



· 深度学习视觉处理与分析 ·

结合注意力与自适应特征融合的遥感图像道路提取

李云红,侯乐乐,陈伟重,苏雪平,张刚,刘欢,李昀媿

西安工程大学 电子信息学院,陕西 西安 710048

摘要 针对遥感图像道路提取任务中道路特征提取不充分、细节信息丢失以及类别不平衡等问题,提出了一种结合注意力与自适应特征融合(attention and adaptive feature fusion unet, AAF-Unet)的遥感图像道路提取模型。该模型以 VGG16 为骨干网络,首先,在 Unet 网络跳跃连接处加入跨空间多尺度注意力(cross-spatial efficient multi-scale attention, CE-MA)模块,有效融合全局上下文信息,提升了对复杂背景和遮挡区域的特征感知能力。其次,在编码器和解码器桥接处加入自适应卷积融合(adaptive convolution mix, ACmix)模块,通过动态卷积与自注意力机制的结合,实现了局部细节特征与全局上下文信息的高效集成,显著提升了模型在边缘和细节恢复方面的性能。结合 Dice 和 Focal 设计复合损失函数,优化了模型在小目标和边缘模糊区域分割任务中的表现。最后,在不同数据集上进行实验验证,AAF-Unet 模型的 mIoU 指标在 DeepGlobe 数据集上达到了 82.01%,在 CHN6-CUG 数据集上达到了 79.36%,在 2 个数据集上相较于 Unet、DeepLabv3+、Hrnet 模型均有所提升,在保证模型泛化能力的同时取得了较好的分割精度。

关键词 图像处理;遥感图像;道路提取;特征融合;注意力机制

中图分类号: TP751 **DOI:** 10.16152/j.cnki.xdxzbzr.2026-03-009

Road extraction from remote sensing images based on attention and adaptive feature fusion

LI Yunhong, HOU Lele, CHEN Weizhong, SU Xueping,
ZHANG Gang, LIU Huan, LI Yunti

School of Electronics and Information, Xi'an Polytechnic University, Xi'an 710048, China

Abstract To tackle the challenges of inadequate feature representation, detail degradation, and imbalanced class distribution in road extraction from remote sensing imagery, this research introduces a road extraction model from remote sensing images based on attention and adaptive feature fusion Unet (AAF-Unet). The model uses VGG16 as the backbone network. Firstly, the cross-spatial efficient multi-scale attention (CEMA) module is added at the skip connection of the Unet network, effectively integrating global contextual information and enhancing the model's feature perception ability for complex backgrounds and occluded areas. Secondly, add the adaptive convolution mix (ACmix) module at the bridge between the encoder and the decoder.

收稿日期:2025-12-28 修回日期:2026-01-10

基金项目:国家自然科学基金青年基金(62403368);陕西省自然科学基金基础研究计划重点项目(2022JZ-35);西安市科学技术局“科学家+工程师”队伍建设项目(25KGYB00029)。

第一作者:李云红,女,博士,教授,博士生导师,从事图像处理、人工智能等研究,hitliyunhong@163.com。

By combining dynamic convolution with self-attention mechanisms, this module enables efficient integration of local detail features and global context information, significantly improving the model's performance in edge and detail recovery. In the loss function section a composite loss function combining Dice loss and Focal loss is proposed, optimizing the model's performance in segmentation tasks involving small targets and edge-blurred regions. Lastly, through the experimental verification on different dataset, the mIoU index of AAF-Unet model reached 82.01% on DeepGlobe dataset and 79.36% on the CHN6-CUG dataset. Compared with Unet, DeepLabv3+ and Hrnet models, the extraction results were increased respectively on two dataset. Good segmentation accuracy was achieved while ensuring the generalization ability of the model.

Keywords image processing; remote sensing image; road extraction; feature fusion; attention mechanism

遥感图像道路提取是遥感领域中的重要研究课题,在实际应用中发挥着至关重要的作用^[1],例如城市规划^[2]、无人驾驶^[3]、交通管理^[4]和灾害监测^[5]等领域。但遥感图像中道路的背景复杂、障碍物遮挡等问题,给高精度道路提取带来了巨大的挑战。近年来,深度学习技术取得了显著发展,尤其是基于卷积神经网络(CNN)的语义分割方法逐渐成为遥感图像道路提取领域的主流研究方向^[6],现有研究大致可以归纳为以下 3 个类别。

1) 基于经典编码-解码结构的道路提取方法。该类方法以 FCN、Unet 等网络为代表,通过构建端到端的语义分割框架实现道路区域的像素级预测。Long 等提出的全卷积网络(FCN)首次将卷积网络应用于语义分割任务,实现了高效的端到端分割^[7];随后,Zhong 等将 FCN 应用于道路与建筑物提取,有效提高了道路提取精度^[8];Ronneberger 等提出的 U-Net 通过编码器与解码器之间的跳跃连接增强了特征恢复能力,在遥感道路提取任务中取得了较好的效果^[9];佟喜峰等进一步对 Unet 进行改进,提升了复杂背景下的道路提取能力^[10]。然而,这类方法在处理细长道路、小目标以及复杂场景时,仍容易出现细节丢失、边界模糊和道路断裂等问题。

2) 融合注意力机制与多尺度特征增强的道路提取方法。针对传统卷积网络特征表达能力不足的问题,通过引入注意力机制、残差结构以及多尺度特征融合模块,增强网络对关键道路特征的感知能力。何哲等在 U-Net 基础上结合注意力残差结构与空间金字塔池化模块,提高了复杂场景下的道路特征提取能力^[11];齐然等提出 RAED-Net,在编码阶段引入改进的通道注意残差模块,以增强重要通道特征并抑制背景干扰^[12];郝明等提出 ResT-UNet,通过融合 ResNeSt 与多尺度特征

信息,进一步提升了道路边缘连续性与小目标分割精度^[13];曲海成等则利用双分支融合模块实现全局语义信息与局部细节特征的协同建模^[14]。这类方法能够有效改善道路连续性和复杂背景下的提取效果,但普遍存在模型结构复杂、计算量较大以及复杂场景下易过拟合等问题。

3) 针对复杂场景与特殊问题的改进型道路提取方法。该类研究主要聚焦于遮挡干扰、多光谱影像以及变化边界等特定问题的处理。唐伟等基于 Snake 模型理论提出高分辨率多光谱影像道路提取方法,通过四元数表达与分水岭分割实现精准道路轮廓提取^[15];Zhang 等提出 B2CNet,引入变化边界感知模块与深度特征提取模块,以增强变化区域边界信息表达并减少信息损失^[16];徐虹等则针对植被遮挡问题,提出双域特征均衡方法,通过联合遮挡移除与道路提取子网络提升复杂遮挡条件下的道路提取能力^[17]。这类方法在特定场景下具有较强针对性,但通常存在计算复杂度较高、泛化能力有限以及对数据质量依赖较强等问题。

