引文格式：彭劲松,赵林峰,刘慧娟,等. DSMC：全局—局部特征混合编码的遥感影像道路分割模型[J]. 地理空间信息, 2026, 24(5): 1-5. doi:10.3969/j.issn. 1672-4623.2026.05.001

DSMC：全局—局部特征混合编码的遥感影像道路分割模型

彭劲松[1]，赵林峰[1]，刘慧娟[1]，陈　琼[1]，许　慧[1]，陈海佳[2]
（1. 湖南环境生物职业技术学院，湖南 衡阳 421005；2. 武汉象印科技有限责任公司，湖北 武汉 430200）

摘　要：针对现有方法分割遥感影像内部道路结构时，存在的较多分割结果断裂、细小地物错误分割等问题，提出了一种全局—局部特征混合编码的遥感影像道路分割模型。采用深度可分离卷积与第二代窗口转换器对道路的细节与全局上下文信息进行混合编码；解码阶段采用内容感知特征重组与混合局部通道注意力单元对小尺寸特征进行负样本过滤与精密尺寸上采样；以统一焦点损失为训练损失函数，通过参数加权引导模型学习道路样本特征。结果表明，该模型在中国代表性城市新型大尺度卫星遥感影像道路数据集与深度全球—道路数据集上的F_1-score分别为90.68%与91.35%，比窗口视觉转换器—U型网络模型提高了4.26%与3.81%，且性能表现优于金字塔场景解析网络、增强U型网络等模型。

关键词：遥感道路分割；全局—局部特征混合编码；内容感知特征重组；特征过滤；统一焦点损失
中图分类号：P237　　**文献标志码：**B　　**文章编号：**1672-4623（2026）05-0001-05

DSMC: A Remote Sensing Image Road Segmentation Model Based on Global-Local Feature Hybrid Encoding

PENG Jinsong[1], ZHAO Linfeng[1], LIU Huijuan[1], CHEN Qiong[1], XU Hui[1], CHEN Haijia[2]
(1. Hunan Environmental Biology Vocational and Technical College, Hengyang 421005, China;
2. Wuhan Xiangyin Technology Co., Ltd., Wuhan 430200, China)

Abstract: Aiming at the problems of fragmented segmentation results and error segmentation of small objects when existing methods being used to segment the internal road structure of remote sensing images, we proposed a remote sensing image road segmentation model based on global-local feature hybrid encoding. We used deep separable convolution and shift window TransformerV2 to hybrid encode the details of roads and global context information. In the decoding stage, we employed content-aware reassembly of features and hybrid local channel attention units to filter negative samples and perform fine-scale upsampling of small-size features, and used unified focal loss as the training loss function to guide the model to learn road sample features through parameter weighting. Experimental results show that the F_1-scores of this model on the public datasets CHN6-CUG and DeepGlobe Road are 90.68% and 91.35%, respectively, which are 4.26% and 3.81% higher than those of Swin-UNet model. Compared with PSPNet, UNet++ and other models, the proposed model exhibits the best performance.

Key words: remote sensing road segmentation, global-local feature hybrid encoding, content-aware reassembly of features, feature filter, unified focal loss

随着在轨卫星数量的增加和星载相机硬件的快速发展[1-2]，遥感影像的时空分辨率不断提高。遥感影像智能解译技术能提取遥感影像内部像素级的道路结构[3-4]，为城市建设规划、地图绘制更新、应急救灾等场景提供技术支持[5-6]。为实现大尺度场景下的道路结构分割，采用轮廓等特征构建提取算子以及机器学习算法匹配遥感影像内的道路目标[7]；但该类方法构造的特征过多依赖人工先验知识，且传统机器学习无法处理复杂的上下文特征，易受相似特征地物干扰，出现错误分割或道路结构断裂情况[8]。随着深度学习技术的进步，具备强大深层特征学习能力的卷积神经网络（CNN）和视觉转换器（VIT）模型，被应用于多类遥感影像语义分割领域。

