

基于多尺度差异聚合机制的遥感影像道路提取^①



许 明, 周春晖, 姜彦吉

(辽宁工程技术大学 软件学院, 葫芦岛 125105)

通信作者: 许 明, E-mail: xum.2016@tsinghua.org.cn

摘 要: 针对高分辨率遥感图像中地物背景复杂多样, 成像过程中道路区域易受树木、建筑物遮挡影响, 从中提取道路时易出现局部断连和细节缺失问题, 设计并实现了一种基于多尺度差异聚合机制的道路提取网络模型 (MSDANet). 网络模型整体采用编码-解码器结构, 使用 Res2Net 模块作为编码器骨干网络获取细粒度多尺度特征信息, 增大特征提取感受野; 同时结合道路形态特征提出一种门控轴向引导模块, 用于突出道路特征的表达, 改善道路提取长距离断裂现象; 此外, 设计了一种应用于编解码器之间的多尺度差异聚合模块, 用以提取浅层与深层特征间的差异信息并将其聚合, 并通过特征融合模块将聚合特征与解码特征融合, 促进解码器准确还原道路特征; 在高分辨率遥感数据集 DeepGlobe 和 CHN6-CUG 上进行模型实验评估, 所提方法的 $F1$ 值分别为 80.37%、78.17%, IoU 分别为 67.18%、64.17%, 均优于对比模型.

关键词: 道路提取; 遥感影像; 多尺度差异聚合机制; 门控轴向引导; Res2Net

引用格式: 许明, 周春晖, 姜彦吉. 基于多尺度差异聚合机制的遥感影像道路提取. 计算机系统应用, 2024, 33(9):95-104. <http://www.c-s-a.org.cn/1003-3254/9603.html>

Road Extraction from Remote Sensing Image Based on Multi-scale Difference Aggregation Mechanism

XU Ming, ZHOU Chun-Hui, JIANG Yan-Ji

(Software College, Liaoning Technical University, Huludao 125105, China)

Abstract: In the extraction of roads from high-resolution remote sensing images, problems such as local disconnections and the loss of details are common due to the complex backgrounds and the presence of trees and buildings covering the roads during the image formation process. To solve these problems, this study proposes a road extraction model called MSDANet, based on a multi-scale difference aggregation mechanism. The model has an encoder-decoder structure, using the Res2Net module as the backbone network of the encoder to obtain information with fine-grained and multi-scale features from the images and to expand the receptive field for feature extraction. Additionally, a gated axial guidance module, in conjunction with road morphological features, is applied to highlight the representation of road features and improve the connectivity of long-distance roads in road extraction. Furthermore, a multi-scale difference aggregation module is used between the encoder and decoder to extract and aggregate the different information between shallow and deep features. The aggregated features are then fused with the decoded features through a feature fusion module to facilitate the decoder to accurately restore road features. The proposed method has been evaluated on two high-resolution remote sensing datasets: DeepGlobe and CHN6-CUG. The results show that the $F1$ score of the MSDANet model is 80.37% and 78.17% respectively, and the IoU is 67.18% and 64.17% respectively. It indicates that the proposed model outperforms the comparison models.

① 基金项目: 辽宁省教育厅科学技术研究项目 (LJKZ0338)

收稿时间: 2024-03-01; 修改时间: 2024-04-01; 采用时间: 2024-04-10; csa 在线出版时间: 2024-07-26

CNKI 网络首发时间: 2024-07-29

Key words: road extraction; remote sensing image; multi-scale difference aggregation mechanism; gated axial guidance; Res2Net

高分辨率遥感图像包含丰富的道路网络信息,从中高效、准确地提取道路信息对城市规划建设^[1]、公共交通管理^[2]和灾害监测^[3]等具有重要作用.遥感图像中道路具有结构复杂、跨度大和狭窄曲折等特性且易受到建筑阴影、林木和铁路等背景信息的干扰^[4],因此,从高分遥感图像中准确高效地提取道路信息仍是一项极具挑战的任务^[5].

传统道路提取如基于形态学、基于手工特征等方法可有效获取道路形状特征,在遥感影像道路提取领域取得了一定的成果,但这些算法提取过程繁杂,自动化程度较低^[6],当前,以卷积神经网络为代表的深度学习算法因其良好的性能和泛化能力而广泛应用于遥感影像智能化处理研究中^[7].Ronneberger等^[8]提出的U-Net通过对称的上采样-下采样结构,在准确度与处理速度之间取得了平衡,LinkNet^[9]和D-LinkNet^[10]选用残差网络作为编码器主干网络以提高感受野,并将中间特征拼接部分替换成跳跃连接结构以减少参数量,后者加入并行级联的空洞卷积模块,使模型在保留中心特征信息的同时最大化感受野.但遥感影像中存在植被、建筑物阴影遮挡现象时,以卷积为主的类U-Net模型易出现道路提取结果连续性差及细小道路缺失现象,其原因如下:(1)受卷积神经网络归纳偏置影响,类U-Net模型在提取局部有效信息时相对高效,但缺乏对通道维度之间关系的显式建模^[11],提取全局信息时表现较差,导致提取结果存在断裂问题.(2)跳跃连接结构中采用逐元素相加或拼接操作时易造成冗余信息堆叠和关键信息丢失等问题,编码器中不同深度的特征层具有不同性质特征信息,浅层特征包含更多空间位置信息,便于捕获道路的细微边界,深层特征则富含语义信息^[12],而跳跃连接结构使特征传递局限于相同层级间,导致部分空间关联性信息损失,提取复杂道路网络时易出现误提和漏提现象.