针对上述问题,本文采用注意力机制与自适应特征融合的方法,构建了一种新型遥感图像道路提取模型 AAF-Unet(attention and adaptive feature fusion Unet)。在 DeepGlobe 和 CHN6-CUG 数据集上,通过与现有主流模型进行对比实验,验证了该模型在道路提取任务中的有效性。该方法能够有效整合局部细节特征与全局上下文信息,显著提升了道路边缘与细节特征的恢复能力。

1 模型概述

AFF-UNet 网络主要由编码器和解码器 2 部分构成,整体结构如图 1 所示。编码器部分基于 VGG16 架构,包含 10 个卷积层和 4 个最大池化,

用于逐步提取图像的浅层特征。每一层末端通过池化操作将特征图尺寸减半,在降低计算量的同时扩大感受野,从而增强模型对全局上下文信息的感知能力。在跳跃连接处,引入了跨空间多尺度注意力模块(CEMA),该模块借助并行子网络提取多尺度特征,并结合跨空间学习方法,增强模型对多尺度道路特征的感知能力,从而进一步提升分割

性能。在编码器与解码器之间的桥接部分,使用ACmix模块替换原始卷积块,ACmix模块融合了动态卷积与自注意力机制,既保留了卷积操作提取局部特征的能力,又引入了自注意力机制的全局上下文建模功能,显著改善了模型在边缘细节恢复方面的表现。最终,经解码器处理后的特征图通过一个 1×1 卷积层进行像素级分类,输出每个像素的类别概率,完成道路提取任务。

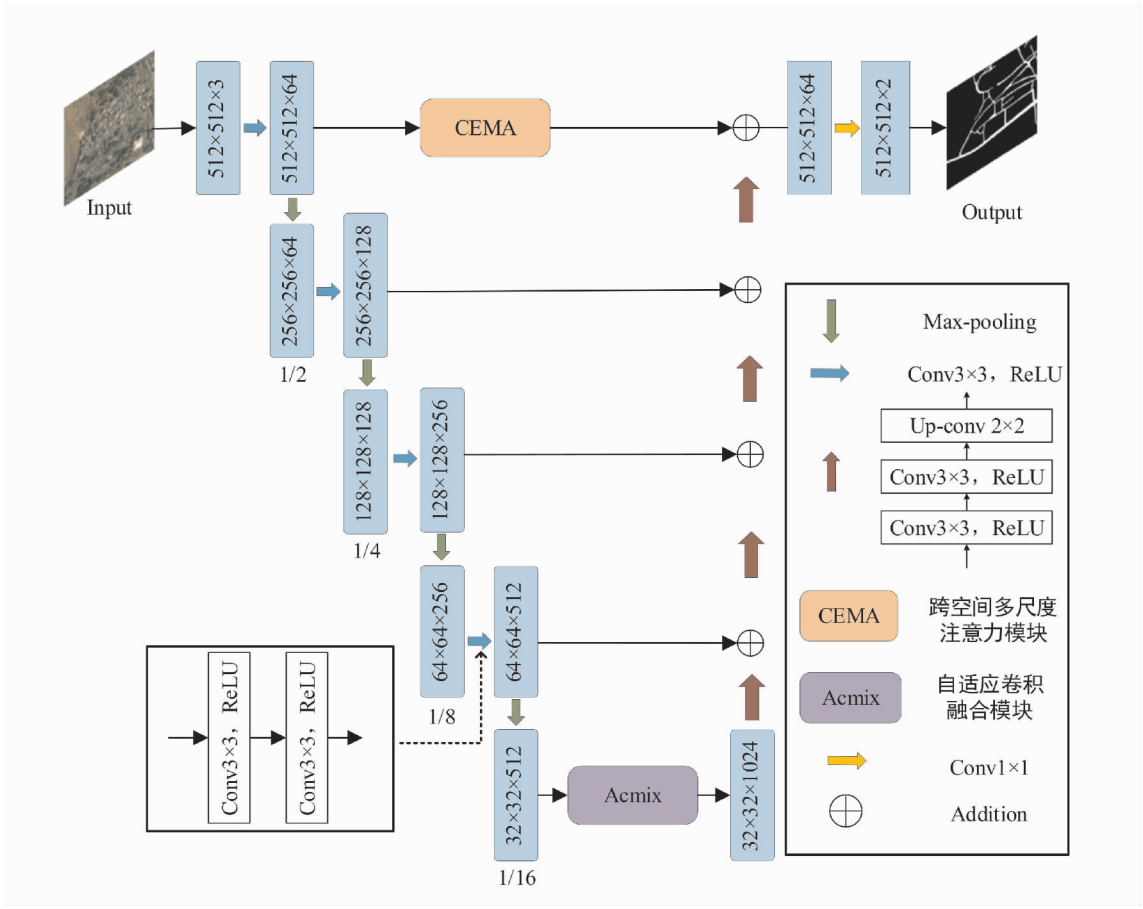


图1 AFF-Unet 网络结构

Fig.1 AFF-UNET network structure

1.1 跨空间多尺度注意力模块

传统注意力机制在有效捕捉多尺度特征方面存在局限,且易受复杂背景干扰而导致特征混淆。为此,本文提出跨空间多尺度注意力模块CEMA,通过并行子网络提取不同尺度的特征,并结合跨空间学习方法,增强模型对多尺度道路特征的感知能力,同时有效融合全局上下文信息,显著提升了道路提取的精度,尤其在复杂背景和遮挡区域的识别上表现更加出色。

跨空间多尺度注意力模块结构图如图2所示。对于任意给定的输入特征图 $\mathbf{x} \in \mathbf{R}^{C \times H \times W}$,先将输入特征图 $\mathbf{x}$ 分成 G 个子特征图 $\mathbf{x} = [\mathbf{x}_0, \mathbf{x}_1, \dots,$

$\mathbf{x}_{G-1}]$, $\mathbf{x}_i \in \mathbf{R}^{C/G \times H \times W}$ 。其中: C 表示通道数; H 为高度; W 为宽度。然后将分组后的子特征图 $\mathbf{x}_i$ 经过水平全局池化和垂直全局池化提取多尺度信息,分别得到 $\mathbf{x}_h \in \mathbf{R}^{C/G \times W \times 1}$ 、 $\mathbf{x}_w \in \mathbf{R}^{C/G \times H \times 1}$,接下来将 $\mathbf{x}_h$ 和 $\mathbf{x}_w$ 在通道维度进行拼接,得到形状为 $(C/G, H+W, 1)$ 的特征张量。接着进行 3×3 卷积,进一步捕捉局部空间信息,再通过 1×1 卷积调整通道数。通过分割操作,将该特征图重新分为 $\mathbf{x}_h$ 和 $\mathbf{x}_w$,分别进行Sigmoid激活。之后再对分组特征图 $\mathbf{x}_i$ 和激活后的特征进行加权。接着,模块使用分组归一化(group normalization, GN)操作,对加权后的特征进行归一化,得到 $\mathbf{x}_1 \in \mathbf{R}^{C/G \times H \times W}$ 。