现阶段基于深度学习的遥感道路提领模型已取得不少研究进展[9-11]，但仍存在一些问题：在特征编码阶段，CNN层关注道路细节特征，无法进行全局上下文建模，而VIT层虽能捕获全局上下文注意力，但容易忽略道路的边缘细节特征；在特征解码阶段，现有模型大多通过双线性插值方式完成上采样，无法充分恢复图像内的道路细节特征；在训练过程中，道路与无

项目来源：2023年度湖南省教育厅科学研究资助项目（23C0634）；湖南环境生物职业技术学院支柱工程。
第一作者简介：彭劲松（1975—），硕士，副教授，高级工程师，研究方向为园林工程测量、建筑工程测量教学及GPS全球定位、无人机摄影测绘技术应用推广，E-mail：JsPeng1975@163.com。

关背景的占比严重失衡，无法引导模型高效学习道路目标特征。鉴于此，本文设计了一种全局—局部特征混合编码的遥感影像道路分割模型。首先，采用深度可分离卷积（DSC）单元搭建双层跨阶段部分融合瓶颈卷积（C2f）架构的细节特征提取层，并在提取层间采用哈尔小波下采样（HWD）算子保留更多道路细节特征；采用第二代窗口转换器（Swin TransformerV2）搭建全局特征提取层，并在编码器深层对道路特征进行全局上下文建模；结合DSC与Swin TransformerV2，搭建混合特征编码网络（DSNet）。其次，结合混合局部通道注意力（MLCA）单元与内容感知特征重组（CARAFE）单元，构建带有特征过滤与精密上采样机制的解码器网络（MCNet），在解码过程中准确还原道路细节特征。最后，以统一焦点损失（UFL）为训练损失函数，抑制正负样本不均衡问题给模型学习带来的负样本偏置，为便于表述，将本文模型命名为DSMC。

1 研究方法

1.1 混合特征编码器

为协同捕获道路细节与全局上下文特征，结合DSC与SwinTransformerV2单元，搭建DSNet，各结构见图1。DSNet前两层采用卷积算子提取特征更丰富的浅层特征图。提取层采用C2f结构搭建，输入的特征图经过通道切分后，一半特征图由3组核心算子DSC与1×1逐点卷积（PW）组成的瓶颈结构模块（DSC Block）挖掘细节特征，并将逐层提取结果与未提取处理特征图融合。在局部特征提取层间，为避免传统池化或2步长卷积核下采样导致的道路细节信息大量丢失，采用HWD单元对特征图下采样。HWD将特征图沿水平方向进行一阶段变换，即

$$\begin{cases} Y_L^{(1)}(i,j)=\dfrac{X(i,2j)+X(i,2j+1)}{\sqrt{2}} \\ Y_H^{(1)}(i,j)=\dfrac{X(i,2j)-X(i,2j+1)}{\sqrt{2}} \end{cases} \tag{1}$$

式中，$X(i,2j)$为坐标为$(i,2j)$的像素点；$Y_L^{(1)}$为一阶低频分量；$Y_H^{(1)}$为一阶高频分量。

在$Y_L^{(1)}$与$Y_H^{(1)}$基础上，再次开展列维度哈尔小波变换，将特征图分解为二阶低频Y_{LL}、水平高频Y_{LH}、垂直高频Y_{HL}、对角高频Y_{HH}四组分量，则有：

$$\begin{cases} Y_{LL}(i,j)=\dfrac{Y_L^{(1)}(2i,j)+Y_L^{(1)}(2i+1,j)}{\sqrt{2}} \\ Y_{LH}(i,j)=\dfrac{Y_L^{(1)}(2i,j)-Y_L^{(1)}(2i+1,j)}{\sqrt{2}} \\ Y_{HL}(i,j)=\dfrac{Y_H^{(1)}(2i,j)+Y_H^{(1)}(2i+1,j)}{\sqrt{2}} \\ Y_{HH}(i,j)=\dfrac{Y_H^{(1)}(2i,j)-Y_H^{(1)}(2i+1,j)}{\sqrt{2}} \end{cases} \tag{2}$$