针对CNN归纳偏置影响,可通过将卷积神经网络与注意力机制结合以增强全局特征信息提取能力,张亚宁等^[13]将残差网络与CBAM注意力模块^[14]结合使用,促进上下文信息融合,提高全局语义特征提取能力;NL-LinkNet^[15]将Non-Local注意力^[16]模块与残差网络

组合构成编码器主体网络以捕获长程依赖关系,张立恒等^[17]引入双注意力机制以捕获道路信息的全局特征依赖.针对跳跃连接无法充分利用空间位置信息的问题,陈国军等^[18]通过在跳跃连接处加入特征融合模块,提高模型全局特征提取能力.陈果等^[19]通过特征金字塔模块将多尺度下跳跃连接特征进行聚合,为解码器提供更有效的特征信息.

结合以上问题,本文提出一种基于多尺度差异聚合机制的道路提取方法MSDANet (multi-scale difference aggregation network),主要工作如下.

(1)以多尺度残差网络Res2Net^[20]作为编码器主体网络,提取细粒度特征信息,提高网络分割性能.(2)结合道路形态特性,提出一种门控轴向引导模块,引导模型专注于提取横向或纵向道路特征信息.(3)提出一种多尺度差异聚合模块,用以减少跳跃连接结构在特征传递时的空间信息损失.(4)提出一种特征融合模块,融合差异聚合模块输出与解码特征,提升道路提取准确性.

1 网络结构

本文方法整体结构如图1所示,MSDANet网络模型以LinkNet为基础,主要由编码器、门控轴向引导(gated axial guidance, GAG)模块、多尺度差异聚合(multi-scale difference aggregation, MSDA)模块、特征融合(feature fusion, FF)模块和解码器构成.遥感图像输入后不同于传统 7×7 卷积层和最大池化层,本文采用3个 3×3 卷积和SoftPool^[21]下采样层进行特征粗提取,增加感受野的同时减少下采样过程中的信息损失.特征输入Res2Net模块编码模块,得到4个尺度下的细粒度编码特征 E_i ,再经门控轴向引导模块处理后输入多尺度差异聚合模块,得到互补增强特征 K_i ,与对应尺度解码特征 D_i 融合,最后通过上采样模块将特征图恢复至原始尺寸,得到最终道路分割图.

1.1 编码器模块

编码器模块通常用于提取道路细节特征信息和语义信息,多数以残差神经网络ResNet作为编码骨干网络的模型特征表达能力不足,为实现更高效的特征提取,本文采用多尺度深度残差网络Res2Net作为编码

骨干网络, 增加网络层的感受野, 提取更细粒度道路特征.

如图2所示输入特征图经过 1×1 卷积降维后, 将通道平均拆分成 s 个子集即 x_i , 其中 $i \in \{1, \dots, s\}$, 对应输出为 y_i ; 除 x_1 和 x_2 外, 其余子集在经过 3×3 卷积层 K_i 前

均与上一子集输出 y_{i-1} 进行加和操作, 具体表达式如式(1)所示:

$$y_i = \begin{cases} x_i, & i=1 \\ K_i(x_i), & i=2 \\ K_i(x_i + y_{i-1}), & 3 \leq i \leq s \end{cases} \quad (1)$$

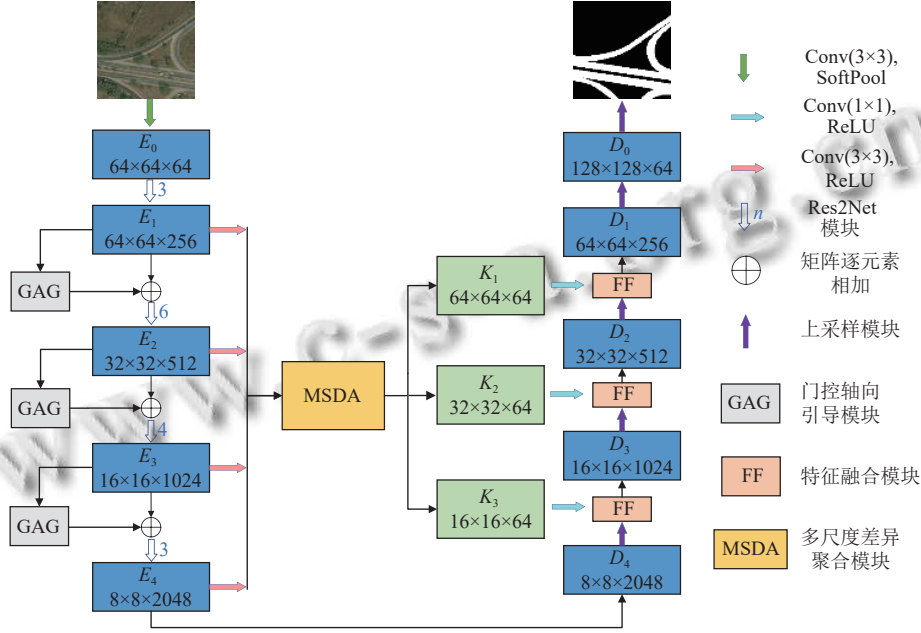


图1 MSDANet 网络整体结构

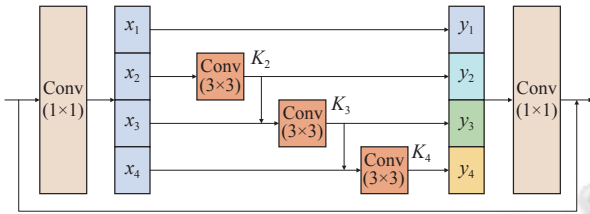


图2 Res2Net 模块

所有中间特征 y_i 拼接后经过 1×1 卷积层升维后得到最终多尺度特征.

Res2Net 沿用了残差网络的恒等映射结构用以缓解网络模型在深度增加时伴随出现的退化问题, 并对中间特征采取先分组, 后分布递进提取的方式, 获取不同信息融合度的子集特征, 最后将其合并, 这种策略有效地控制了计算开销, 并通过不同尺度的上下文信息减少局部信息不足造成的歧义, 提高整体网络模型多尺度表达能力.