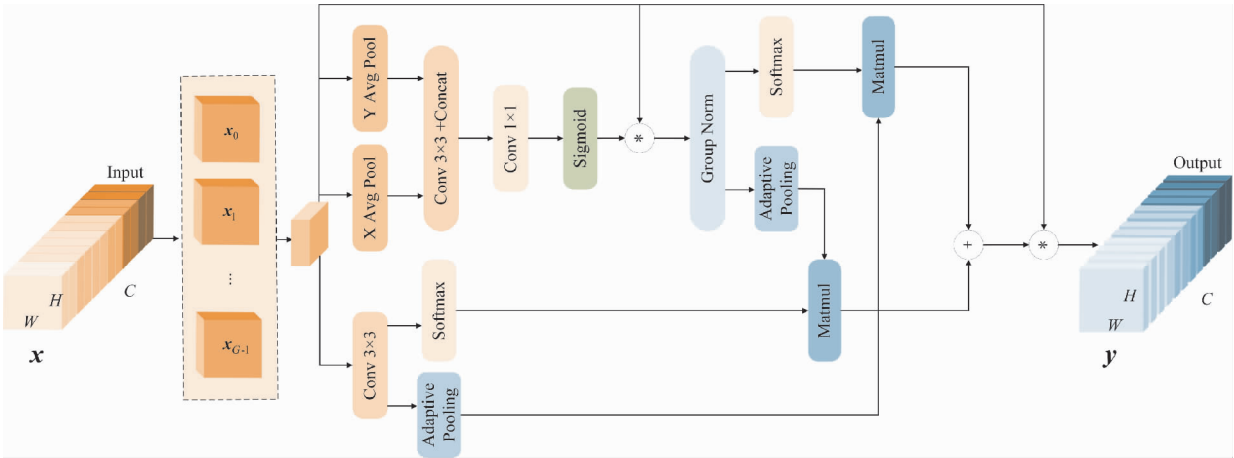


图 2 跨空间多尺度注意力模块

Fig. 2 Cross-spatial efficient multi-scale attention module

在另一条支路中使用 3×3 卷积操作对分组后的 $\mathbf{g}_x$ 进行处理,得到特征 $\mathbf{x}_2 \in \mathbf{R}^{C/G \times H \times W}$,然后对 $\mathbf{x}_1$ 和 $\mathbf{x}_2$ 进行自适应池化 (adaptive pooling, AP),得到全局特征 $\mathbf{z}_1 \in \mathbf{R}^{C/G \times 1 \times 1}$ 和 $\mathbf{z}_2 \in \mathbf{R}^{C/G \times 1 \times 1}$,进一步增强了全局特征的表达能力。这些全局特征随后经过 Softmax,得到归一化后的权重 $\mathbf{u}_1 \in \mathbf{R}^{1 \times C/G}$ 、 $\mathbf{u}_2 \in \mathbf{R}^{1 \times C/G}$ 。接下来计算加权系数,通过矩阵乘法 MatMul 得到最终的加权重 $\mathbf{U} \in \mathbf{R}^{G \times H \times W}$ 。最后,将计算出的加权系数 $\mathbf{U}$ 应用到原始的分组特征图 $\mathbf{g}_x$ 上,得到加权后的输出特征图 $\mathbf{y} \in \mathbf{R}^{G \times H \times W}$,上述过程表示如下:

$$\mathbf{U} = (\text{MatMul}(\mathbf{u}_1, \mathbf{x}_2) + \text{MatMul}(\mathbf{u}_2, \mathbf{x}_1)) \tag{1}$$

$$\mathbf{y} = (\mathbf{g}_x \cdot \sigma(\mathbf{U})) \tag{2}$$

通过这种加权机制,CEMA 模块能自适应调整特征图中每个位置的重要性,从而在处理复杂特征和细化局部信息方面提供更精细的控制。

1.2 自适应卷积融合模块

卷积和自注意力机制在特征提取中分别具备独特优势。卷积擅长捕捉局部细节特征,而自注意力机制则能够联系全局上下文信息。在自适应卷积融合模块 ACmix 中,输入特征先经过 1×1 卷积操作,再将得到的输出共享给 2 个模块同时使用,将卷积和自注意力机制高效结合,以提取出更加完整、连续的道路。自适应卷积融合模块如图 3 所示。

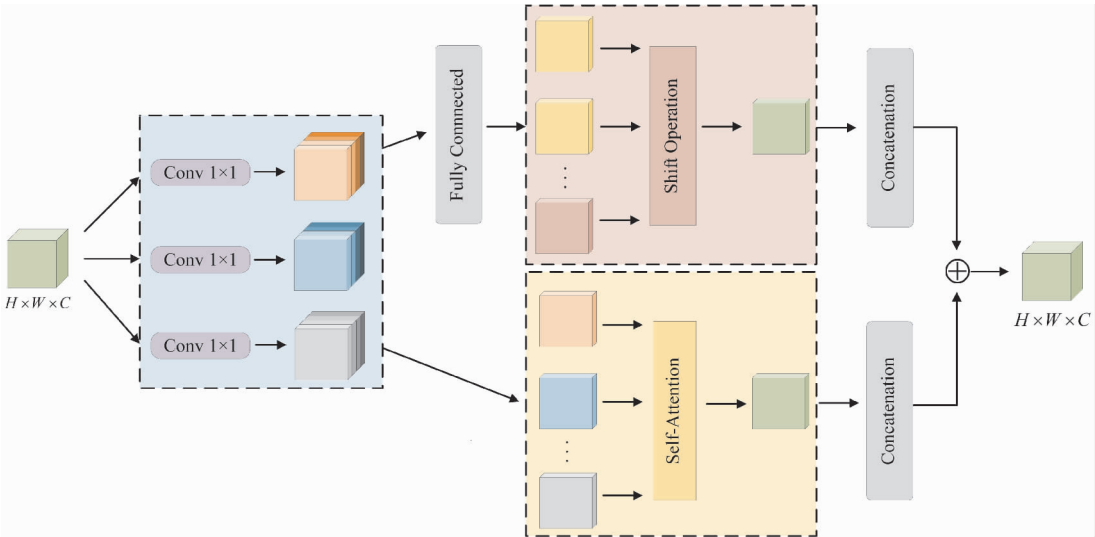


图 3 自适应卷积融合模块

Fig. 3 Adaptive convolution mix module

给定输入特征图 $\mathbf{F} \in \mathbf{R}^{C \times H \times W}$,首先通过 3 个 1×1 卷积生成 $\mathbf{Q}, \mathbf{K}, \mathbf{V}$ 。随后分别通过自注意力路