4组二阶分量的平面尺寸降至原图的1/2，通道数量则扩张一倍。利用1×1卷积算子进行通道级特征图融合，再通过批量归一化层与乙型修正函数对特征进行归一——非线性化处理。网络深层采用Swin TransformerV2搭建全局上下文建模层。前一层输出的特征图经过补丁化与线性投影转换为查询$\boldsymbol{Q}$、键$\boldsymbol{K}$、值$\boldsymbol{V}$三组向量，即

$$\begin{cases} \boldsymbol{Q}=\boldsymbol{X}\times\boldsymbol{W}^{Q} \\ \boldsymbol{K}=\boldsymbol{X}\times\boldsymbol{W}^{K} \\ \boldsymbol{V}=\boldsymbol{X}\times\boldsymbol{W}^{V} \end{cases} \tag{3}$$

式中，$\boldsymbol{X}$为线性投影后的特征向量；$\boldsymbol{W}^{Q}$、$\boldsymbol{W}^{K}$、$\boldsymbol{W}^{V}$为查询、键、值对应的转换矩阵。

获取3组向量后，Swin TransformerV2采用缩放余弦注意力计算任意像素间的自注意力权重，计算过程为：

$$\mathrm{SCA}(q_i,k_j)=\frac{\cos(q_i,k_j)}{\tau}+B_{ij} \tag{4}$$

式中，q_i、k_j为像素i的查询头和像素j对应的键；cos()为余弦相似度；τ为缩放余弦相似度的可学习参数；B_{ij}为像素i、j之间的相对位置偏差。

为使各窗口间相对位置偏差更平滑的传递，Swin TransformerV2通过一个多重感知机（MSA）构成的小型网络，计算任意像素间相对位置偏差。计算多头注意力后，再对输出特征进行层归一化操作，缓解特征内部的协变量偏移问题。Swin TransformerV2的完整计算过程为：

$$\begin{cases} \hat{x}^{l}=LN\left(\mathrm{MSA}\left(x^{l-1}\right)\right)+x^{l-1} \\ x^{l}=LN\left(\mathrm{MLP}\left(\hat{x}^{l}\right)\right)+\hat{x}^{l} \\ \hat{x}^{l+1}=LN\left(\mathrm{MSA}\left(x^{l}\right)\right)+x^{l} \\ x^{l+1}=LN\left(\mathrm{MLP}\left(\hat{x}^{l+1}\right)\right)+\hat{x}^{l+1} \end{cases} \tag{5}$$

式中，x^{l-1}为输入特征；x^{l}为赋权特征；$\hat{x}^{l+1}$为MSA层计算的多头注意力；x^{l+1}为输出特征；$LN()$为层归一化。

1.2 精密上采样过滤解码器

为充分恢复道路细节特征并过滤无关样本，结合样本过滤机制与精密上采样模块，搭建解码器网络

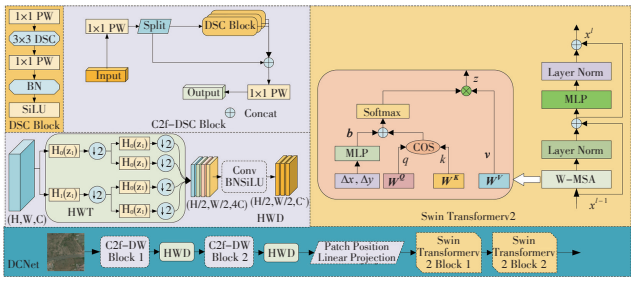


图 1　DSNet 网络各结构图

MCNet。MCNet 以编码器各层输出结果为输入，通过由 MLCA 单元与 DSC 单元组成的样本过滤模块（MADB）过滤无关背景特征。MLCA 单元先通过平均池化将输入特征图尺寸由 C、W、H 压缩至 C、ks、ks；再通过独立分支分别计算全局和局部注意力；最后采用逐元素相乘混合全局—局部注意力图，并通过 UNAP 将混合注意力恢复至原始尺寸，为输入特征图赋权。在全局注意力分支，利用全局平均池化操作和 1×1 PW 算子将特征图压缩为一维向量，并计算全局注意力信息，再通过反池化（UNAP）将分辨率恢复至 C、ks、ks。在局部注意力分支，利用转换和 1×1 PW 算子将特征图变形至 1、1、$C\times ks^2$ 大小，并计算局部注意力信息，再通过 UNAP 恢复至输入分辨率。

