 3.51%。该方法有效克服了遥感分割中存在的问题,具备显著的高精度、强稳定性和优秀的实际应用潜力。

关键词:遥感影像分割;矩形自校准;状态空间模型;多尺度特征融合;边缘细节优化;动态上采样

中图分类号: TP79 文献标识码: A

doi:10.37188/OPE.20253319.3121

CSTR:32169.14.OPE.20253319.3121

Detailed segmentation of high-resolution remote sensing images calibrated with omnidirectional rectangle

CHEN Sirui, HU Fangmin, WANG Hongliang, XIE Tao*

(School of Civil Aviation and Aeronautics, Kunming University of Science and Technology,
Kunming 650500, China)

* Corresponding author, E-mail: 77183118@qq.com

Abstract: To address incomplete feature extraction, blurred boundaries, and omission of small targets in high-resolution remote sensing image segmentation, a detail-preserving segmentation method based on omnidirectional rectangular calibration, termed the Omnidirectional Rectangular Calibration Network (ORCNet), is proposed. First, an Omnidirectional Rectangular Calibration State Space Module (ORSM) is introduced to enhance geometric adaptability and target retention via octagonal scanning and geometry-sensitive weight calibration. Next, a Complementary Filtering Hybrid Attention Fusion Module (CFHAF) is developed, integrating channel-, spatial-, and pixel-level attention mechanisms to enable adaptive multi-scale feature fusion and improved semantic discrimination. Finally, Dynamic Point Upsampling (DySample) is incorporated to refine boundary detail recovery. The model is trained with a Focal-Dice hybrid

收稿日期: 2025-07-01; 修订日期: 2025-07-28.

基金项目: 国家自然科学基金资助项目(No. 62163021); 云南省基础研究专项资助项目(No. 202301AT070420)

loss. On the Massachusetts Buildings dataset, the method achieves an F1 score of 84.64% and an mIoU of 77.07%. On the DeepGlobe Road dataset, an F1 score of 85.32% is obtained, representing a 3.51% improvement over RSMamba. Experimental results indicate that the proposed approach effectively addresses the three primary challenges in remote sensing segmentation, delivering high precision, robust performance, and strong potential for practical application.

Key words: remote sensing image segmentation; rectangular self-calibration; state space model; multi-scale feature fusion; edge detail enhancement; dynamic upsampling

1 引言

高分辨率光学遥感卫星作为现代对地观测体系的核心载荷,其亚米级空间分辨率为实现地表精细识别提供了数据基础。随着我国高分专项等重大工程的实施,国产遥感数据质量已进入国际先进行列,但成像分辨率的提升导致数据量呈现指数级增长,传统人工解译方法面临巨大挑战。当前高分辨率遥感影像分割主要存在以下3个问题:(1)小目标漏检:在建筑密集区域,由于目标尺寸较小且易受周围地物遮挡,传统方法在识别小目标时漏检率较高;(2)特征提取不全:遥感影像中存在光谱特征高度相似的区域,如玻璃幕墙与水体,导致传统模型难以有效区分,特征提取不完整,影响语义分割的准确率;(3)边缘模糊:高分辨率影像对小尺度特征和边界细节灵敏度较高,但传统方法如双线性插值无法有效恢复这些信息,导致分割结果边界模糊、结构松散。这对高分辨遥感影像智能化、自动化目标提取技术提出了更高要求^[1-3]。

现有的影像分割研究可分为传统图像处理与深度学习两大技术路线。传统方法依赖先验知识,包括数学形态学^[4]、马尔可夫随机场^[5]等,适用于规则目标,但难以处理复杂非线性特征^[6-7]。深度学习方法分为三个阶段:卷积神经网络主导期的 DeepLabv3+^[8]和 U-Net^[9], Transformer 时期的 Swin-Unet^[10],及 VMamba^[11]等基于选择性状态空间模型的架构,以实现线性计算与长程依赖建模的平衡。

在遥感影像分割领域,研究人员提出了很多创新解决方案。李智杰^[12]等通过结合空间注意力和时间注意力机制设计了空间特征提取模块,但仍存在特征提取不全和道路拓扑结构还原细节不足的问题。刘帅^[13]等通过双解耦策略改善

了边缘细节保留,但边界模糊和小尺度目标漏检问题仍未完全解决。此外,遥感影像中目标尺度多变,要求算法具备多尺度特征处理能力。ZHANG^[14]等采用多尺度特征融合策略提升了模型对尺度变化的敏感性,但在处理极端尺度变化和相似目标时仍存在显著的误分现象。随着技术的更迭,注意力机制被广泛用于遥感影像分割领域^[15-16],通过动态特征权重调整增强关键特征表示。最新的研究当中,ZHU^[17]等引入时空线索增强对作物区域的感知,LI^[18]等依靠面训练的开放词汇语义分割网络,通过 SimFeatUp 上采样实现多种遥感任务的广泛实验。然而,复杂场景下的关键特征定位与语义混淆仍是当前研究的主要瓶颈。

综上所述,尽管智能高分辨率遥感影像分割技术已取得显著进展,但当前主流方法仍存在小目标漏检、特征提取不全、边缘模糊等瓶颈问题。本文针对上述问题,提出了全向矩形校准网络(Omnidirectional Rectangular Calibration Network, ORCNet),将全向矩形校准状态空间模块(Omnidirectional Rectangular Calibration State Space Module, ORSM)作为主干网络,设计特征融合模块(Complementary Filtering Hybrid Attention Fusion Module, CFHAF)高效融合多尺度特征,结合动态点上采样模块(Dynamic Point Upsample, DySample)进一步提升了特征表达能力,显著改善了边缘信息丢失问题,增强了遥感影像分割精度。

2 原理

2.1 网络整体结构

ORCNet 结构如图 1 所示。在编码部分设计全向矩形自校准选择性扫描模块 ORSM 作为主

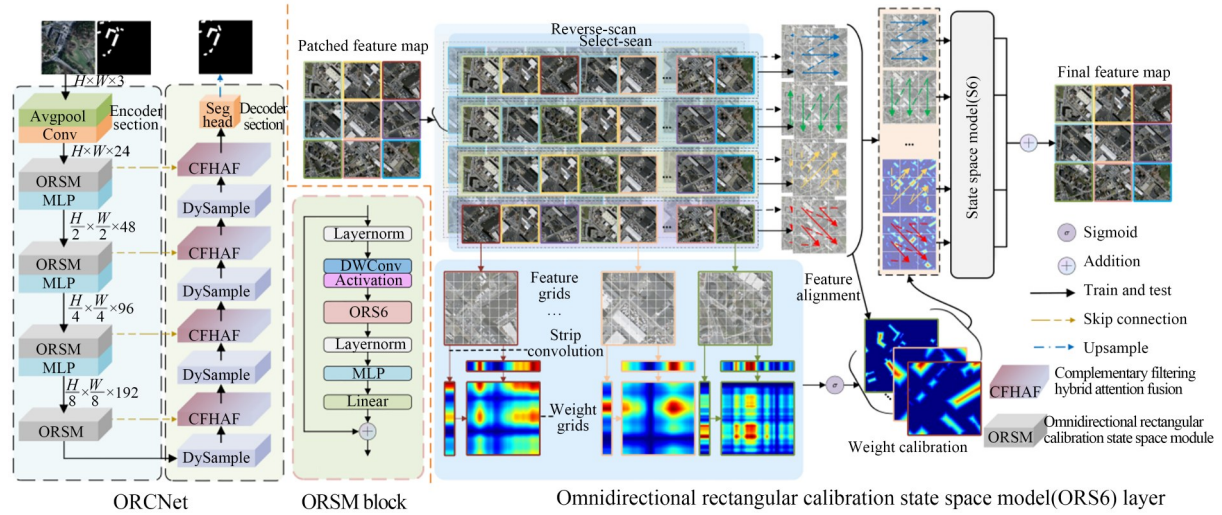


图 1 ORCNet 及核心组件结构

Fig. 1 ORCNet and core component structure

干网络,提取遥感影像的主要特征信息。在解码部分,通过CFHAF融合多尺度特征并强化核心特征权重,DySample优化细节恢复能力,增强网络特征表达能力,从而提升多尺度目标的分割效果。