径和卷积路径对特征进行处理。在自注意力路径中,输入特征首先被重塑为多头形式。模块的头数

为 N , 每个头的特征维度为 $d = C/N$, 则 $\mathbf{Q}$ 、 $\mathbf{K}$ 、 $\mathbf{V}$ 被重塑为:

$$\mathbf{Q}_{\text{att}} = \mathbf{Q} \in \mathbf{R}^{B \times N \times d \times H \times W} \quad (3)$$

$$\mathbf{K}_{\text{att}} = \mathbf{K} \in \mathbf{R}^{B \times N \times d \times H \times W} \quad (4)$$

$$\mathbf{V}_{\text{att}} = \mathbf{V} \in \mathbf{R}^{B \times N \times d \times H \times W} \quad (5)$$

式中, B 为批次大小。

接着, 通过展开操作将 $\mathbf{K}$ 和 $\mathbf{V}$ 映射到局部区域:

$$\text{Unfold}(\mathbf{K}_{\text{att}}) \in \mathbf{R}^{B \times N \times d \times k^2 \times H' \times W'} \quad (6)$$

$$\text{Unfold}(\mathbf{V}_{\text{att}}) \in \mathbf{R}^{B \times N \times d \times k^2 \times H' \times W'} \quad (7)$$

式中: k 为自注意力核大小; H' 和 W' 为输出特征图的空间分辨率。

注意力权重通过 $\mathbf{Q}$ 和 $\mathbf{K}$ 的点积计算, 并经过 Softmax 归一化得到 $\mathbf{A}$, 最后, 通过注意力权重对值进行加权求和, 生成自注意力路径的输出。上述过程表示为:

$$\mathbf{A} = \text{Soft max} \left(\frac{\mathbf{Q}_{\text{att}} \cdot \mathbf{K}_{\text{att}}^T}{\sqrt{d}} \right) \in \mathbf{R}^{B \times N \times k^2 \times H' \times W'} \quad (8)$$

$$\mathbf{F}_{\text{att}} = \mathbf{A} \cdot \mathbf{V}_{\text{att}} \in \mathbf{R}^{B \times C \times H' \times W'} \quad (9)$$

在卷积路径中, 输入特征通过全连接层生成卷积核权重。最后, $\mathbf{F}_{\text{att}}$ 和 $\mathbf{F}_{\text{conv}}$ 通过 2 个可学习的权重 α 和 β 进行加权求和, 生成最终的输出特征。上述过程表示为:

$$\mathbf{F}_{\text{all}} = \text{FC}(\text{Concat}(\mathbf{Q}, \mathbf{K}, \mathbf{V})) \in \mathbf{R}^{B \times (k^2 \times N) \times H \times W} \quad (10)$$

$$\mathbf{F}_{\text{conv}} = \text{DepthwiseConv}(\mathbf{F}_{\text{all}}) \in \mathbf{R}^{B \times C \times H' \times W'} \quad (11)$$

$$\mathbf{F}_{\text{out}} = \alpha \cdot \mathbf{F}_{\text{att}} + \beta \cdot \mathbf{F}_{\text{conv}} \in \mathbf{R}^{B \times C \times H' \times W'} \quad (12)$$

式中: FC 表示全连接层; k 为卷积核大小。

ACmix 模块通过并行卷积和自注意力机制的方法, 既保留了卷积的局部特征提取能力, 又引入了自注意力的全局上下文建模能力, 显著提升了模型在图像识别和分割任务中的性能。

1.3 复合损失函数

为了解决样本数量不平衡和提高模型对复杂结构的敏感性, 提出了一种复合损失函数, 将 Dice 损失和 Focal 损失相结合。Dice 损失主要用于评估模型预测与真实标签之间的相似性, 特别适用于像素级分类任务。Focal 损失通过减少易分类样本的贡献, 增强难分类样本的影响, 从而缓解类别不平衡问题。复合损失函数计算公式如下,

$$L = \alpha \cdot L_{\text{Dice}} + (1 - \alpha) \cdot L_{\text{Focal}} \quad (13)$$

式中: 超参数 α 用于调节 Dice 损失和 Focal 损失之间的权重比例。

$$L_{\text{Dice}} = 1 - \frac{2 \cdot \sum_{i=1}^N p_i \cdot y_i}{\sum_{i=1}^N p_i^2 + \sum_{i=1}^N y_i^2} \quad (14)$$

$$L_{\text{Focal}} = -\lambda (1 - p_i)^\gamma \log(p_i) \quad (15)$$

式中: N 为类别数; p_i 为模型对类别 i 的预测概率; y_i 为真实标签; p_i 表示模型对目标类别的预测概率; λ 是调节正负样本权重的超参数; γ 则控制难分类样本的关注程度。

Dice 损失常用于类别不平衡的语义分割任务中, 通过计算模型预测的概率 p_i 与真实标签 t_i 之间的重叠度来计算损失值。该损失函数的设计目的是最大化预测区域与真实区域的重叠, 推动模型在类别不平衡的情况下生成更多的正确重叠区域, 从而提高分割精度。Focal 损失的设计目的是让模型聚焦于难以分类的样本。在类别不平衡的任务中, 容易分类的样本占据主导地位, 导致模型没有充分学习到困难样本的特征, 而 Focal 损失通过加大对难分类样本的关注缓解了这一问题, 提升模型在复杂场景下的表现。

2 实验结果与分析

2.1 数据集介绍

实验采用了 DeepGlobe 和 CHN6-CUG 数据集验证本文所提算法的有效性。DeepGlobe 数据集是由 DeepGlobe 2018 挑战赛提供的高分辨率遥感图像数据集。数据集中的图像分辨率为 $1\,024 \times 1\,024$ 像素。数据集包含 6 226 张图像, 其中, 训练集图像为 5 603 张, 验证集图像为 623 张。CHN6-CUG 数据集包含来自北京、上海、武汉、广州、成都、重庆 6 个城市的遥感图像, 覆盖面积约 2 000 km^2 , 图像分辨率为 512×512 像素。数据集包含 4 511 张图像, 其中, 训练集图像为 4 059 张, 验证集图像为 452 张。

2.2 实验条件

2.2.1 实验环境

实验基于 PyTorch 2.4.1 深度学习框架和 CUDA12.4 库, 使用 Python 3.8 环境, 训练时使用 NVIDIA GeForce RTX 4060, 显存为 8 GiB。训练参数 BatchSize 设置为 4, 初始学习率为 0.000 2, 训练总轮次 Epochs 为 200 轮次。

2.2.2 评价指标

实验中采用精确度 (Precision, 式中简记 P)、

召回率 (Recall, 式中简记 R)、F1 分数 (F1-score, 式中简记 F_1)、平均交并比 (mIoU)。参数 (Params) 和运算量 (FLOPs) 6 个指标进行评估。具体公式为:

$$P = \frac{N_{TP}}{N_{TP} + N_{FP}}$$

(16)

$$R = \frac{N_{TP}}{N_{TP} + N_{FN}}$$

(17)

$$F_1 = \frac{2 \times P \times R}{P + R}$$

(18)

$$mIoU = \frac{1}{k + 1} \sum_{i=0}^k \frac{N_{TP}}{N_{FN} + N_{FP} + N_{TP}}$$

(19)