1.2 门控轴向引导模块

Res2Net 模块在空间维度上提取特征信息时, 因卷积神经网络归纳偏置影响, 全局特征提取能力不足, 同

时缺少对通道维度相关性的显式建模, 进而限制编码器特征提取能力, 因此本文提出了门控轴向引导模块, 其结构如图3所示, 分为轴向特征提取和轴向权重嵌入两部分, 结合图1结构图可知, 该模块以不同尺度下的编码特征 E_i 作为输入, 轴向特征提取部分以门控轴向注意力^[22]机制作为基础, 用以突出道路形态特征, 轴向权重嵌入部分通过提取通道相关性, 捕获轴向特征映射中重要通道信息, 最后将其嵌入编码特征中, 引导编码器准确提取道路特征信息.

轴向特征提取部分通过头尾 1×1 卷积层控制中间特征图维度, 减少模型参数和计算开销, 通过高宽轴向门控注意力, 计算特征图中两个特定位置间相似关联性并分配相应关注度, 其中, 门控轴向注意力机制为自注意力机制^[23]变体, 具体结构如图4所示, 可解构为两部分即门控机制和轴向注意力, 以沿宽度轴方向为例门控轴向注意力机制, 计算过程如式(2):

$$F_{ij}^G = \sum_{w=1}^W \text{Softmax}(q_{ij}^T k_{iw} + G_Q q_{ij}^T r_{iw}^q + G_k k_{iw}^T r_{iw}^k)(G_1 r_{iw}^v + G_2 v_{iw}) \quad (2)$$

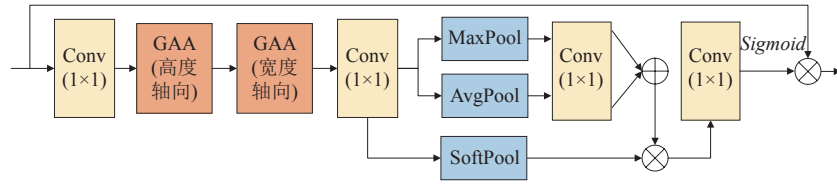


图3 门控轴向引导模块

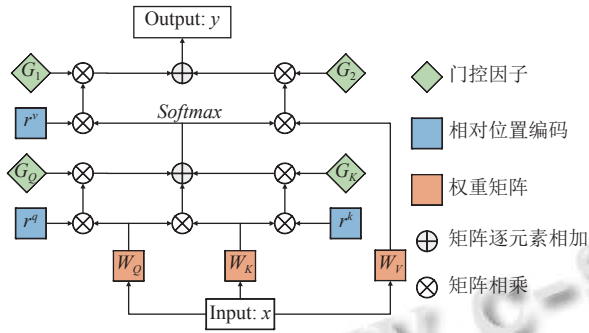
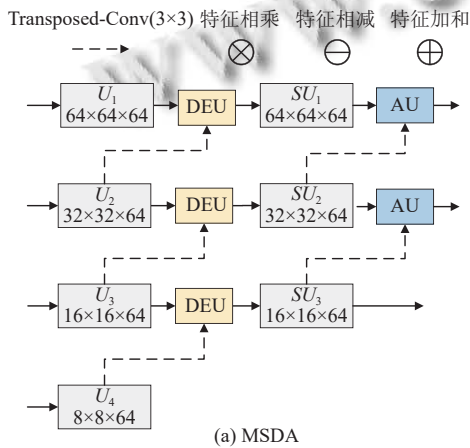


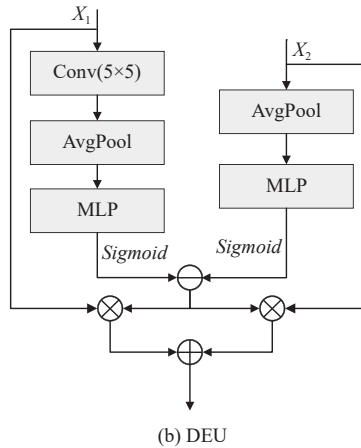
图4 门控轴向注意力

给定输入特征图为 $x \in \mathbb{R}^{C_{in} \times H \times W}$, 其中 C_{in} 代表输入特征图通道数, $F^G \in \mathbb{R}^{C_{out} \times H \times W}$ 表示轴向特征提取部分的输出, C_{out} 代表输出特征图通道数, 查询 $q = W_Q x$ 、键 $k = W_K x$ 和值 $v = W_V x$ 表示输入特征 x 的线性函数映射, W_Q 、 W_K 和 W_V 表示对应映射的可学习权重矩阵, q_{ij} 、 k_{ij} 和 v_{ij} 表示在任意位置的对应矩阵, 其中 $i \in \{1, \dots, H\}$, $j \in \{1, \dots, W\}$; r^d 、 r^k 和 r^v 表示相对位置编码^[24]矩阵, 增强局部模式捕获能力; G_Q 、 G_K 、 G_1 、 G_2 表示可学习门控因子, 门控机制通过引入可学习参数, 对相对位置编码进行准确性判断并赋予相应权重, 控制其带来的负面影响, 若相对位置编码对于道路特征能准确学习, 门控权重相应增大, 反之则减小。