2.2 全向矩形自校准选择性扫描模块

针对遥感影像中存在的目标尺度多样性问题及传统卷积网络易丢失小目标信息的缺陷,本文在RSMamba^[19]模型全向选择性扫描机制(Omnidirectional Selective Scan Mechanism, OS-SM)的基础上,提出ORSM。该模块通过方向感知的特征解耦与几何自适应的细节增强相结合,实现多尺度目标的高效表征,其核心在于将全局上下文建模并与局部几何特征校准相融合。

该模块以层归一化开始,通过深度可分离卷积和激活函数提取特征,再经ORS6层建模全局-局部依赖关系,接着经过层归一化、多层感知机(Multilayer Perceptron, MLP)和线性层,最后与原输入进行残差连接输出。该层通过8个方向的选择性扫描机制将输入特征图沿宽、高及对角线方向进行双向序列化解耦,生成方向敏感的patch组 S_i ,得到:

$$S_i = \text{flatten}(X, i), \quad (1)$$

其中: $X \in \mathbb{R}^{B \times C \times H \times W}$ 为输入特征图, $i \in (1, 2, \dots, 8)$ 表示从八个不同方向进行扫描, S_i 为序列化扫描后各组Patch的集合。

每组patch通过状态空间模型(State Space

Model, SSM, S6)建立全局特征关联,捕获长程空间依赖关系。该模型在设计上继承了RS-Mamba中OSSM模块的全局建模优势,但此种方式只能学习patch之间的关系,并不能有效获取patch内部的特征依赖关系。受建筑物与道路目标边缘连续且线性,具有显著梯度变化的特性启发,本文集成了具有通道自适应的矩形自校准注意力机制(Rectangular Self-Calibration Attention, RCA)^[20],对每个patch内部的特征进行边缘细节提取。先使用水平池化和垂直池化分别来计算每个patch的轴向全局上下文,生成两个独立轴向向量。再通过广播加法建模矩形感兴趣区域后,采用条带卷积(Strip Convolution, $k \times 1$ 和 $1 \times k$ 核)以解耦方式生成轴向注意力权重图 W_{ij}^h 和 W_{ij}^v 。最终,通过外积运算构建几何敏感权重图 W_{ij} ,即:

$$\begin{cases} W_{ij}^h = \text{Conv}_{k \times 1}(\text{Horizontalpool}(S_i)) \\ W_{ij}^v = \text{Conv}_{1 \times k}(\text{Verticalpool}(S_i)) \\ W_{ij} = \text{softmax}(W_{ij}^h \otimes W_{ij}^v) \end{cases}, \quad (2)$$

其中: S_i 来自全向选择性扫描序列,每个方向对应一组输入特征图; W_{ij}^h 为水平方向权重图, W_{ij}^v 为垂直方向权重图; W_{ij} 为几何敏感权重图,他们均为可学习的权重; $\otimes$ 为矩阵乘法,几何敏感权重采用零初始化,在训练过程中通过残差连接逐步学习几何增强方式; i, j 分别为权重矩阵中的行索引和列索引, i 表示第 i 组权重, j 表示第 i 组权重中的第 j 个元素。

将几何敏感权重图 W_{ij} 与原图进行对齐叠加,二者取内积生成校准权重图 T_i 。校准权重图 T_i 与对应序列 S_i 拼接后通过选择性扫描模块获得最终特征图输出 Y_i 。强化 patch 内部的特征关联以进行边缘细节提取。

$$T_i = S_i \odot W_{ij}, \quad (3)$$

$$Y_i = \text{SSM}(\text{concat}(T_i; S_i)), \quad (4)$$

$$Y = \sum_{i=1}^8 Y_i, \quad (5)$$

其中: T_i 为生成校准后的权重图, Y_i 代表对应序列的最终特征图, Y 代表最终特征图的集合。SSM 模块采用选择性扫描核心,它采用状态空间方程^[21]建立 patch 间的相关性,具体工作原理如下:

$$\begin{cases} h'(t) = Ah(t) + Bu(t) \\ y(t) = Ch(t) + Du(t) \end{cases} \quad (6)$$

其中: $h'(t)$ 是隐藏状态, $h(t)$ 和 $u(t)$ 分别代表隐状态向量和输入信号,状态矩阵、输入矩阵、输出矩阵和直连项分别由 A , B , C 和 D 表示, $A \in \mathbb{R}^{d_{\text{channel}} \times d_{\text{state}}}$ 为固定初始化矩阵, $D \in \mathbb{R}^{d_{\text{channel}} \times d_{\text{channel}}}$, B 和 C 形状均为 $d_{\text{state}} \times d_{\text{channel}}$,三者均为可学习参数,参数量由状态空间的维度 d_{state} 和输入通道数 d_{channel} 共同决定, $y(t)$ 代表最终输出。对连续模型通过欧拉近似^[22]进行时间步长离散化,获得其最终工作原理如下:

$$h_t = (I + \Delta_t A)h_{t-1} + \Delta_t Bu_t, \quad (7)$$

其中: $\Delta_t \in \mathbb{R}^p$ 为自适应步长, h_t 代表第 t 步离散隐状态, u_t 代表第 t 个扫描位置的输入, I 代表单位矩阵。

RSMamba 虽具备优秀的线性复杂度和多尺度表征能力,但在语义边界重建和小目标识别方面仍存在一定局限,尤其无法模拟遥感影像中如建筑物边缘、道路拓扑等具有显著几何特征的区域对细节信息的依赖关系。这是由于其核心模块 OSSM 仅依赖序列化扫描和状态空间模型,但忽略了 patch 内部的几何敏感特征校准问题。因此,本文提出 ORSM 模块,通过全向扫描建立全局特征关联,结合矩形自校准注意力机制,对 patch 内部的空间几何特征进行显式建模,在保持大尺度目标检测精度的同时,显著提升了小目标特征的保留率与表征能力。

2.3 互补滤波混合特征融合模块

在遥感影像分割任务中,高分辨率影像的多尺度特性(如建筑物边缘锐度、道路网络连续性等)对特征融合提出特殊要求。传统特征金字塔网络(Feature Pyramid Networks, FPN)^[23]在浅层特征融合时存在语义混叠和细节模糊化问题,尤其难以平衡大范围上下文与局部精细特征的关系。因此,本文设计 CFHAF 模块通过多级特征补偿与注意力协同机制实现细节增强,如图 2 所示。本文定义了一种自适应加权策略融合深浅层特征如下:

$$F_1 = F_{\text{high}} + w(F_{\text{low}} - F_{\text{high}}), \quad (8)$$

其中: F_1 为加权融合后的输入特征, F_{high} 为深层解码器特征, F_{low} 为浅层解码器特征, w 为混合注意力特征融合模块(Hybrid Attention Feature Fusion Module, HAFFM)学习的动态权重图。