2.3 实验结果分析

2.3.1 权重参数分配

所提算法中采用了 Dice 损失和 Focal 损失的复合损失函数,分别将式(13)中的 α 设置为 0.1、0.3、0.5、0.7 和 0.9 进行实验,并通过模型性能评价指标 mIoU 分析不同权重比例下的网络性能。表 1 为不同权重比例下的性能对比结果。

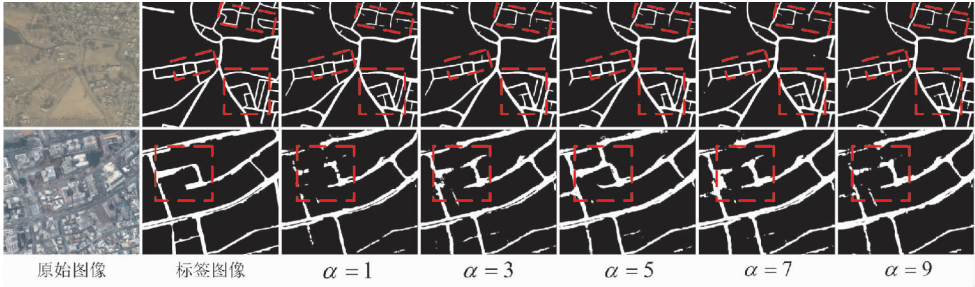


图 4 损失函数不同权重实验结果

Fig. 4 Experimental results of different weights of loss function

2.3.2 DeepGlobe 数据集结果对比

将本文所提模型与多个主流模型在 DeepGlobe 和 CHN6-CUG 数据集上进行对比实验。参与对比的模型包括 Unet^[9]、Deeplabv3 +^[18]、PSP-

表 1 损失函数不同权重实验对比
Tab. 1 Experimental comparison of loss function with different weights
单位:/%

α	mIoU	
	DeepGlobe 数据集	CHN6-CUG 数据集
0.1	81.59	78.29
0.3	81.84	78.86
0.5	82.01	79.36
0.7	81.73	79.27
0.9	81.68	78.92

实验结果表明,当 $\alpha = 0.5$ 时,在 2 个数据集上模型的 mIoU 相较于其他权重比例达到最优,分别为 82.01% 和 79.36%。图 4 是在 2 个数据集上各选取一张图片的可视化结果对比,从图 4 中可以看出,当 $\alpha = 0.5$ 时,模型的分割结果在边界清晰度和目标完整性上表现最佳。

net^[19]、Hrnet^[20]、SGCN^[21]、EGE-Unet^[22]、VM-Unet^[23]、CFRNet^[24]、Segman^[25] 等。在 DeepGlobe 数据集上对比实验结果如表 2 所示。

表 2 在 DeepGlobe 数据集上的道路提取结果

Tab. 2 Road extraction results on the DeepGlobe dataset

模型	Precision/%	Recall/%	F1-score/%	mIoU/%	Params/ 10^6	FLOPs/ 10^9
Unet ^[9]	89.08	86.88	87.61	80.26	31.03	14.86
Deeplabv3 + ^[18]	83.03	84.68	83.85	75.02	41.25	22.54
Pspnet ^[19]	81.40	66.89	73.44	62.90	46.52	26.84
Hrnet ^[20]	88.36	81.45	84.76	75.92	59.30	35.84
SGCN ^[21]	94.86	53.38	68.31	74.67	36.82	24.36
EGE-Unet ^[22]	76.93	69.44	72.99	65.59	15.38	9.82
VM-Unet ^[23]	78.10	71.55	74.68	75.28	22.15	12.53

续表 2

方法	Precision/%	Recall/%	F1-score/%	mIoU/%	Params/ 10^6	FLOPs/ 10^9
CFRNet ^[24]	85.91	78.62	82.11	78.90	28.50	16.83
Segman ^[25]	91.56	82.89	87.02	80.44	45.00	9.94
Ours	90.44	88.09	89.25	82.01	32.72	17.50

从表 2 可以看出,所提模型的 Recall、F1-score 和 mIoU 分别达到了 88.09%、89.25%、82.01%,均优于其他模型。但 Precision 达到 90.44%,次于最优 4.42%,原因在于 SGCN 的核心是图卷积操作,这使得模型对典型地物的预测置信度很高,误判背景为目标的概率极低,因此表现出优异的 Precision。但是这一优势也伴随着对局部细节敏感性的削弱,在面对难分类样本时,模型倾向于保守预测,将其判定为背景,从而造成大量漏检,直接拉低了 Recall 指标。所提模型参数量为 32.72 M,相较于基础模型增加约 5.4%的参数,但获得更好的分割精度和交并比,在性能与计算开销之间取得了良好平衡。总的来说,相较于其他经典模型,论文所提算法取得了一定优势,综

合来看性能优越。

图 5 给出了部分对比模型与本文所提模型 AFF-Unet 在 DeepGlobe 数据集上的可视化对比结果。图 5 第 2 行中 Unet 模型出现了分割不完整的问题,Deeplabv3 +、Hrnet、SGCN、VM-Unet 在图像中道路连接处分割效果较差,Pspnet 模型的分割结果中出现道路间断的问题。在第 3、4 行有障碍物遮挡和复杂背景的情况下,对比模型不同程度地出现了漏检和错误提取的情况,其中,SGCN 和 VM-Unet 出现的错误提取较为严重,本文所提模型虽然在道路连接处存在少许间断,但仍有效保障了道路的连续性和完整性,整体表现优于其他模型。