在利用 MADB 模块过滤特征图的基础上，采用 CARAFE 对大尺度特征图进行精细上采样重建。CARAFE 上采样包括核预测（KPM）与内容感知重组（CARM）两部分。KPM 模块首先通过 1×1 PW 与 3×3 深度卷积（DW）组合算子，对特征图进行通道压缩与内容编码；然后通过柔性最大归一化得到动态上采样核，即

$$W=\mathrm{Softmax}\left(DW_{3,3}\left(PW_{1\times1}(F)\right)\right) \tag{6}$$

式中，F 为输入特征图；$PW_{1\times1}$、$DW_{3,3}$ 分别为 PW 与 DW 卷积核。

CARM 模块基于 W 完成上采样重建，即

$$F_{l'}=\sum_{n=-r}^{r}\sum_{n=-r}^{r}W_{(m,n)}F_{(i+m,j+n)} \tag{7}$$

1.3　组合损失函数

当前语义分割模型大多通过二元交叉熵（BCE）函数、骰子（Dice）函数来计算模型训练损失。BCE 函数可用于计算模型对各类像素的分类精度，Dice 函数可用于计算整体分割目标与真实掩膜间的相似度。然而，在遥感影像场景下，道路与其余样本间的像素比例差异较大，而 Dice、BCE 函数均未抑制负样本在训练阶段的特征表达，将导致训练阶段无法充分学习道路样本特征，因此本文以融合校正焦点特沃斯基（mFT）和修正焦点（mF）函数的 UFL 作为训练损失函数。UFL 对 mFT 与 mF 函数的超参数进行合并，并通过各类目标加权的方式抑制无关背景在训练阶段的表达，完整函数表达式为：

$$\begin{cases}L_{\mathrm{UFL}}=\lambda\times L_{\mathrm{mF}}+(1-\lambda)L_{\mathrm{mFT}}\\ L_{\mathrm{mF}}=\delta(1-p_t)^{1-\gamma}\\ L_{\mathrm{mFT}}=\sum_{k=1}^{K}(1-\mathrm{mTI})^{\gamma}\end{cases} \tag{8}$$

式中，$\lambda\in[0,1]$ 为两组损失函数的参与权重，本文设置 $\lambda=0.5$；δ 为针对输出不均衡问题的控制超参数，本文设置 $\delta\in[0.1,0.9]$；p_t 为模型对道路像素类别的预测概率；γ 为控制难分类样本对损失贡献的超参数，本文设置 $\gamma\in[0.1,0.9]$；K 为像素类别总数；mTI 为修正特沃斯基指标，通过为假阳性和假阴性样本分配权重来克服正负样本间不平衡问题。

$$\mathrm{mTI}=\frac{\sum_{i=1}^{N}p_{0i}g_{0i}}{\sum_{i=1}^{N}p_{0i}g_{0i}+\delta\sum_{i=1}^{N}p_{0i}g_{1i}+(1-\delta)\sum_{i=1}^{N}p_{1i}g_{0i}} \tag{9}$$

式中，p_{0i}、p_{1i} 为预测像素 i 属于道路样本、背景的概率；g_{0i}、g_{1i} 为真实像素 i 属于背景、目标的概率。

2　实验与结果分析

2.1　实验数据集

本文选取中国代表性城市新型大尺度卫星遥感影像道路数据集（CHN6-CUG）[9]和深度全球-道路（DeepGlobe Road）数据集[10]开展模型性能验证实验。CHN6-CUG数据集包含4 511张尺寸为512像素×512像素的遥感影像，单幅影像空间分辨率为50 cm；DeepGlobe Road数据集包含6 226张尺寸为1 024像素×1 024像素的遥感影像，单幅影像空间分辨率为50 cm。将两个数据集以 8∶1∶1 的比例，随机划分为训练、验证、测试3组子数据集。