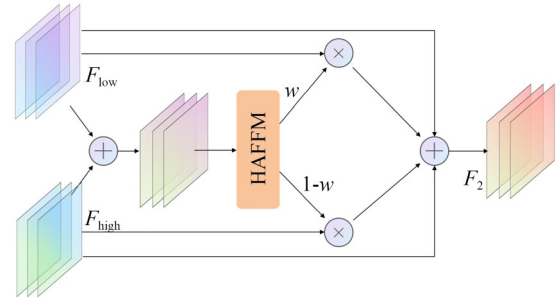


图 2 特征融合模块结构

Fig. 2 Feature fusion module structure

基于残差学习思想^[24],进一步构建特征重组结构,尽可能加强细节特征提取的同时保留原特征,防止特征丢失并降低深度网络训练过程中梯度异常的风险。在实际中,融合过程的式(8)变形为式(9),后续残差结构如下:

$$F_1 = wF_{\text{low}} + (1 - w)F_{\text{high}}, \quad (9)$$

$$F_2 = F_1 + F_{\text{high}} + F_{\text{low}}, \quad (10)$$

其中 F_2 为加权融合后的输入特征。

如图 3 所示,混合注意力模块 HAFFM 通过通道注意模块、空间注意模块和像素注意模块协同实现像素级特征优化。同步在通道维度筛选关键特征和空间维度突出重要区域,最终在像素维度完成细节增强。

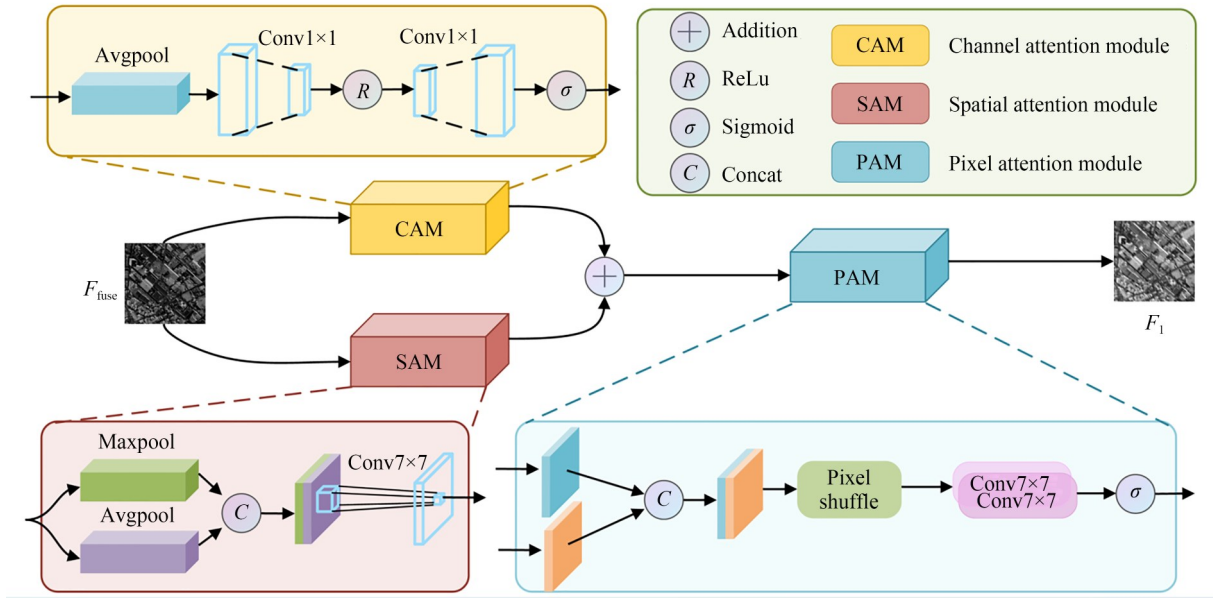


图3 混合注意力机制融合模块结构

Fig. 3 Hybrid Attention Feature Fusion Module(HAFFM) structure

在CAM中,通过全局平均池化将输入特征 F_{fuse} 进行压缩,从而生成通道级统计特征图以捕获全局上下文信息。随后采用 1×1 卷积进行通道降维,通过ReLU激活函数增强非线性表达能力。最后,使用 1×1 卷积恢复原始通道维度并经sigmoid函数归一化,得到通道权重 w_c ,其数学表达式如下:

$$w_c = \sigma(\text{Conv}_{1 \times 1}(\text{ReLU}(\text{Conv}_{1 \times 1}(\text{AvgPool}(F_{\text{fuse}}))))), \quad (11)$$

其中: $w_c \in \mathbb{R}^{B \times C \times 1 \times 1}$ 表示通道注意映射, σ 为sigmoid函数。

在SAM中,采用双路池化(平均池化和最大池化)分别沿通道维度压缩输入特征 F_{fuse} 。随后,通过 7×7 卷积融合池化信息,大感受野可确保局部与全局特征的协同建模,最终输出空间权重 w_s ,其数学表达式如下:

$$w_s = \text{Conv}_{7 \times 7}([\text{MaxPool}(F_{\text{fuse}}); \text{AvgPool}(F_{\text{fuse}})]), \quad (12)$$

其中: $w_s \in \mathbb{R}^{B \times 1 \times H \times W}$ 代表空间注意映射, $;$ 代表通道拼接。

本文定义了通道-空间权重整合方法,将上述两个模块生成的权重通过统计特性约束实现自适应特征选择,获得融合权重,其数学表达式

如下:

$$w_{\text{fuse}} = \begin{cases} \max(w_c, w_s), & \text{if } w_c \geq \mu_c \& w_s \geq \mu_s \\ \frac{w_c + w_s}{2}, & \text{otherwise} \\ \min(w_c, w_s), & \text{if } w_c \leq \mu_c \& w_s \leq \mu_s \end{cases}, \quad (13)$$

其中: w_{fuse} 表示融合权重, μ_c 和 μ_s 分别对应输入权重 w_c 和 w_s 的均值。当输入权重 w_c 和 w_s 均大于各自对应均值 μ_c 和 μ_s 时,取较大值,凸显跨纬度一致性重要特征;当输入权重 w_c 和 w_s 均小于其对应均值 μ_c 和 μ_s 时,取较小值,弱化双维度共同低响应区域;当输入权重与对应均值相比一大一小时,说明在空间和通道维度对该特征的重视程度不同,采用算术平均实现平滑过渡,以平衡不同维度的权重分配。这种设计能够自适应地调节特征融合,同时依赖简单且可计算的均值约束条件来实现高效融合。

在PAM中,将输入特征 F_{fuse} 与融合权重 w_{fuse} 进行通道维度拼接后,通过像素混洗(Pixel Shuffle, PS)提升特征交互效率,再经过 7×7 卷积层实现像素级特征优化,最终通过sigmoid函数归一化生成精细化掩码 w_p ,其数学过程如下:

$$w_p = \sigma(f^{7 \times 7}(\text{PS}([F_{\text{fuse}}; w_c + w_s]))), \quad (14)$$