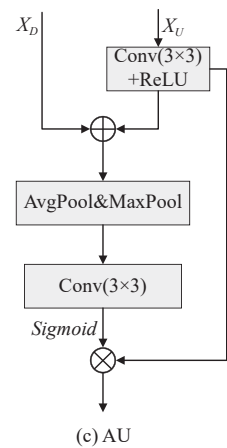
轴向权重嵌入部分将 F^G 分别输入 3 种池化层中,



(a) MSDA



(b) DEU



(c) AU

图5 多尺度差异聚合模块

用以获取更全面的通道相关性信息, 通过平均池化层、最大池化层提取统计特征, 将池化结果分别输入 1×1 卷积层, 两者加和得到 F_{MA}^G , 同时通过 SoftPool 池化分支得到 F_P^G , 计算过程如式 (3) 和式 (4):

$$F_{MA}^G = \sigma(C_1(\text{MaxPool}(F^G))) + \sigma(C_1(\text{AvgPool}(F^G))) \quad (3)$$

$$F_S^G = \sigma(\text{SoftPool}(F^G)) \quad (4)$$

其中, σ 表示 ReLU 函数, $C_1(\cdot)$ 表示 1×1 卷积层. 将 F_{MA}^G 和 F_P^G 相乘结果输入 1×1 卷积层, 经过 Sigmoid 函数 δ 后与编码特征 E_i 进行矩阵乘法, 最终得到 GAG 模块输出 T_i , 计算过程如式 (5):

$$T_i = \delta(C_1(F_P^G \otimes F_{MA}^G)) \otimes E_i \quad (5)$$

1.3 多尺度差异聚合模块

当前以编码-解码结构为基础的语义分割模型, 多采用单一的跳跃连接/拼接操作来传递编码特征, 其过程中忽略了不同尺度下中间特征间的差异性信息, 本文提出一种多尺度差异聚合模块, 通过若干个差异聚合单元在多个尺度下进行差异信息提取, 将深层丰富感受野嵌入到浅层特征中, 同时保留一部分浅层特征的道路信息, 实现浅层单一特征向深层次多元特征转化, 以此输出不同差异融合度的特征, 具体结构如图 5(a) 所示。

结合图1可知编码特征 E_i 分别经过 3×3 降维卷积层得到 U_i ,不同层级特征 U_i 通过差异嵌入单元(difference embedding unit, DEU)得到对应层级中间特征,以 $S U_1$ 计算过程为例,将当前层特征图 U_1 (对应图5(b) X_1)与上采样后的深层特征图 U_2 (对应图5(b) X_2)输入DEU中,各分支计算过程如式(6):

$$\begin{cases} X'_1 = \delta(W_{12}\sigma(W_{11}(\text{AvgPool}(C_5(X_1)))))) \\ X'_2 = \delta(W_{22}\sigma(W_{21}(\text{AvgPool}(X_2)))) \end{cases} \quad (6)$$

其中, X'_1 和 X'_2 表示对应分支权重向量, W_{11} 、 W_{12} 、 W_{21} 和 W_{22} 表示全连接层.不同于通道或空间注意力机制,将权重向量与输入特征相乘,DEU首先采用减法操作,提取两条分支即浅层与深层特征间差异权重向量 P ,并将其嵌入到对应输入,将两者加和,最终输入到 3×3 卷积和ReLU层得到差异嵌入单元输出 $S U_1$,计算过程如式(7)和式(8):

$$P = |(X'_1 \ominus X'_2)| \quad (7)$$

$$S U_1 = \sigma(C_3(X_1 \otimes P + X_2 \otimes P)) \quad (8)$$

其中, $\ominus$ 表示逐元素减法运算, $||$ 表示绝对值计算, $C_3(\cdot)$ 表示 3×3 卷积操作,将不同层级 $S U_i$ 输入聚合单元(aggregation unit, AU)进行加和,如图5(c)所示, X_D 表示通过转置卷积上采样后的深层差异特征, X_U 表示浅层差异特征,计算过程如式(9)和式(10):

$$\begin{cases} X_F = \sigma(C_3(X_U)) + X_D \\ X'_F = c(\text{MaxPool}(X_F), \text{AvgPool}(X_F)) \end{cases} \quad (9)$$

$$K = \text{Sigmoid}(C_3(X'_F)) \otimes \sigma(C_3(X_U)) \quad (10)$$

其中, $c(\cdot, \cdot)$ 表示特征拼接,将不同层级差异特征加和,通过聚合单元抑制融合特征中无关区域的噪声干扰,最终得到互补增强特征 K_i .

1.4 特征融合模块

解码特征通常与中间传递特征直接加和或拼接,此类融合方式忽略了二者间感受野的差异,使解码过程中特征重建恢复程度有限,易导致浅层解码器难以优化,为促进解码器准确还原道路细节,本文提出一种特征融合模块,用于融合解码特征 D_i 与互补增强特征 K_i .

如图6所示,通过不同尺度卷积层分支,提高互补增强特征 K_i 感受野,并与解码特征 D_i 进行拼接,各分支拼接后进行加和融合,将融合特征通过 1×1 卷积层对齐特征维度,最终将输出特征与 K_i 进行残差连接,计算

过程如式(11)和式(12):

$$M'_i = c(K_i, D_i) + c(S(C_3(K_i)), D_i) + c(S(C_5(K_i)), D_i) \quad (11)$$

$$D'_i = C_1(M'_i) + K_i \quad (12)$$

其中, S 表示维度调整卷积层,用于减少 3×3 和 5×5 卷积分支计算开销,采用 1×1 卷积调整特征维度.

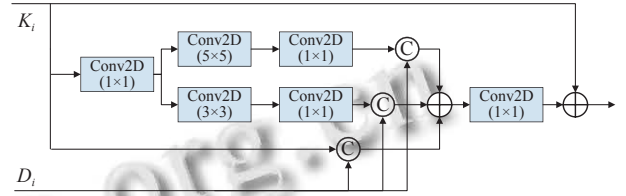


图6 特征融合模块

2 实验设置与评价指标

2.1 实验数据集

为综合评估MSDANet的有效性,本文在公开数据集DeepGlobe^[25]和CHN6-CUG^[26]上进行模型性能评估.