2.2　实验配置

实验设备为云端服务器，CPU 为 Xeon Platinum 8352V，GPU 为 Nvidia RTX4090，运行内存为 24 GB，操作系统为 Ubuntu 22.04。训练阶段采用带权重衰减的自适应矩估计参数优化器，设置一阶动量参数为 0.925，权重衰减系数为 0.000 5，初始学习率为 0.001，训练迭代次数为 100 次，单次参与训练样本数为 18 个。采用随机水平和垂直翻转策略，将训练图像样本及其对应掩膜以 20% 的概率随机翻转后送入模型参与训练，以提高模型泛化性能。测试阶段选取精确率、召回率、F_1-score、平均交并比（mIoU）作为模型的分割精度评价指标。

2.3　实验结果分析

为评估各模型组件在道路分割任务中的性能，首先在 CHN6-CUG 数据集上，针对混合特征编码、特征过滤与精密上采样开展结构对比实验；再以 UNet++为

基线模型，开展针对各模块的消融实验；最后为验证DSMC的先进性，选取A、B、C、D、E五组模型作为对照组，在CHN6-CUG与DeepGlobe Road数据集上开展多模型对比实验。

2.3.1　结构对比实验

为验证DSNet对道路特征的提取效果，在DSMC中将其与全部由DSC单元搭建的编码器网络（DSCNet），以及全部由Swin TransformerV2搭建的编码器网络（Swinv2Net）开展对比实验，3组模型的检测结果见图2。结果表明，DSCNet、Swinv2Net、DSNet的F_1-score分别为83.58%、86.76%、90.68%，DSCNet、Swinv2Net、DSNet的mIoU分别为80.73%、84.25%、88.35%，说明混合编码器在道路特征提取方面的性能优于独立CNN、VIT结构的编码器。

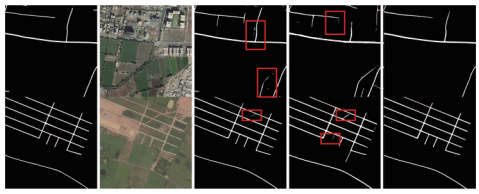


图2　混合编码器模型分割结果

为验证MCNet中精密上采样与特征过滤模块的效果，将其与仅使用CARAFE上采样（MCNet-CARAFE）以及仅使用MLCA模块（MCNet-MLCA）的解码器放入原模型中开展对比实验。结果表明（图3），MCNet-CARAFE、MCNet-MLCA、MCNet的F_1-score分别为85.24%、86.71%、90.68%，MCNet-CARAFE、MCNet-MLCA、MCNet的mIoU分别为83.62%、84.83%、88.35%，说明在特征解码过程中，同时对特征图进行过滤和精密上采样，能取得最理想的分割结果。

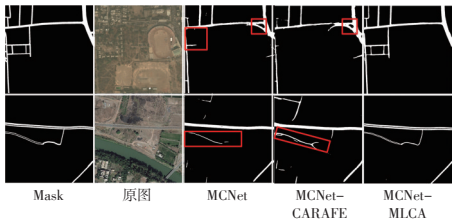


图3　不同解码器模型分割结果

为验证UFL损失函数对道路分割精度的提升效果，将UFL拆分后分别与BCE、Dice函数组合为新损失函数，在DSMC模型中开展对比实验。结果表明（图4），仅替换mF函数或mFT函数作为新损失函数，对应F_1-score有小幅提升，分别达到88.18%与88.59%，而UFL函数的F_1-score为90.68%，证明3种组合方式均能提升当前模型的道路分割精度。