其中: $w_p \in \mathbb{R}^{B \times 1 \times H \times W}$ 表示通道注意映射, $f^{7 \times 7}$ 为由多个 7×7 核组成的卷积层。

$$\text{DiceLoss} = 1 - \frac{2 \sum_{i=1}^N y_i \hat{y}_i + \epsilon}{\sum_{i=1}^N y_i + \sum_{i=1}^N \hat{y}_i + \epsilon}, \quad (17)$$

$$\text{Loss} = \lambda_1 \text{DiceLoss} + \lambda_2 \text{FocalLoss}, \quad (18)$$

其中: p_i 为模型预测的概率; α_i 是平衡因子, 用于调节正负样本权重; γ 是聚焦参数, 用于调节难以样本权重; y_i 是真实标签; $\hat{y}_i$ 是预测概率; ϵ 仅用于避免分母为零; N 是像素总数; λ_1 和 λ_2 为两种损失函数的权重因子。

3 数据集构建及评价指标

3.1 实验环境与参数配置

在模型训练过程中对参数进行设置, Learning rate 初始值设置为 0.001, batch-size 为 2, 最大 epoch 为 100。实验环境配置见表 1。

表 1 实验环境配置

Tab. 1 Test environment configuration

名称	参 数
CPU	12 vCPU Intel(R) Xeon(R) Platinum 8352V
GPU	vGPU-32GB(32GB) * 1
Python version	Python 3.10.14
操作系统	Ubuntu 22.04
深度学习框架	Pytorch 2.1.1

3.2 数据集构建

为验证本文方法的有效性和泛化能力, 分别在 Massachusetts buildings^[30] 和 deepglobe-road-dataset^[31] 数据集上进行遥感影像分割实验, 二者均为真实世界的高分辨率光学遥感影像。

Massachusetts Buildings 数据集为常用于建筑物分割的高分辨率遥感数据集, 覆盖约 341 平方公里的区域, 包含 151 张 1 500 × 1 500 像素的 RGB 航拍图像, 空间分辨率为 1 米/像素。图像经过滑动窗口裁剪为 500 × 500 像素, 并随机切割为 256 × 256 像素样本, 剔除无效标签后, 形成 994 张训练图像和 110 张验证图像。Deepglobe-Road Dataset 空间分辨率为 0.5 米/像素, 采用相同处理策略, 将 6 226 张 1 024 × 1 024 训练图像生成 22 416 张训练样本和 2 488 张验证样本, 如图 5 所示。

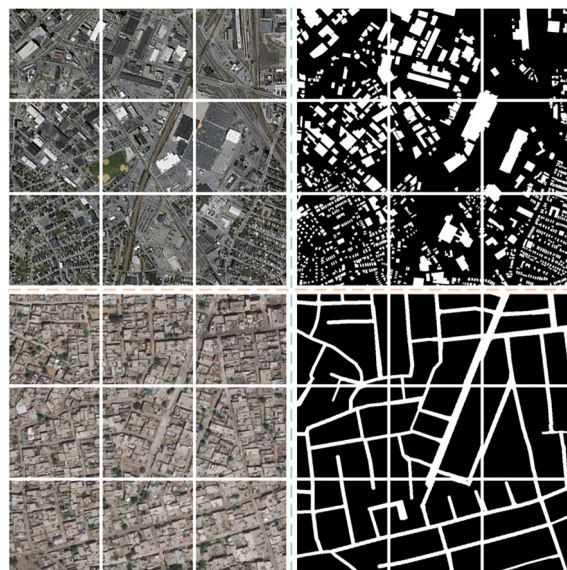


图 5 部分数据集展示

Fig. 5 Partial dataset display

3.3 评价指标

实验采用的评价指标有准确率(A), $F1$, $mIOU$ (Mean Intersection-over-union, $mIoU$)、预测精确度(P)和召回率(R)的计算公式如下:

$$A = \frac{TP + TN}{TP + FP + TN + FN}, \quad (19)$$

$$F1 = \frac{2TP}{2TP + FN + FP}, \quad (20)$$

$$mIOU = \frac{TP}{(TP + FP + FN)}, \quad (21)$$

$$P = \frac{TP}{(TP + FP)}, \quad (22)$$

$$R = \frac{TP}{(TP + FN)}, \quad (23)$$

其中: TP , FP , TN , FN 分别表示正确分类为正类的像素个数、错误分类为正类的像素个数、错误分类为负类的像素个数、错误分类为负类的像素个数, $F1$ 为 2 倍精确率与召回率之积与精确率与召回率之和的比值。

4 实验结果与分析

4.1 消融实验

为验证各模块对模型性能的影响, 基于已划分的 Massachusetts Buildings 数据集, 采用 OS-SM 作为特征提取模块、简单相加方式作为跳层连接的特征融合策略, 并采用线性插值作为上采

样方法,构建了基准模型(方法 1)。随后逐步引入 CFHAF, DySample 以及 ORSM 进行消融实验。实验结果见表 2,其中“√”表示该模块被加入,指标中加粗为该指标的最优模型,下划线为次优模型。可视化对比结果如图 6 所示,注意力权重如图 7 所示。

实验结果表明,各个模块均对模型性能提升具有独特贡献。方法 2 单独引入 CFHAF,使

F1score 提升 0.55%,mIoU 提升 1.88%,表明该模块在特征融合方面有效缓解假阳性问题;方法 3 单独使用 DySample,使 F1score 提升 1.25%,mIoU 提升 2.74%,验证了其在提升边缘细节恢复能力方面的有效性,但可能导致目标粘连现象;方法 4 将 OSSM 替换为 ORSM,带来了 1.29% 的 F1score 和 2.32% 的 mIoU 提升,体现了其在处理中等尺寸目标方面的优势。

表 2 消融实验结果对比
Tab. 2 Comparison of ablation results

Method	CFHAF	DySample	ORSM	F1 score/%	mIoU/%	P/%	R/%	Accuracy/%
1				82.79	73.35	77.24	89.19	95.19
2	√			83.34	75.23	<u>81.41</u>	85.36	95.58
3		√		84.04	76.09	81.09	87.20	95.71
4			√	84.08	75.67	80.25	88.30	95.67
5	√	√		84.30	<u>76.38</u>	80.72	88.20	95.74
6	√		√	<u>84.38</u>	76.26	82.58	86.27	<u>95.84</u>
7	√	√	√	84.64	77.07	81.15	<u>88.43</u>	95.86