DeepGlobe公开道路数据集是2018年道路提取挑战赛提出的一组高分辨率遥感图像,每张影像尺寸为 1024×1024 像素,其地面分辨率为 0.5 m/像素 ,标签为二值化图像,其道路像素值为1,背景像素值为0. DeepGlobe数据集包含6226张带有标注数据的图像,本文从该数据集中随机选取4696张图像用于训练,1530张图像用于测试,并将所有图像以256步长和512步长进行无重叠裁剪,由此得到两组不同像素尺寸大小的训练集和测试集.

CHN6-CUG公开道路数据集是中国首套城市大尺度卫星遥感影像道路数据集,包含4511张 512×512 像素的遥感影像,其地面分辨率为 0.5 m/像素 ,其中训练集为3608张,测试集为903张.该数据集存在部分无效影像,去除后得到3401张训练集和505张测试集,并将其无重叠裁剪为 256×256 像素大小影像,最终得到12164张训练集图像和2020张测试集图像以及对应数量的标签.

2.2 实验环境与参数设置

本文实验环境为Ubuntu操作系统, GPU为24 GB显存的NVIDIA GeForce RTX 3090,采用PyTorch作为深度学习训练框架.训练过程中,DeepGlobe数据集的初始学习率设置为0.001,批处理大小为32, CHN6-

CUG 数据集初始学习率设置为 0.000 5, 批处理大小为 16, 迭代周期均为 80, 采用 Adamw^[27] 优化函数. 为了更好地优化网络的训练过程, 采用二分类交叉熵 BCE (binary cross entropy) 损失函数和 Dice 损失函数共同作为损失函数.

2.3 评价指标

实验指标准确率 (Precision)、召回率 (Recall)、F1 值和交并比 (IoU) 可有效对模型性能进行评估. 本文选用准确率、F1 值和 IoU 作为主要评价指标. F1 值表示准确率和召回率的调和均值; IoU 表示预测值和真实值两个集合的交集与并集之比. TP 表示正确预测为道路的像素, FP 表示将背景错误预测为道路的像素, TN 表示正确预测为背景的像素, FN 表示将道路错误预测为背景的像素.

3 实验分析

3.1 DeepGlobe 数据集实验分析

为综合分析模型道路提取能力, 在相同的实验环境下使用 256×256 像素数据集 I 和 512×512 像素数据集 II 对各模型进行训练, 通过数据集 I 可视化结果评估模型局部小范围提取精准性, 通过数据集 II 可视化结果评估模型较大范围场景下道路提取能力, 如图 7 和图 8 所示为部分道路提取结果, 红色表示正确提取区域, 绿色表示遗漏提取区域, 蓝色表示错误提取区域.

3.1.1 数据集 I 实验结果分析

由表 1 可知本文模型在 F1 值、IoU 和准确率上提升明显, 相较于对比模型方法中最高数据, 分别提升 1.19%、1.64% 和 1.13%, 相较于基线模型 LinkNet 分别提升 1.50%、2.07% 和 3.05%.

表 1 不同模型在数据集 I 的道路提取结果对比 (%)			
模型	准确率	F1值	IoU
MSDANet	81.72	80.37	67.18
D-LinkNet	78.99	79.18	65.54
LinkNet	78.67	78.87	65.11
U-Net	77.28	73.92	58.63
DeepLabv3+	80.09	79.16	65.51
NL-LinkNet	80.59	79.04	65.34

如图 7 所示, (a) 组标注区域因光谱纹理不明显, U-Net、LinkNet 和 D-LinkNet 均无法识别特征不显著道路区域而造成漏提现象, DeepLabv3+ 和 NL-LinkNet 可提取到部分道路但存在断连现象, MSDANet 通过多尺度编码器模块更好地提取了道路细节和轮廓特征, GAG 模块使提取结果更加连贯. (b) 组道路区域存在林木遮挡现象, 从目视解译角度标注区域仍与主干道有明显连结地带, LinkNet、D-LinkNet 和 DeepLabv3+ 仅能提取到该区域部分道路, 本文方法提取结果相对准确完整, 体现了模型在遮蔽环境下的抗干扰能力; (c) 组存在建筑阴影遮蔽现象, 由提取结果可知 MSDANet 预测结果相对平滑完整, 道路交叉口区域轮廓完整清晰, 但仍存在漏提问题.

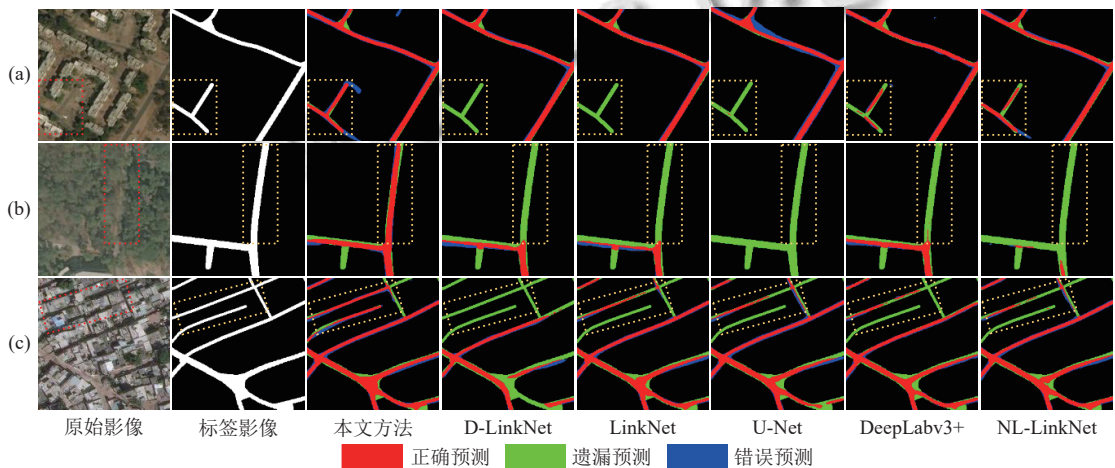


图 7 DeepGlobe 数据集 I 分割结果

3.1.2 数据集 II 实验结果分析

由表 2 可知本文模型在 F1 值、IoU 上提升明显, 相

较于对比模型方法中最高数据, 分别提升 0.78% 和 1.11%, 相较于基线模型 LinkNet 分别提升 1.17% 和 1.65%.