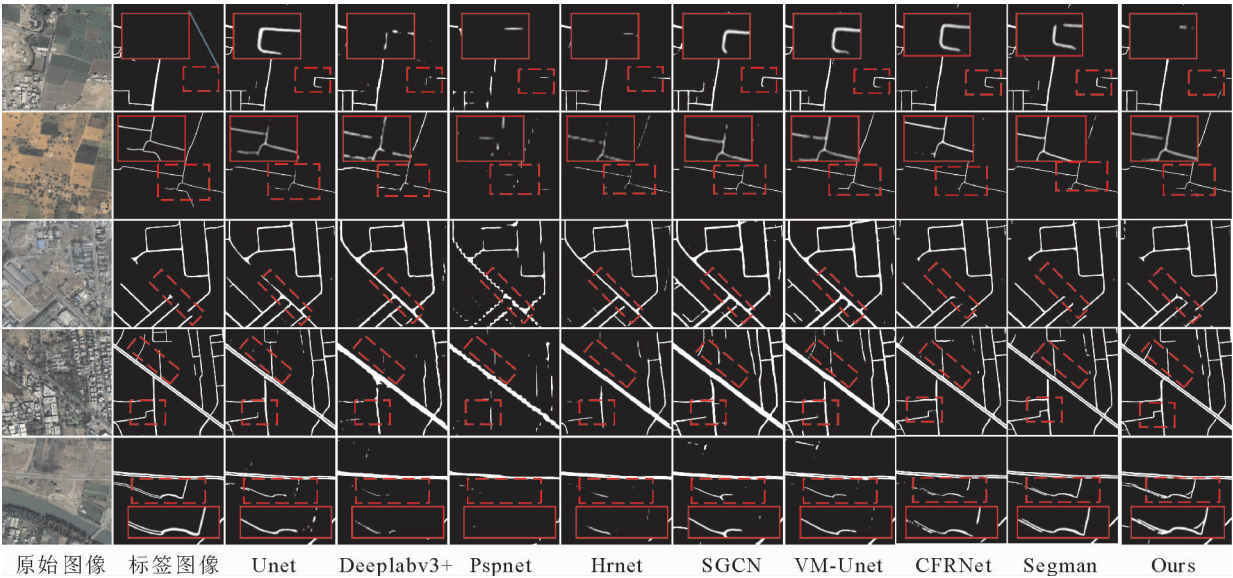


图 5 AFF-Unet 与对比模型在 DeepGlobe 数据集提取结果

Fig. 5 Extraction results of AFF-Unet and contrast model in DeepGlobe dataset

2.3.3 CHN6-CUG 数据集结果对比

CHN6-CUG 数据集上对比结果如表 3 所示。从表 3 中可以看出,本文所提模型在 4 个评价指标上分别达到了 90.32%、85.02%、87.59% 和 79.36%。SGCN 的 Precision 指标为 88.97%,但其 Recall 和 F1-score 明显较低,分别为 52.64% 和

66.14%,表明该模型在漏检问题上较为严重,综合表现不佳。Unet、Deeplabv3 +、Hrnet 等经典网络虽然在 Precision 和 Recall 上表现较好,但在 mIoU 上仍有一定差距。所提模型综合性能上相较于其他经典模型表现更加优越,进一步验证了所提模型的有效性。

表 3 在 CHN6-CUG 数据集上的道路提取结果

Tab. 3 Road Extraction Results on the CHN6-CUG Dataset

单位:/%

模型	Precision	Recall	F1-score	mIou
Unet ^[9]	89. 29	84. 09	86. 61	78. 05
Deeplabv3 + ^[18]	88. 02	84. 87	86. 42	77. 89
Pspnet ^[19]	88. 43	84. 01	86. 16	77. 51
Hrnet ^[20]	89. 94	83. 33	86. 51	77. 82
SGCN ^[21]	88. 97	52. 64	66. 14	67. 10
EGE-Unet ^[22]	48. 91	49. 99	49. 45	60. 91
VM-Unet ^[23]	73. 33	61. 95	67. 17	67. 86
CFRNet ^[24]	85. 23	79. 88	82. 47	75. 42
Segman ^[25]	88. 95	84. 32	86. 57	78. 82
Ours	90. 32	85. 02	87. 59	79. 36

图 6 展示了对比模型与本文所提模型 AFF-Unet 在 CHN6-CUG 数据集上的可视化结果对比,并用红色矩形框标注了差异部分。从第 1、3、5 行的结果可以看出,论文所提算法在提取细长道路时表现出了较好的完整性和连续性,而其他对比模型在不同程度上存在道路提取中断或无法提取的情况。在第 1、3 行有房屋阴影、树木等障碍物

遮挡时,Hrnet 和 SGCN 的漏检情况比较严重。在第 4 行提取图像边缘道路时,部分对比模型出现了大范围的间断问题,而本文所提模型则能够较为完整地提取道路,确保了较好的连续性和完整性。总体来说,本文所提模型在分割效果上明显优于其他对比模型。

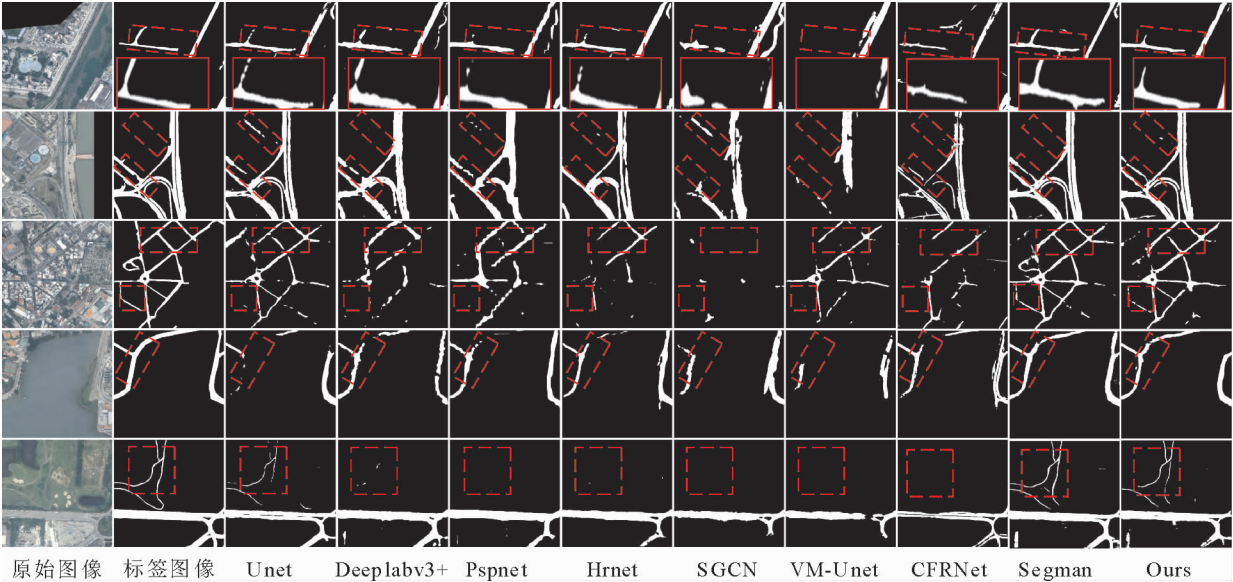


图 6 AFF-Unet 与对比模型在 CHN6-CUG 数据集提取结果

Fig. 6 Extraction results of AFF-Unet and comparison model in CHN6-CUG dataset

图 5 和图 6 分别展示了在 DeepGlobe 和 CHN6-CUG 数据集上采用不同模型的道路提取结果。通过对 2 个具有显著不同地理特征与场景类型的遥感图像数据集进行实验测试,模型在不同环境条件下均保持了稳定的分割精度,表现出较强的适应性与一致性。该结果充分验证了所提模