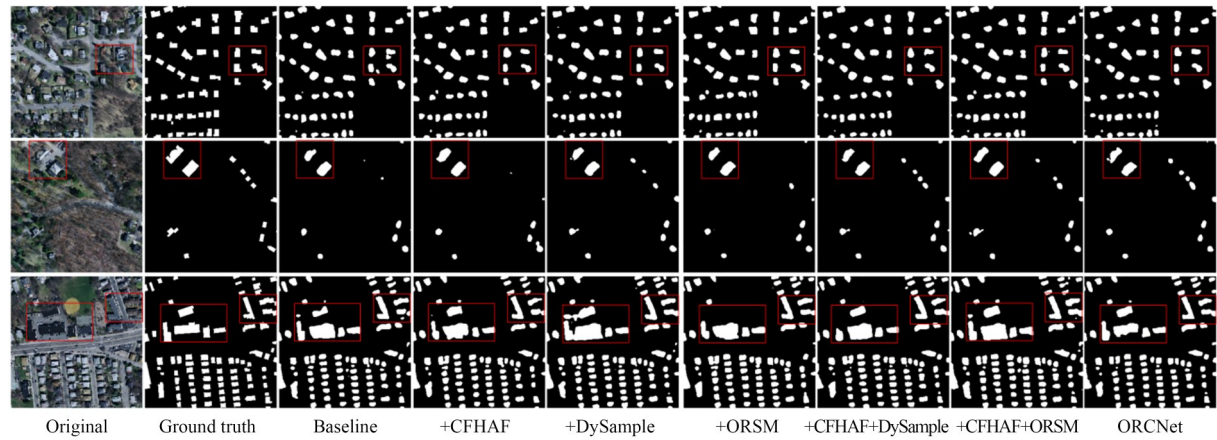


图 6 消融实验结果对比
Fig. 6 Comparison of ablation results

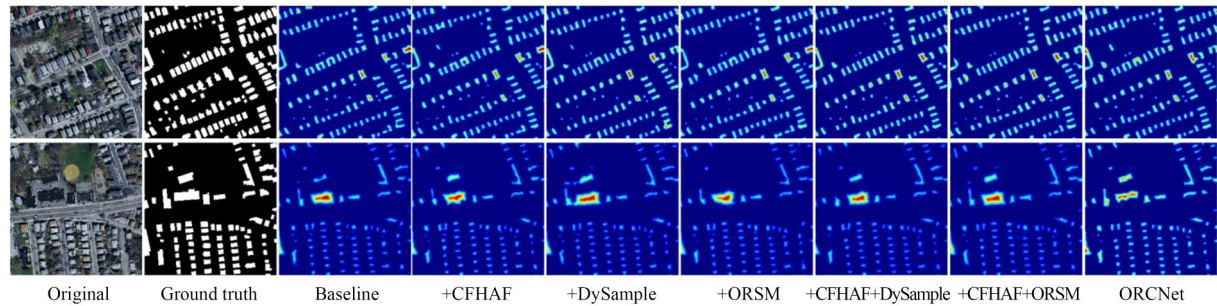


图 7 消融实验注意力权重可视化
Fig. 7 Visualization of attention weight for ablation results

在方法 2 的基础上,方法 5 联合使用 CFHAF 和 DySample,相较单独使用,提升了 1.51% 的 F1score (84.30%) 以及 3.03% 的 mIoU (76.38%),能够在缓解目标黏连的同时实现对中型目标的更精准定位;方法 6 联合 CFHAF 与 ORSM,以牺牲部分召回率为代价,使 F1score 提升 1.59%,mIoU 提升 2.91%,展示了全局-局部特征关联对空间表达精度的提升作用;最终完整模型(方法 7)整合所有模块,获得 84.64% 的 F1score、77.07% 的 mIoU、81.15% 的精确度、88.43% 的召回率和 95.86% 的准确率,三项核心指标均达到最优,比基准模型分别提升了 1.85%、3.72% 和 0.67%。从图 7 注意力权重的可视化分析可以看出,ORCNet 在注意力分布上更具平衡性,能更准确地识别中小目标边缘,避免了大尺度目标权重过高带来的局部细节丢失问题,并在细长条状建筑物识别方面表现出更强的结构保持能力,具有更高效的特征表征与细节重建能力。

4.2 损失函数

针对本文提出的 ORCNet 模型所使用的损失函数的权重因子进行分析实验,结果见表 3。为尽可能使 Focal Loss 与 Dice Loss 为同一数量级,因此取值均取 0.5,1,1.5。当 $\lambda_1=1$ 且 $\lambda_2=1$ 时,模型性能达到最优,F1score 和 mIoU 分别为 84.64% 和 77.07%。实验表明,增加 Dice Loss 权重通常能提升模型性能,而增加 Focal Loss 权重则对性能产生负面影响。因此,在 ORCNet 模型中采用 $\lambda_1=1$ 且 $\lambda_2=1$ 的参数设置,以获得最佳的模型性能。

表 3 权重因子分析实验

Tab. 3 Weight factor analysis experiment

Methods	λ_1	λ_2	F1/%	mIoU/%
1	0.5	0.5	82.51	71.90
2	0.5	1	80.13	65.15
3	0.5	1.5	79.49	66.39
4	1	0.5	83.04	73.78
5	1	1	84.64	77.07
6	1	1.5	82.51	72.51
7	1.5	0.5	83.93	74.99
8	1.5	1	81.48	69.69
9	1.5	1.5	79.90	67.89

4.3 不同模型对比实验

本研究提出的 ORCNet 模型在 Massachusetts buildings 数据集上进行系统验证,通过与传统卷积架构下 FCN^[32],DANet^[33],DeeplabV3+,Fastscnn^[34],Emanet^[35]等模型和 Transformer 架构下 Swin-UPerNet^[36],TransUNet^[37],Segformer^[38]以及 Mamba 框架下 RSMamba 共 9 种主流影像分割网络模型进行对比,对比结果见表 4。本文模型与其他对比模型的分割结果如图 8 所示。

表 4 Massachusetts buildings 数据集分割指标对比

Tab. 4 Comparison of segmentation metrics on Massachusetts buildings dataset (%)

方 法	F1	mIoU	P	R	Accuracy
FCN	76.55	66.02	82.81	72.73	92.27
DANet	77.37	66.90	83.70	73.48	92.52
Deeplab v3+	81.76	71.79	85.29	79.05	93.54
Emanet	81.93	71.96	84.37	79.92	93.44
Swin-UPerNet	82.24	72.36	85.61	79.60	93.68
RSMamba	82.41	73.15	79.63	86.33	95.37
Pspnet	82.86	73.17	88.34	79.06	94.14
TransUNet	83.11	73.45	87.03	80.14	94.05
Fastscnn	84.12	74.65	85.50	82.89	94.07
Segformer	84.14	74.73	87.92	81.22	94.38
ORCNet	84.64	77.07	81.15	88.43	95.84

由表 4 可知,ORCNet 在 F1score, mIoU 和准确率三项指标上分别达到 84.64%, 77.07% 和 95.84%, 均为最优表现。相比传统模型 FCN, F1 score 和准确率分别提升 8.09% 和 3.57%, mIoU 提升 11.05%; 相较 DANet 和 DeeplabV3+ 等场景分割模型, F1score 和 mIoU 均有显著提升; 与 RSMamba, TransUNet 和 Segformer 等遥感分割模型相比, ORCNet 的 F1score 分别提升 2.23%, 1.53% 和 0.5%, mIoU 提升 3.92%, 3.62% 和 2.34%。整体而言, ORCNet 在均衡精度与交并比表现上均优于其他方法, 充分验证了所提方法的有效性和优越性。

图 8 可视化结果分析表明, 不同框架模型呈现明显性能差异。传统分割模型 FCN 等传统卷积模型在小目标识别中易出现大面积误判, 而 Transformer 系列模型 (Swin-UPerNet, TransU-

Net 和 Segformer)均在场景一大尺寸目标目标边缘分割中明显存在不同程度的缺失。其中, TransUNet 凭借多尺度特征提取能力展现出相对较好的边缘识别效果,但在场景二、三等细节区域仍存在边缘模糊问题。RSMamba 框架虽然