表 2 不同模型在数据集 II 的道路提取结果对比 (%)

模型	准确率	F1值	IoU
MSDANet	81.90	81.12	68.24
D-LinkNet	78.17	80.10	66.80
LinkNet	81.30	79.95	66.59
U-Net	78.44	73.56	58.18
DeepLabv3+	79.41	79.75	66.32
NL-LinkNet	80.89	80.34	67.13

如图 8 所示, 本文模型在大范围场景下提取结果连贯完整, 整体道路网络断裂现象较少, 局部道路细节

还原较好, 道路交汇和连接处的提取相对准确, 综合数据集 I 提取结果分析, 相较于其他对比模型, 本文方法提高了道路提取结果的整体效果.

3.1.3 交叉验证实验结果分析

为评估 MSDANet 泛化能力, 本文使用数据集 I 进行 4 折交叉验证, 将数据集随机分为 4 份, 按 3:1 的比例分为训练集和测试集, 每一折实验环境均与数据集 I 各模型训练测试环境相同.

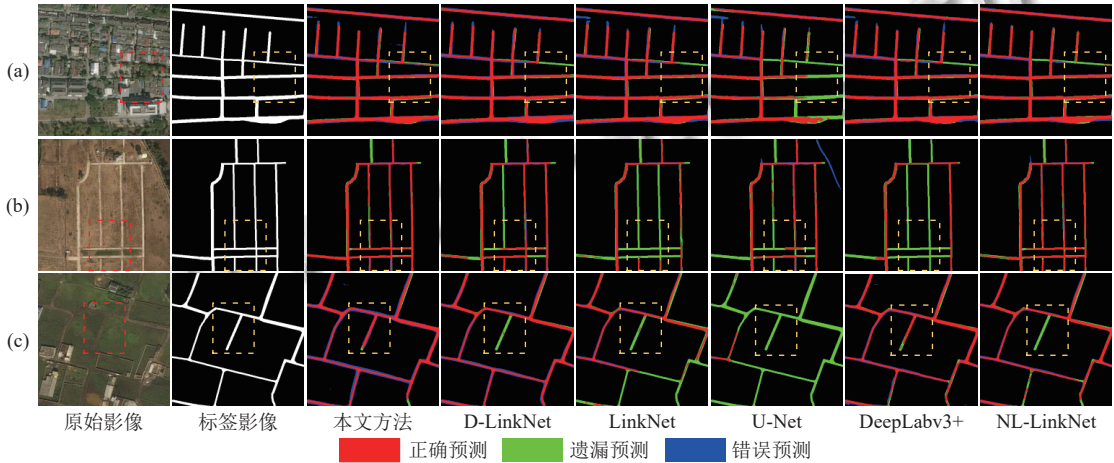


图 8 DeepGlobe 数据集 II 分割结果

由表 3 中 4 折交叉验证结果可知, 在 3 项指标上 4 折结果中的 3 折略低于数据集 I, 由平均值可知各项评价指标整体与数据集 I 结果相近, 因此本文模型具有较强泛化能力.

表 3 交叉验证结果对比 (%)

折数	准确率	F1值	IoU
第1折	81.22	80.35	67.15
第2折	80.41	80.03	66.71
第3折	81.76	80.51	67.37
第4折	81.39	80.22	66.98
平均值	81.20	80.28	67.05

3.2 CHN6-CUG 数据集实验结果分析

为进一步验证模型在其他数据集上的泛化能力, 选用 CHN6-CUG 数据集进行测试, 所有实验均在相同实验环境下训练和评估, 部分提取结果如图 9 所示.

由表 4 可知模型在 F1 值和 IoU 两项指标上均有提升, 相较于对比方法中最高数据分别提升 0.77%、1.04%, 相较于基线方法 LinkNet 分别提升 0.91%、1.22%, 准确率略高于基线方法.

如图 9 所示, (a) 组提取结果体现了本文模型能从

复杂背景下辨识不同拓扑结构的道路网络, 标注区域背景复杂, 仅 MSDANet 和 NL-LinkNet 能完整提取; (b) 组标注区域道路交汇连接处较窄, 所有模型均有漏提现象, MSDANet 预测结果相对完整, 进一步体现本文方法在面对复杂路段时有较好的提取能力; (c) 组整体道路区域清晰, U-Net、LinkNet 和 NL-LinkNet 因采用跳跃拼接/连接结构将单一的中间特征信息传递到解码器, 使高权重特征信息不断叠合, 部分低权重关键特征信息被稀释, 导致部分与背景颜色、纹理相似的道路区域易出现漏提和细节缺失现象, 本文方法通过 MSDA 模块获取更多上下文信息, 使模型对道路特征提取更加准确. 综合提取结果可知, 本文模型泛化能力以及对于不规则道路提取能力优于对比模型.