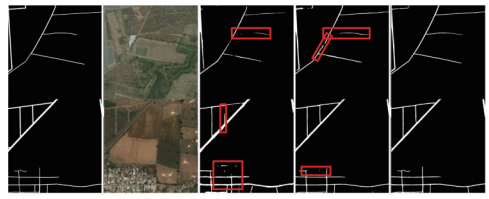


图4　不同损失函数模型分割结果

2.3.2　消融实验

以UNet++为基线模型，分别加入DSNet、MCNet、UFL三组模块，通过结构消融实验验证各模块的重要性。各消融模型的实验结果见表1，可以看出，UNet++基线模型在测试数据集上的F_1-score为79.82%，分别将原生编码器、解码器、BCE-Dice组合损失函数替换为相应消融模块后，对应消融模型的F_1-score分别升至84.63%、82.36%、81.29%，说明3组优化模块能提高模型对道路目标的分割精度；将DSNet、MCNet、UFL以不同的组合形式引入基线模型，对应3组消融模型的F_1-score升至87.36%、86.94%、84.47%，说明各种优化结构的组合使用，仍能有效提升模型的道路分割精度。消融模型分割结果见图5。

表1　消融实验结果

消融模型	Method			F_1-score/%	mIoU/%
	DSNet	MCNet	UFL		
UNet++	–	–	–	79.82	75.49
1	√	–	–	84.63	81.85
2	–	√	–	82.36	78.37
3	–	–	√	81.29	77.63
4	√	√	–	87.36	85.52
5	√	–	√	86.94	84.48
6	–	√	√	84.47	81.56



图5　消融模型分割结果

2.3.3　多模型对比实验

为验证本文模型在道路分割方面的性能优势，以当前主流分割模型金字塔场景解析网络（PSPNet）[11]、稀疏图卷积网络（SGCN）[12]、UNet++[13]、密集连接网络（D-LinkNet）[14]、窗口视觉转换器-U型网络（Swin-UNet）[15]为对照组，在全部采用UFL作为损失函数的情况下，分别在CHN6-CUG与DeepGlobe Road数据集上开展对比实验，结果见表2，可以看出，DSMC在CHN6-CUG数据集上，较对照组模型的F_1-score分别提高了15.34%、10.86%、9.32%、5.94%、4.26%；在DeepGlobe Road数据集上，较对照组模型的F_1-score分别提高了13.96%、10.78%、

8.92%、5.67%、3.81%，说明DSMC在遥感道路分割任务上的精度优于当前的主流语义分割模型，且在不同的场景下分割性能表现稳定。

表2　模型对比实验评价结果/%

数据集	模型	精确率	召回率	F_1-score	mIoU
CHN6-CUG	PSPNet	77.56	79.91	75.34	73.68
	UNet++	80.38	80.95	79.82	77.62
	SGCN	82.54	83.75	81.36	78.35
	D-LinkNet	85.23	85.73	84.74	81.58
	Swin-UNet	88.79	91.29	86.42	84.76
	DSMC	90.42	90.16	90.68	88.35
DeepGlobe Road	PSPNet	79.42	75.46	77.39	75.21
	UNet++	83.51	77.83	80.57	77.82
	SGCN	84.97	80.04	82.43	80.59
	D-LinkNet	88.34	83.18	85.68	83.87
	Swin-UNet	89.26	85.89	87.54	86.33
	DSMC	92.58	90.15	91.35	89.21

进一步通过对比各组模型在CHN6-CUG与DeepGlobe Road数据集多个场景下的道路分割结果，直观说明DSMC的道路分割性能，部分提取结果见图6，可以看出，对照组模型的道路分割结果存在不同程度的线型断裂、非道路结构误提取问题，而DSMC则能准确分割出城市、郊区等不同地块中的道路结构，在道路结构的完整性和细小道路的分割准确性方面表现更优。