获得较高召回率,但预测精确度却降至 79.63%。相比之下,ORCNet 通过校准建模与特征融合机制,全面提升预测精度,有效解决了传统模型的小目标混淆问题,验证了所提方法在高分辨率遥感影像分割中的实用价值。

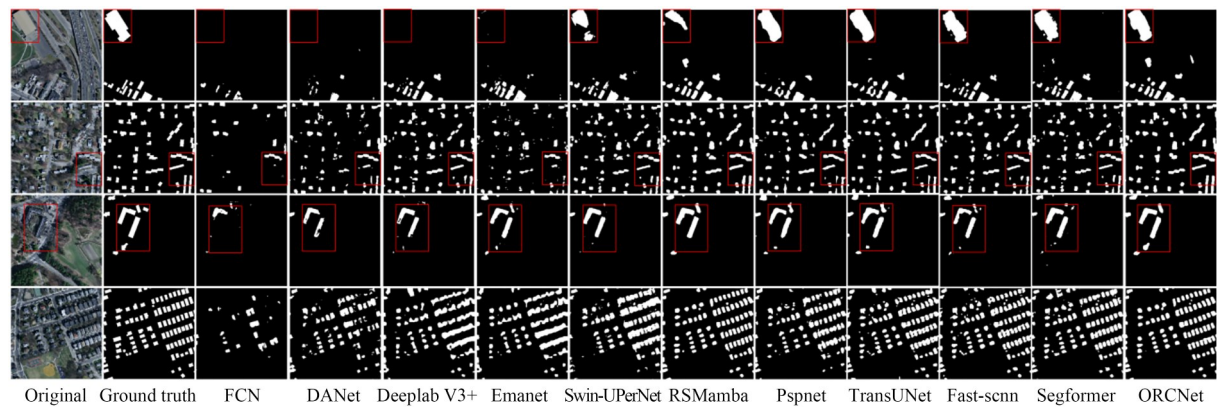


图 8 Massachusetts buildings 数据集分割结果对比

Fig. 8 Comparison of segmentation results on massachusetts buildings dataset

4.4 泛化能力测试

为验证本文提出的 ORCNet 模型的泛化能力,采用与建筑物识别的相同配置在 deepglobe-road-dataset 数据集上进行泛化实验,并对不同模型的分割精度进行了对比分析,实验结果如表 5 所示。

如图 9 所示,在 Deepglobe-road-dataset 数据集上,ORCNet 在复合场景中保持稳定识别,在

海域港口与民居混叠场景中完整保留道路拓扑结构,在农田细长小路场景中优于传统卷积模型(如 FCN, Deeplab v3+),避免输出断裂;相较于 Transformer 模型,其避免了 patch 划分对连续性的破坏;相比 Mamba 框架模型,其在遮挡场景中更准确还原结构。通过多尺度特征建模与结构化约束机制,ORCNet 显著提升了道路连续性与复杂场景适应能力,体现了其跨域性能优势。

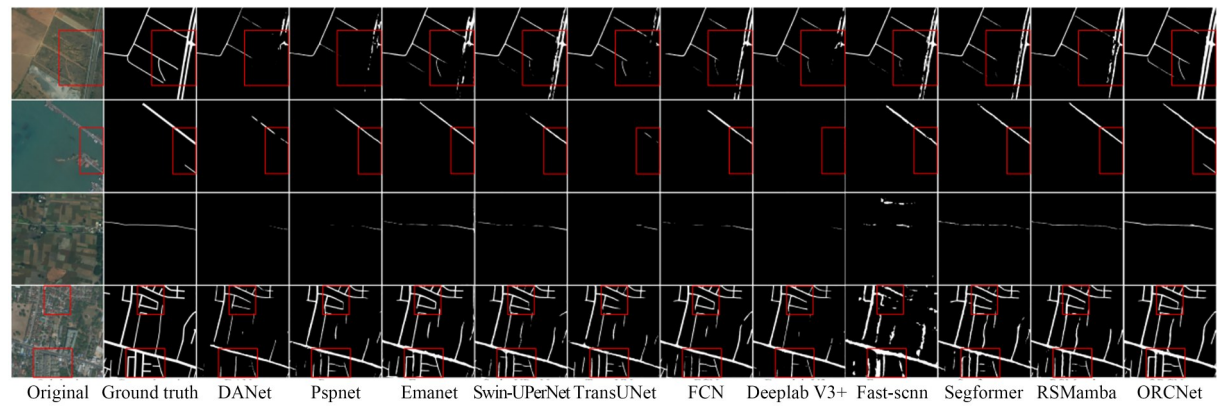


图 9 deepglobe-road-dataset 数据集分割结果对比

Fig. 9 Comparison of segmentation results on deepglobe-road-dataset

表 5 显示,ORCNet 在道路数据集上表现优于全部对照模型,F1score 达 85.32%,高于 RS-

Mamba 模型 3.51%,mIoU 为 69.73%,优于 Segformer 和 FastSCNN。其准确率为 98.02%,高于

建筑物数据集的 95.84%,主要因道路像素占比低,背景分类影响更大。DySample 优化了细长线性特征,ORSM 提升几何敏感性。尽管 mIoU 略有下降,但较传统模型 Deeplab v3+ 仍有显著优势。ORCNet 在保持高精确率(85.44%)的同时,召回率提升至 85.19%,优于 RSMamba 6.69%,体现良好的正负样本平衡能力。

表 5 Deepglobe-road-dataset数据集指标对比
Tab. 5 Comparison of metrics for deepglobe-road-dataset (%)

方法	F1	mIoU	P	R	Accuracy
DANet	60.22	54.33	90.32	56.43	96.01
Pspnet	60.56	54.55	90.95	56.64	96.03
Emanet	64.10	56.79	74.90	60.22	95.65
Swin-UPerNet	64.69	57.38	90.4	59.53	96.23
TransUNet	65.34	57.84	88.47	60.11	96.23
FCN	66.24	58.52	90.82	60.68	96.33
Deeplab v3+	66.34	58.59	90.25	60.78	96.32
Fastscnn	73.27	64.23	89.19	66.96	96.71
Segformer	77.41	68.06	85.47	72.55	96.84
RSMamba	81.81	65.61	85.43	78.5	97.64
ORCNet	85.32	69.73	85.44	85.19	98.02

为进一步验证本文提出的 ORCNet 模型的泛化能力,取 WHU-CD 数据集^[39]上的 2016 年图像作为 WHU 数据集再次进行泛化实验,并对不同模型的分割精度进行了性能对比分析,结果如表 6 所示。WHU 数据集的分割结果如图 10 所示。

表 6 展示了在 WHU 数据集上,多种图像分割模型在 F1score 和 mIoU 指标上的对比结果。本表中还列出了模型的参数量(Params)、计算复杂度(GFLOPs)以及推理速度(FPS),以全面评估不同方法在精度、效率和可扩展性方面的综合性能。结果显示,ORCNet 在 F1 分数和 mIoU 指标上显著优于其他模型,同时保持了较高的推理速度和较低的计算开销,体现出其在高分辨率遥感图像分割任务中的高效性和优越性。

图 10 展示了 ORCNet 在 WHU 数据集上的分割结果,直观地呈现了目标边界清晰度以及整体分割质量效果。上述跨域实验结果表明,ORCNet 不仅具备处理建筑物等大尺度目标的能