3.3 消融实验

本文通过 DeepGlobe 数据集进行消融实验, 验证和分析 GAG 模块、MSDA 模块以及 FF 模块对 MSDANet 的性能影响, 所有消融实验均采用相同训练和测试环境, 消融实验定量分析如表 3 所示. 其中, No.1 为基线模型 LinkNet 其编码器主干网络为 ResNet, No.2-No.4

以 LinkNet 为基础分别加入对应模块, No.5 将主干网络替换为 Res2Net, No.6–No.11 以 No.5 为基础加入对应模块以及模块组合。

如表 5 对 FF 模块消融实验定量分析所示, ResNet 对比组中的 No.2 相较于 No.1 在准确率、F1 值和 IoU 上分别提升 2.15%、0.5% 和 0.68%, Res2Net

对比组中 No.6 相较于 No.5 分别提升 0.74%、0.36% 和 0.46%, 通过对比可知, 因主干网络替换为 Res2Net 后整体精度提升, Res2Net 对比组中整体提升幅度降低, FF 模块与 GAG 模块和 MSDA 模块组合使用时有助于促进中间传递特征与解码特征融合, 进一步提升模型提取准确性。

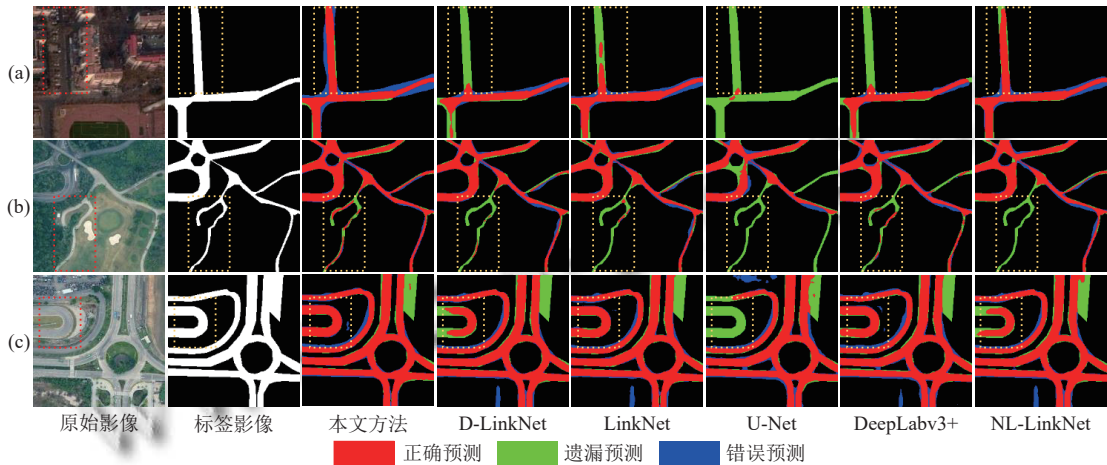


图9 CHN6-CUG数据集分割结果

表4 不同模型在 CHN6-CUG 数据集的道路提取结果对比 (%)

模型	准确率	F1值	IoU
MSDANet	79.69	78.17	64.17
D-LinkNet	80.84	77.06	62.67
LinkNet	79.44	77.26	62.95
U-Net	77.73	70.39	54.31
DeepLabv3+	78.47	77.13	62.77
NL-LinkNet	78.91	77.40	63.13

表5 消融实验结果对比

编号 (No.)	骨干网络	FF 模块	GAG 模块	MSDA 模块	准确率 (%)	F1值 (%)	IoU (%)
1	ResNet	—	—	—	78.67	78.87	65.11
2	ResNet	√	—	—	80.82	79.37	65.79
3	ResNet	—	√	—	80.93	79.87	66.49
4	ResNet	—	—	√	80.52	79.68	66.22
5	Res2Net	—	—	—	80.17	79.49	65.96
6	Res2Net	√	—	—	80.91	79.82	66.42
7	Res2Net	—	√	—	81.29	80.03	66.71
8	Res2Net	—	—	√	81.16	80.14	66.85
9	Res2Net	√	√	—	80.77	80.24	67.00
10	Res2Net	√	—	√	81.23	80.28	67.06
11	Res2Net	—	√	√	81.36	80.20	66.94

通过对 GAG 模块消融实验定量分析可知, ResNet 对比组中的 No.3 相较于 No.1 在准确率、F1 值和 IoU 上分别提升 2.26%、1.0% 和 1.38%, Res2Net 对比组

中 No.7 相较于 No.5 分别提升 1.12%、0.54% 和 0.75%, 通过对比可知, 该模块在道路分割任务中表现较好; GAG 模块通过突出道路形态特征, 引导模型提取道路元素间的关联关系, 因此应用于不同骨干网络时均提升明显, 适用于道路提取任务。

通过对 MSDA 模块消融实验定量分析可知, ResNet 对比组中的 No.4 相较于 No.1 在准确率、F1 值和 IoU 上分别提升 1.85%、0.81% 和 1.11%, Res2Net 对比组中 No.8 相较于 No.5 分别提升 0.99%、0.65% 和 0.89%, 各指标在主干网络更换后提升幅度略有降低, 进一步对比 No.8 与 No.10、No.11 可知, 将 FF 模块和 GAG 模块与 MSDA 模块组合使用时 F1 值和 IoU 均略有提升, 体现了三者组合使用具有一定协调性, 共同促进模型提取能力进一步提升; 综合上述分析可知, FF、GAG 和 MSDA 模块均可有效增强网络模型表征能力和特征提取融合能力。

4 结论

本文针对高分遥感图像中, 因背景复杂、道路拓扑结构多变等因素, 提取结果准确性低、连续性差问题, 提出一种基于多尺度差异聚合机制的道路提取网络 MSDANet, 采用 Res2Net 作为编码骨干网络, 通过

门控轴向引导模块引导编码器道路形态特征, 利用多尺度差异聚合模块提取跨层级差异特征, 最终通过特征融合模块将互补增强特征与解码特征融合. 为评估网络模型性能, 在 DeepGlobe 和 CHN6-CUG 数据集上设置了定量分析实验, 实验结果表明, MSDANet 整体优于对比模型, 有效改善了错分和漏分现象. 未来研究工作将从以下方面展开: ① 优化模型算法, 进一步平衡模型参数数量和分割精准度. ② 扩大应用范围, 使其应用于其他地物提取任务中.