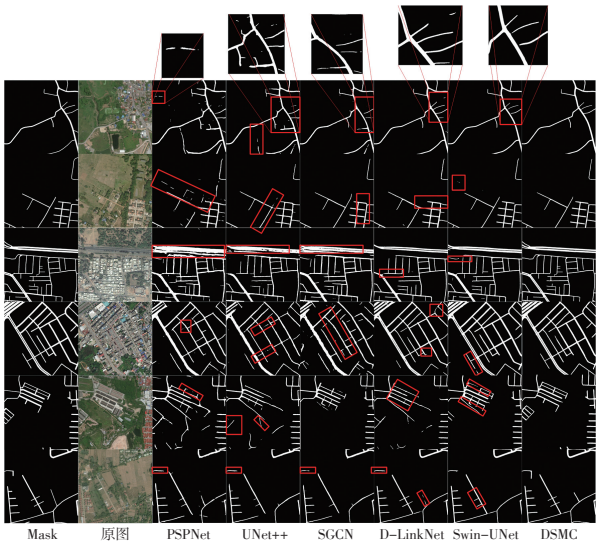


图6　主流模型分割效果对比

3　结　语

针对现有模型分割道路结构时存在的结构断裂、细节不完整、相似地物错误分割等问题，提出了DSMC模型。在CHN6-CUG与DeepGlobe Road数据集上的结果表明，DSMC的F_1-score分别为90.68%与91.35%，mIoU分别为88.35%与89.21%；在不同数据集上的分割精度表现一致，且各指标均优于对照组模型，能为城市规划、应急灾害管理等场景，提供有力的技术支持。后续将进一步优化模型结构，引入多模态遥感影像数据，进一步提高对大场景中道路结构的分割精度。

参考文献

[1] 沈佳洁,曹侃,祁信舒. 自然资源要素遥感智能解译样本分类体系研究[J]. 地理空间信息,2023,21(4):103-106

[2] 张永军,李彦胜,党博,等. 多模态遥感基础大模型:研究现状与未来展望[J]. 测绘学报,2024,53(10):1942-1954

[3] 高玉慧,朱武,张锐煊,等. D-Net:一种动态模拟道路形态的路网提取模型[J]. 地理空间信息,2024,22(6):25-28

[4] 周天舒,周绍光. 分组无监督域适应的遥感影像道路提取[J]. 测绘科学,2024,49(9):115-124

[5] 王子明. 基于注意力机制的遥感影像道路提取方法[J]. 地理空间信息,2025,23(1):13-16

[6] Chen Z, Deng L, Luo Y, et al. Road Extraction in Remote Sensing Data: A Survey[J]. International Journal of Applied Earth Observation and Geoinformation, 2022, 112: 102833

[7] 孙根云,孙超,张爱竹. 融合多尺度与边缘特征的道路提取网络[J]. 测绘学报,2024,53(12):2233-2243

[8] Lian R, Wang W, Mustafa N, et al. Road Extraction Methods in High-resolution Remote Sensing Images: A Comprehensive Review[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2020, 13:5489-5507

[9] 余果,李大成,杨毅. 结合空洞卷积与注意力机制的道路提取方法[J]. 中国空间科学技术(中英文),2024,44(5):175-185

[10] 何士俊,肖提荣,夏既胜. 基于改进DeepLabv3+模型的农村道路提取方法研究[J]. 云南大学学报(自然科学版),2024,46(3):486-495

[11] Li K, Tan M, Xiao D, et al. Research on Road Extraction from High-resolution Remote Sensing Images Based on Improved UNet++[J]. IEEE Access, 2024, 12:50300-50309

[12] Zhu Q, Zhang Y, Wang L, et al. A Global Context-aware and Batch-independent Network for Road Extraction from VHR Satellite Imagery[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2021, 175:353-365

[13] Lu X, Zhong Y, Zheng Z, et al. Cross-domain Road Detection Based on Global-local Adversarial Learning Framework from very High Resolution Satellite Imagery[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2021, 180: 296-312

[14] Zhou J, Hao M, Zhang D, et al. Fusion PSPnet Image Segmentation Based Method for Multi-focus Image Fusion[J]. IEEE Photonics Journal, 2019, 11(6):1-12

[15] Shi L, Wang L, Long C, et al. SGCN: Sparse Graph Convolution Network for Pedestrian Trajectory Prediction[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021:8994-9003