表 6 WHU 数据集上的性能分析
Tab. 6 Performance analysis on WHU dataset

方 法	F1/%	mIoU /%	Param s(M)	GFLO Ps	FPS
FCN	89.60	81.17	47.13	49.47	10.7
DANet	89.88	81.62	47.48	50.97	8.83
Swin-UPerNet	90.60	82.81	64.05	59.31	9.5
Pspnet	91.13	83.71	46.61	44.71	11.4
TransUNet	91.22	83.86	28.99	50.96	0.8
Fastscnn	91.69	84.66	1.40	0.23	143.6
Segformer	91.74	84.74	3.72	1.78	76
RSMamba	93.01	86.93	30.69	25.16	32.71
Emanet	93.29	87.42	39.99	42.39	4.9
Deeplab v3+	93.44	87.68	41.23	44.16	11.33
ORCNet	94.68	89.90	36.20	25.33	30.57

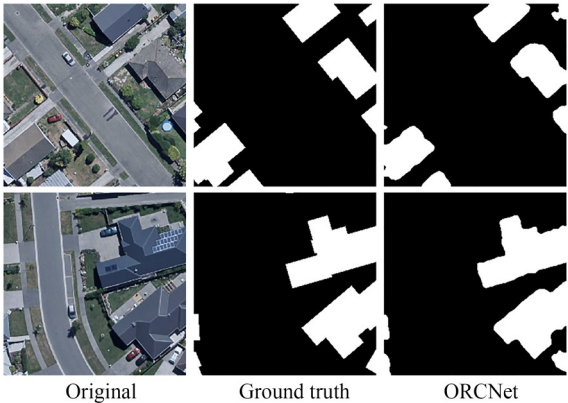


图 10 WHU 数据集上的分割结果
Fig. 10 Segmentation results on WHU dataset

力,在面对道路这类细长、连续性目标时,同样表现出优秀的特征捕捉和场景适应能力,验证了模型架构设计的泛用性价值。

5 结 论

针对高分遥感影像分割中特征提取不全、边缘细节丢失及多尺度目标分割效果不佳等问题,本文构建了一种全向矩形校准网络(ORCNet)。设计 ORSM 作为主干网络,解决了传统 OSSM 模块 patch 内部特征依赖关系建模问题,结合矩形自校准机制建立几何敏感权重,改善相似地物误分情况,提升了小尺度目标特征的保留率与表

征能力,高效提取主要特征信息;设计了互补滤波混合特征融合模块,该模块采用加权融合并结合残差结构,集成CAM,SAM和PAM模块以捕获更多空间信息,在保留原特征的同时增加细节补充,提升关键信息敏感度并融合多尺度特征,提高了网络的整体性能;集成动态点上采样模块进行图像恢复与细节增强,强化像素级感知能力,提升边缘恢复精度,从而进一步增强网络的特征表达能力。在Massachusetts buildings数据集上的实验表明,ORCNet在F1score,mIoU和Accuracy分别达到84.64%,77.07%和95.84%,均优于TransUNet,Segformer,RS-Mamba等对比模型。在deepglobe-road-dataset数据集上的测试进一步验证了其泛化能力,

F1score, mIoU 和 Accuracy 依旧可以保持在85.32%,69.73%和98.02%。证明了ORCNet在像素级分割精度、边缘细节保留及多尺度目标识别方面的优越性。

未来,将探索更多新型网络结构和优化策略,进一步优化高分遥感影像分割网络,在尽可能不牺牲Precision的情况下提升模型召回率,减少模型误报概率。

作者贡献声明:

陈思睿:论文构思、撰写及可视化呈现;

胡方敏:论文审核与提供资源;

王洪亮:论文审核与获取资助;

谢涛:论文审核与编辑写作。

参考文献:

- [1] ZHANG G Z, JIANG W Y. Remote sensing image semantic segmentation method based on a deep convolutional neural network and multiscale feature fusion [J]. *International Journal on Semantic Web and Information Systems*, 2023, 19(1): 1-16.
- [2] CHENG J, DENG C J, SU Y Z, *et al.* Methods and datasets on semantic segmentation for Unmanned Aerial Vehicle remote sensing images: a review [J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2024, 211: 1-34.
- [3] ZHU Q W, LI K, ZHANG G J, *et al.* GDSR: Global-Detail Integration through Dual-Branch Network with Wavelet Losses for Remote Sensing Image Super-Resolution[J/OL]. *Arxiv*, preprint Arxiv: 2501.01460, 2024. <https://arxiv.org/html/2501.01460>.
- [4] 朱长青,王耀革,马秋禾,等. 基于形态分割的高分辨率遥感影像道路提取[J]. *测绘学报*, 2004, 33(4): 347-351.
ZHU CH Q, WANG Y G, MA Q H, *et al.* Road extraction from high-resolution remotely sensed image based on morphological segmentation[J]. *Acta Geodaetica et Cartographica Sinica*, 2004, 33(4): 347-351. (in Chinese)
- [5] 汪闯,骆剑承,周成虎,等. 结合高斯马尔可夫随机场纹理模型与支撑向量机在高分辨率遥感图像上提取道路网[J]. *遥感学报*, 2005, 9(3): 271-276.
- WANG M, LUO J CH, ZHOU CH H, *et al.* Extraction of road network from high resolution remote sensed imagery with the combination of Gaussian Markov random field texture model and support vector machine[J]. *Journal of Remote Sensing*, 2005, 9(3): 271-276. (in Chinese)
- [6] SHI J B, MALIK J. Normalized cuts and image segmentation [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, 22(8): 888-905.
- [7] KASS M, WITKIN A, TERZOPOULOS D. Snakes: active contour models [J]. *International Journal of Computer Vision*, 1988, 1(4): 321-331.
- [8] CHEN L C, ZHU Y K, PAPANDREOU G, *et al.* Encoder-decoder with atrous separable convolution for semantic image segmentation[C]. *Computer Vision-ECCV 2018*. Cham: Springer, 2018: 833-851.
- [9] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional Networks for Biomedical Image Segmentation [M]. *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015*. Cham: Springer International Publishing, 2015: 234-241.
- [10] CAO H, WANG Y Y, CHEN J, *et al.* Swin-unet: Unet-like Pure Transformer for Medical Image Segmentation [M]. *Computer Vision-ECCV 2022 Workshops*. Cham: Springer Nature Switzerland, 2023: 205-218.
- [11] LIU Y, TIAN Y, ZHAO Y, *et al.* Vmamba: Vi-