型具有良好的泛化能力,能够有效应对不同地貌特征和复杂场景下的道路提取任务。

2. 3. 4 消融实验

为了评估各个模块对模型性能的影响,分别对 CEMA 模块和 ACmix 模块进行消融实验。首先将 CEMA 嵌入网络不同位置,并在 2 个数据集

上进行实验,结果如表 4 所示。当 CEMA 嵌入在第 1 层跳跃连接时,在 DeepGlobe 数据集上 Precision 提升了 0.54%, Recall 提升了 1.07%, F1-score 提升了 1.17%, mIoU 提升了 1.11%。在 CHN6-CUG 数据集上 4 个指标相较于原模型分别提升了 0.08%、0.14%、0.11%、0.15%。CEMA 模块可以在特征学习的早期阶段对输入特征进行平滑,减少噪声、避免梯度消失,并增强初步特征的稳定性。这有助于网络从源头上提取干净且有意义的低层特征,为后续的高层特征学习提供更好的基础。在确定 CEMA 模块的嵌入位置之后,再比较在原模型上添加 ACmix 前后的分割效果,从表 4 可知,添加 ACmix 前后在 DeepGlobe 数据集上 4 个指标分别提升了 1.01%、0.96%、

1.34%、1.33%,均得到了不同程度的提升。在 CHN6-CUG 数据集上分别提升了 0.33%、0.18%、0.25%、0.33%,表明 ACmix 在遥感图像道路提取任务中,通过动态调整特征权重和融合全局与局部信息,显著提升了网络的特征表达能力。其自注意力机制和动态卷积使得网络能够更加灵活和精确地处理复杂的图像特征。加入单个模块后各项评价指标均得到了一定程度的提升,同时加入 2 个模块后,在 DeepGlobe 数据集上, Precision、Recall、F1-score 和 mIoU 相较于加入模块前分别提升了 1.36%、1.21%、1.64%、1.75%,4 个评价指标均达到最优。在 CHN6-CUG 数据集上分别提升了 1.03%、0.93%、0.98%、1.31%。

表 4 消融模型道路提取结果

Tab. 4 Road extraction results of ablation model

单位:/%

方法	加入位置	DeepGlobe 数据集				CHN6-CUG 数据集			
		Precision	Recall	F1-score	mIoU	Precision	Recall	F1-score	mIoU
Unet	—	89.08	86.88	87.61	80.26	89.29	84.09	86.61	78.05
Unet + CEMA	第 1 层	89.62	87.95	88.78	81.37	89.37	84.23	86.72	78.20
Unet + CEMA	第 2 层	89.22	87.27	88.23	80.62	90.08	83.00	86.40	77.62
Unet + CEMA	第 3 层	89.37	85.42	87.35	79.37	88.00	84.44	86.18	77.56
Unet + ACmix	—	90.09	87.84	88.95	81.59	89.62	84.27	86.86	78.38
Unet + CEMA + ACmix	—	90.44	88.09	89.25	82.01	90.32	85.02	87.59	79.36

图 7 为消融实验各模型道路提取的可视化结果。从第 1、2 行可以看出,在将 CEMA 模块加入到第 1 层跳跃连接,并同时加入 Acmix 模块后, AFF-Unet 模型对边缘道路提取更加完整,细节更加充分。从 3、4、6 行可以看出,改进后的模型提取的道路完整性和连续性优于其他消融实验模型。

3 结语

针对遥感图像道路提取任务中道路特征提取不充分、细节信息丢失以及类别不平衡等问题,本文提出一种基于改进 U-Net 的道路提取模型 AFF-Unet。该模型以 VGG16 为骨干网络,通过引入跨空间多尺度注意力模块 CEMA 和自适应卷

积混合模块 ACmix,显著提升了模型对复杂背景和遮挡区域的特征感知能力,并在边缘和细节恢复方面表现出色。此外,为解决样本数量不平衡问题并提高模型对复杂结构的敏感性,将 Dice 损失和 Focal 损失相结合,优化模型在像素级分类任务中的表现。综合实验结果分析,AAF-Unet 模型的 mIoU 指标在 DeepGlobe 数据集上达到了 82.01%,相较于 Unet、Deeplabv3 +、Hrnet 模型分别提高了 1.75%、6.99%、6.09%,在 CHN6-CUG 数据集上达到了 79.36%,相较于 Unet、Deeplabv3 +、Hrnet 模型分别提高了 1.31%、1.47%、1.54%,通过实验验证了算法整体的有效性。下一步研究将围绕算法性能优化和模型轻量化展开,以进一步提升模型的实用性和泛化能力。

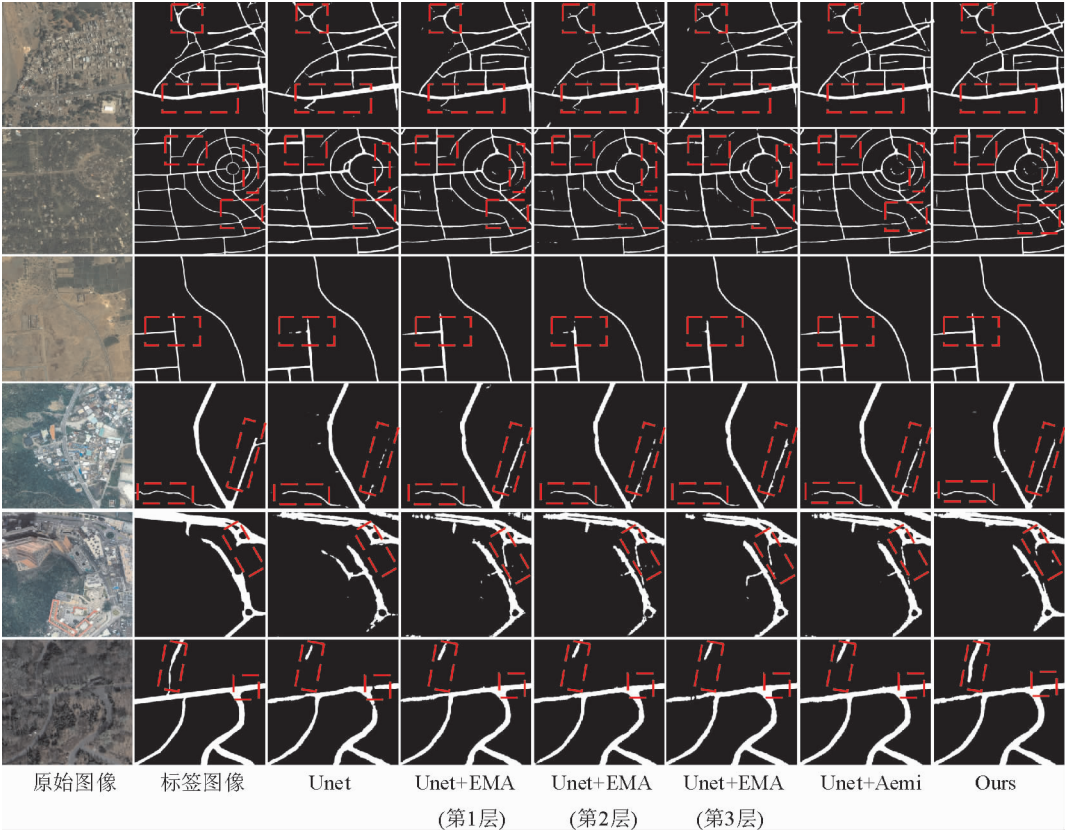


图 7 消融模型可视化提取结果
Fig.7 Visual extraction results of ablation model

参考文献

[1] 李云红, 魏小双, 苏雪平, 等多尺度特征增强与交互融合的遥感小目标检测[J]. 西北大学学报(自然科学版), 2025, 55(2): 277-285.
LI Y H, WEI X S, SU X P, et al. Remote sensing small target detection based on feature multi-scale enhancement and interaction fusion [J]. Journal of Northwest University (Natural Science Edition), 2025, 55(2): 277-285.