参考文献

- 赫晓慧, 李代栋, 李盼乐, 等. 基于 EDRNet 模型的高分辨率遥感影像道路提取. 计算机工程, 2021, 47(9): 297–303, 312.
- 戴激光, 王杨, 杜阳, 等. 光学遥感影像道路提取的方法综述. 遥感学报, 2020, 24(7): 804–823.
- Geng K, Sun X, Yan ZY, *et al.* Topological space knowledge distillation for compact road extraction in optical remote sensing images. Remote Sensing, 2020, 12(19): 3175. [doi: [10.3390/rs12193175](https://doi.org/10.3390/rs12193175)]
- Sarukkai V, Jain A, Uzkent B, *et al.* Cloud removal in satellite images using spatiotemporal generative networks. Proceedings of the 2020 IEEE Winter Conference on Applications of Computer Vision. Snowmass: IEEE, 2020. 1785–1794.
- Dai L, Zhang GY, Zhang RT. RADANet: Road augmented deformable attention network for road extraction from complex high-resolution remote-sensing images. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 1–13.
- Chen ZY, Deng LA, Lu YH, *et al.* Road extraction in remote sensing data: A survey. International Journal of Applied Earth Observation and Geoinformation, 2022, 112: 102833. [doi: [10.1016/j.jag.2022.102833](https://doi.org/10.1016/j.jag.2022.102833)]
- 余佐明, 申勇智, 宋剑虹, 等. 利用经典 CNN 网络方法构建贵阳市道路要素遥感影像自动提取模型. 测绘通报, 2023(4): 177–182.
- Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. Proceedings of the 18th International Conference on Medical Image Computing and Computer-assisted Intervention. Munich: Springer, 2015. 234–241.
- Chaurasia A, Culurciello E. LinkNet: Exploiting encoder representations for efficient semantic segmentation. Proceedings of the 2017 IEEE Visual Communications and Image Processing. St. Petersburg: IEEE, 2017. 1–4.
- Zhou LC, Zhang C, Wu M. D-LinkNet: LinkNet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Salt Lake City: IEEE, 2018. 192–1924.
- He X, Zhou Y, Zhao JQ, *et al.* Swin Transformer embedding UNet for remote sensing image semantic segmentation. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 4408715.
- Zhong GQ, Ling X, Wang LN. From shallow feature learning to deep learning: Benefits from the width and depth of deep architectures. WIREs Data Mining and Knowledge Discovery, 2019, 9(1): e1255. [doi: [10.1002/widm.1255](https://doi.org/10.1002/widm.1255)]
- 张亚宁, 张春亢, 王朝, 等. 融合残差和卷积注意力机制的 U-Net 网络高分影像道路提取. 航天返回与遥感, 2023, 44(3): 119–132. [doi: [10.3969/j.issn.1009-8518.2023.03.013](https://doi.org/10.3969/j.issn.1009-8518.2023.03.013)]
- Hu J, Shen L, Sun G. Squeeze-and-excitation networks. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 7132–7141.
- Wang Y, Seo J, Jeon T. NL-LinkNet: Toward lighter but more accurate road extraction with nonlocal operations. IEEE Geoscience and Remote Sensing Letters, 2022, 19: 3000105.
- Wang XL, Girshick R, Gupta A, *et al.* Non-local neural networks. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 7794–7803.
- 张立恒, 王浩, 薛博维, 等. 基于改进 D-LinkNet 模型的高分遥感影像道路提取研究. 计算机工程, 2021, 47(9): 288–296.
- 陈国军, 朱燕宁, 耿润田, 等. 基于 iHDODC-LinkNet 网络的遥感图像道路提取方法. 光电子·激光, 2024, 35(1): 51–58.
- 陈果, 胡立坤. 基于边缘引导和多尺度感知的遥感图像道路提取. 激光杂志, 1–6. <http://kns.cnki.net/kcms/detail/50.1085.TN.20231103.1339.004.html>. [2024-02-29].
- Gao SH, Cheng MM, Zhao K, *et al.* Res2Net: A new multi-scale backbone architecture. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(2): 652–662. [doi: [10.1109/TPAMI.2019.2938758](https://doi.org/10.1109/TPAMI.2019.2938758)]
- Stergiou A, Poppe R, Kalliatakis G. Refining activation downsampling with SoftPool. Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal: IEEE, 2021. 10337–10346.

- 22 Valanarasu MJ, Oza P, Hacıhaliloglu I, *et al.* Medical Transformer: Gated axial-attention for medical image segmentation. Proceedings of the 24th International Conference on Medical Image Computing and Computer-assisted Intervention. Strasbourg: Springer, 2021. 36–46.
- 23 Vaswani A, Shazeer N, Parmar N, *et al.* Attention is all you need. Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: Curran Associates Inc., 2017. 6000–6010.
- 24 Shaw P, Uszkoreit J, Vaswani A. Self-attention with relative position representations. Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. New Orleans: ACL, 2018. 464–468.
- 25 Demir I, Koperski K, Lindenbaum D, *et al.* DeepGlobe 2018: A challenge to parse the earth through satellite images. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Salt Lake City: IEEE, 2018. 172–17209.
- 26 Zhu QQ, Zhang YN, Li ZQ, *et al.* Oil spill contextual and boundary-supervised detection network based on marine SAR images. IEEE Transactions on Geoscience and Remote Sensing, 2021, 60: 5213910.
- 27 Loshchilov I, Hutter F. Decoupled weight decay regularization. Proceedings of the 7th International Conference on Learning Representations. New Orleans: OpenReview.net, 2019. 1–19.

(校对责编: 孙君艳)