- sual state space model [J/OL]. *Arxiv*, preprint Arxiv: 2401.10166, 2024. <https://arxiv.org/abs/2401.10166>.
- [12] 李智杰, 惠爱婷, 李昌华, 等. 面向遥感图像道路提取的多尺度上下文感知网络[J]. *光学精密工程*, 2025, 33(4): 610-623.
LI ZH J, HUI A T, LI CH H, *et al.* Multi-scale context-aware network for road extraction in remote sensing images [J]. *Opt. Precision Eng.*, 2025, 33(4): 610-623. (in Chinese)
- [13] 刘帅, 李笑迎, 于梦, 等. 高分辨率遥感图像双解耦语义分割网络模型[J]. *测绘学报*, 2023, 52(4): 638-647.
LIU SH, LI X Y, YU M, *et al.* Dual decoupling semantic segmentation model for high-resolution remote sensing images[J]. *Acta Geodaetica et Cartographica Sinica*, 2023, 52(4): 638-647. (in Chinese)
- [14] ZHANG W H, JIAO L C, LIU X, *et al.* Multi-scale feature fusion network for object detection in VHR optical remote sensing images[C]. *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium. July 28-August 2, 2019, Yokohama, Japan*. IEEE, 2019: 330-333.
- [15] CHEN Z X, HE Z W, LU Z M. DEA-net: single image dehazing based on detail-enhanced convolution and content-guided attention[J]. *IEEE Transactions on Image Processing*, 2024, 33: 1002-1015.
- [16] 江宝得, 黄威, 许少芬, 等. 融合分散自适应注意力机制的多尺度遥感影像建筑物实例细化提取[J]. *测绘学报*, 2023, 52(9): 1504-1514.
JIANG B D, HUANG W, XU SH F, *et al.* Multi-scale building instance refinement extraction from remote sensing images by fusing with decentralized adaptive attention mechanism [J]. *Acta Geodaetica et Cartographica Sinica*, 2023, 52(9): 1504-1514. (in Chinese)
- [17] ZHU H, ZHU Y, XIAO J Y, *et al.* Exact: exploring space-time perceptive clues for weakly supervised satellite image time series semantic segmentation [C]. 2025 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 10-17, 2025, Nashville, TN, USA. IEEE, 2025: 14036-14045.
- [18] LI K Y, LIU R X, CAO X Y, *et al.* SegEarth-OV: towards training-free open-vocabulary segmentation for remote sensing images [C]. 2025 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 10-17, 2025, Nashville, TN, USA. IEEE, 2025: 10545-10556.
- [19] ZHAO S, CHEN H, ZHANG X, *et al.* Rs-mamba for large remote sensing image dense prediction [J/OL]. *Arxiv*, preprint Arxiv: 2404.02668, 2024. <https://arxiv.org/abs/2404.02668>.
- [20] NI Z, CHEN X, ZHAI Y, *et al.* Context-Guided Spatial Feature Reconstruction for Efficient Semantic Segmentation [J/OL]. *Arxiv*, preprint Arxiv: 2405.06228, 2024. <https://arxiv.org/abs/2405.06228>.
- [21] CHEN K Y, CHEN B W, LIU C Y, *et al.* RS-Mamba: remote sensing image classification with state space model [J]. *IEEE Geoscience and Remote Sensing Letters*, 2024, 21: 8002605.
- [22] DING Z, LI Y, HE Y, *et al.* Dygmamba: Efficiently modeling long-term temporal dependency on continuous-time dynamic graphs with state space models [J/OL]. *Arxiv*, preprint Arxiv: 2408.04713, 2024. <https://arxiv.org/abs/2408.04713>.
- [23] LIN T Y, DOLLÁR P, GIRSHICK R, *et al.* Feature pyramid networks for object detection[C]. 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. July 21-26, 2017, Honolulu, HI, USA. IEEE, 2017: 936-944.
- [24] TARG S, ALMEIDA D, LYMAN K. Resnet in resnet: Generalizing residual architectures[J/OL]. *Arxiv*, preprint Arxiv: 1603.08029, 2016. <https://arxiv.org/abs/1603.08029>.
- [25] NHI N T U, LE T M, VAN T T. A model of semantic-based image retrieval using C-tree and neighbor graph [J]. *International Journal on Semantic Web and Information Systems*, 2022, 18(1): 1-23.
- [26] LIU W Z, LU H, FU H T, *et al.* Learning to upsample by learning to sample[C]. 2023 *IEEE/CVF International Conference on Computer Vision (ICCV)*. October 1-6, 2023, Paris, France. IEEE, 2023: 6004-6014.
- [27] 胡功明, 杨春成, 徐立, 等. 改进U-Net的遥感图像语义分割方法[J]. *测绘学报*, 2023, 52(6): 980-989.
HU G M, YANG CH CH, XU L, *et al.* Improved U-Net remote sensing image semantic segmentation

- method[J]. *Acta Geodaetica et Cartographica Sinica*, 2023, 52(6): 980-989. (in Chinese)
- [28] LIN T Y, GOYAL P, GIRSHICK R, *et al.* Focal loss for dense object detection[C]. 2017 *IEEE International Conference on Computer Vision (ICCV)*. October 22-29, 2017, Venice, Italy. IEEE, 2017: 2999-3007.
- [29] LI X, SUN X, MENG Y, *et al.* Dice loss for data-imbalanced NLP tasks[J/OL]. *Arxiv*, preprint Arxiv:1911.02855, 2019. <https://arxiv.org/abs/1911.02855>.
- [30] MNIH V. *Machine Learning for Aerial image Labeling*[M]. University of Toronto, Canada, 2013.
- [31] DEMIR I, KOPERSKI K, LINDENBAUM D, *et al.* DeepGlobe 2018: a challenge to parse the Earth through satellite images[C]. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. June 18-22, 2018, Salt Lake City, UT, USA. IEEE, 2018: 172-17209.
- [32] LU Y, CHEN Y R, ZHAO D B, *et al.* Graph-FCN for Image Semantic Segmentation[M]. *Advances in Neural Networks-ISNN 2019*. Cham: Springer International Publishing, 2019: 97-105.
- [33] 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)*. June 15-20, 2019, Long Beach, CA, USA. IEEE, 2019: 3141-3149.
- [34] POUDEL R P K, LIWICKI S, CIPOLLA R. Fast-scnn: Fast semantic segmentation network[J/OL]. *Arxiv*, preprint Arxiv:1902.04502, 2019. <https://arxiv.org/abs/1902.04502>.
- [35] WANG P, HU Y S, PENG S Y, *et al.* EMA-Net: an ancient text detection method based on enhanced-EfficientNet and multidimensional scale fusion[J]. *IEEE Internet of Things Journal*, 2024, 11(19): 32105-32116.
- [36] BENAYAS A, HASHEMPOUR R, RUMBLE D, *et al.* Unified transformer multi-task learning for intent classification with entity recognition[J]. *IEEE Access*, 2021, 9: 147306-147314.
- [37] CHEN J, LU Y, YU Q, *et al.* Transunet: Transformers make strong encoders for medical image segmentation [J/OL]. *Arxiv*, preprint Arxiv: 2102.04306, 2021. <https://arxiv.org/abs/2102.04306>.
- [38] XIE E Z, WANG W H, YU Z D, *et al.* SegFormer: Simple and efficient design for semantic segmentation with transformers [J/OL]. *Arxiv*, preprint Arxiv: 2105.15203, 2021. <https://arxiv.org/abs/2105.15203>.
- [39] WANG D P, DONG X H, HUANG L Y, *et al.* Information entropy uncertainty estimation based domain adaptation for land cover classification from multi-source remote sensing images[J]. *Geomatics and Information Science of Wuhan University*, 2024, 49(10):1940-1952.

作者简介:



陈思睿(1999—),男,黑龙江大庆人,硕士研究生,2022年于昆明理工大学获得学士学位,主要从事遥感影像及机器视觉方面的研究。E-mail: 1113251813@qq.com

通讯作者:



谢涛(1974—),男,云南昭通人,副教授,硕士生导师,2000年于云南大学获得硕士学位,主要研究方向包括图像处理,人工智能技术和新型传感器研制。E-mail: 77183118@qq.com