[2] CHEN R N, LI X, HU Y, et al. Road extraction from remote sensing images in wildland-urban interface areas[J]. IEEE Geoscience and Remote Sensing Letters, 2022, 19: 1-5.

[3] 胡宏宇, 左记祥, 吕颖, 等. 面向自动驾驶的遥感影像路网检测方法[J]. 中国公路学报, 2022, 35(11): 310-317.
HU H Y, ZUO J X, LYU Y, et al. Road network detection of remote sensing images for autonomous driving [J]. China Journal of Highway and Transport, 2022, 35(11): 310-317.

[4] ZHANG X Y, JIANG Y, WANG L Z, et al. Complex mountain road extraction in high-resolution remote sensing images via a light roadformer and a new bench-

mark[J]. Remote Sensing, 2022, 14(19): 4729.

[5] ZHAO K J, LIU J, WANG Q N, et al. Road damage detection from post-disaster high-resolution remote sensing images based on TLD framework [J]. IEEE Access, 2022, 10: 43552-43561.

[6] 李云红, 张富星, 苏雪平, 等. 增强上下文特征交互的实时无人机影像分割算法[J]. 北京航空航天大学学报, 2026, 52(3): 668-677.
LI Y H, ZHANG F X, SU X P, et al. Real-time UAV image segmentation algorithm with enhanced contextual feature interaction[J]. Journal of Beijing University of Aeronautics and Astronautics, 2026, 52(3): 668-677.

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

[8] ZHONG Z L, LI J, CUI W H, et al. Fully convolutional networks for building and road extraction: Preliminary results[C] // 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). Beijing: IEEE, 2016: 1591-1594.

[9] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional networks for biomedical image segmenta-

- tion[M] // Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Munich: Springer International Publishing, 2015: 234-241.
- [10] 佟喜峰,张婉莹. 基于改进 U-Net 模型的遥感影像道路提取方法研究[J]. 计算机与数字工程, 2024, 52(5): 1495-1501.
- TONG X F, ZHANG W Y. Research on road extraction method from remote sensing image based on improved U-Net model[J]. Computer & Digital Engineering, 2024, 52(5): 1495-1501.
- [11] 何哲,陶于祥,罗小波,等. 基于改进 U-Net 的遥感图像道路提取[J]. 激光与光电子学进展, 2023, 60(16): 1628004.
- HE Z, TAO Y X, LUO X B, et al. Road extraction from remote sensing image based on an improved U-Net[J]. Laser & Optoelectronics Progress, 2023, 60(16): 1628004.
- [12] 齐然然,帕力旦·吐尔逊,汤泊川,等. 基于残差注意力编-解码网络的道路提取方法[J]. 计算机工程与科学, 2025, 47(1): 119-129.
- QI R R, PALIDAN T, TANG B C, et al. A road extraction method based on residual attention encoder-decoder network[J]. Computer Engineering & Science, 2025, 47(1): 119-129.
- [13] 郝明,白鹤,徐婷婷. 融合 ResNeSt 和多尺度特征融合的遥感影像道路提取[J]. 光电工程, 2025, 52(1): 240236.
- HAO M, BAI H, XU T T. Remote sensing image road extraction by integrating ResNeSt and multi-scale feature fusion[J]. Opto-Electronic Engineering, 2025, 52(1): 240236.
- [14] 曲海成,王莹,刘腊梅,等. 融合 CNN 与 Transformer 的遥感影像道路信息提取[J]. 自然资源遥感, 2025, 37(1): 38-45.
- QU H C, WANG Y, LIU L M, et al. Information extraction of roads from remote sensing images using CNN combined with Transformer[J]. Remote Sensing for Natural Resources, 2025, 37(1): 38-45.
- [15] 唐伟,赵书河. 基于 GVF 和 Snake 模型的高分辨率遥感图像四元数空间道路提取[J]. 遥感学报, 2011, 15(5): 1040-1052.
- TANG W, ZHAO S H. Road extraction in quaternion space from high spatial resolution remotely sensed images basing on GVF Snake model[J]. Journal of Remote Sensing, 2011, 15(5): 1040-1052.
- [16] ZHANG Z Q, BAO L Y, XIANG S, et al. B2CNet: A progressive change boundary-to-center refinement network for multi temporal remote sensing images change detection[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2024, 17: 11322-11338.
- [17] 徐虹,杨莹洁,文武,等. 融合双域特征均衡的遥感图像道路提取[J]. 电讯技术, 2024, 64(6): 878-886.
- XU H, YANG Y J, WEN W, et al. Road extraction from remote sensing images integrating dual-domain feature equalization[J]. Telecommunication Engineering, 2024, 64(6): 878-886.
- [18] 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. Munich: Springer International Publishing, 2018: 833-851.
- [19] 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. Honolulu, HI: IEEE, 2017: 6230-6239.
- [20] WANG J D, SUN K, CHENG T H, et al. Deep high-resolution representation learning for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(10): 3349-3364.
- [21] ZHOU G D, CHEN W T, GUI Q S, et al. Split depth-wise separable graph-convolution network for road extraction in complex environments from high-resolution remote-sensing images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 5614115.
- [22] RUAN J C, XIE M Y, GAO J S, et al. EGE-UNet: An efficient group enhanced UNet for skin lesion segmentation[C] // Medical Image Computing and Computer Assisted Intervention – MICCAI 2023. Cham: Springer, 2023: 481-490.
- [23] RUAN J C, LI J C, XIANG S C. VM-UNet: Vision mamba UNet for medical image segmentation[J]. ACM Transactions on Multimedia Computing, Communications, and Applications, 2025: 3767748.
- [24] DENG B, LIU D, CAO Y, et al. CFRNet: Cross-attention-based fusion and refinement network for enhanced RGB-T salient object detection[J]. Sensors, 2024, 24(22): 7146.
- [25] FU Y X, LOU M, YU Y Z. SegMAN: Omni-scale context modeling with state space models and local attention for semantic segmentation[C] // 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, IN: IEEE, 2025: 19077-19087.