

融合注意力机制的 SegFormer 遥感影像道路识别^①



王晓杰¹, 陈少康¹, 闫皓炜², 杨鹤猛², 燕正亮², 王森²

¹(国网福建省电力有限公司 电力科学研究院, 福州 350007)

²(天津航天中为数据系统科技有限公司, 天津 300450)

通信作者: 闫皓炜, E-mail: 956681045@qq.com

摘要: 道路信息在遥感影像中具有重要意义和价值, 因此准确提取道路对于许多应用非常关键. 然而, 在进行道路识别时存在两个主要问题. 首先, 卫星影像的背景复杂多变, 而道路的形态也是复杂多样的, 这给道路的自动识别带来了挑战. 其次, 道路像素只占整个影像的很小一部分, 导致类别不平衡的问题. 为了解决上述问题, 本文提出了一种基于改进的 SegFormer 模型的卫星影像道路自动识别算法. 该算法采用了两个主要策略来改进识别效果. 第一, 在 SegFormer 编码器的各个阶段的输出端添加了空间注意力模块. 这个模块有助于减弱复杂背景的干扰, 同时增强对道路区域的关注. 通过引入空间注意力机制, 模型能够更好地捕捉到道路的特征, 从而提高识别准确性. 第二, 采用了一种混合损失函数, 结合了像素对比损失和交叉熵损失. 这样的损失函数能够更好地处理类别不平衡的问题, 使得模型更加关注道路类别的训练. 通过优化训练过程, 模型能够更好地学习到道路的特征表示, 从而提升识别准确率. 通过对比实验分析, 改进后的模型在测试集上的 mIoU 指标提升了约 3.3%.

关键词: 深度学习; 道路; SegFormer; 空间注意力; 像素对比损失函数

引用格式: 王晓杰, 陈少康, 闫皓炜, 杨鹤猛, 燕正亮, 王森. 融合注意力机制的 SegFormer 遥感影像道路识别. 计算机系统应用, 2024, 33(11): 186-193. <http://www.c-s-a.org.cn/1003-3254/9641.html>

Road Recognition in Remote Sensing Images Using SegFormer Fused with Attention Mechanism

WANG Xiao-Jie¹, CHEN Shao-Kang¹, YAN Hao-Wei², YANG He-Meng², YAN Zheng-Liang², WANG Sen²

¹(Electric Power Research Institute, State Grid Fujian Electric Power Co. Ltd., Fuzhou 350007, China)

²(Tianjin Zhongwei Aerospace Data System Technology Co. Ltd., Tianjin 300450, China)

Abstract: Road information is of great significance and value in remote sensing images, and thus the accurate extraction of roads is crucial for many applications. However, there are two main challenges in road recognition. Firstly, the background of satellite images is complex and diverse, while the morphology of roads is also complex and diverse, which poses a challenge to automatic road recognition. Secondly, road pixels only account for a small portion of the entire image, leading to class imbalance. To address these challenges, this study proposes an automatic road recognition algorithm based on an improved SegFormer model. The algorithm employs two main strategies to improve the recognition performance. Firstly, spatial attention modules are added to the output of each stage of the SegFormer encoder. This module helps to weaken the interference from complex backgrounds and enhance the attention to road areas. By introducing spatial attention mechanisms, the model can better capture the features of roads, thereby improving recognition accuracy. Secondly, a hybrid loss function that combines pixel contrast loss and cross-entropy loss is used. Such a loss function can better handle class imbalance problems and make the model place more focus on training road categories. By optimizing the training process, the model can better learn road feature representation, thereby improving recognition accuracy. Comparative experimental analysis shows that the improved model achieves an approximate 3.3%

① 基金项目: 国网福建省电力有限公司科技项目 (52130423000X)

收稿时间: 2024-04-02; 修改时间: 2024-04-29; 采用时间: 2024-05-09; csa 在线出版时间: 2024-09-24

CNKI 网络首发时间: 2024-09-25

improvement in the mIoU metric on the test set.

Key words: deep learning; road; SegFormer; spatial attention; pixel contrast loss function

1 引言

准确提取道路信息是遥感影像处理的重要任务之一,因为道路作为遥感影像中的基本要素,它的提取对于研究城市规划、交通管理等方面具有重要的意义和价值。

以往对卫星影像的地物目标进行识别多采用传统的图像处理方法,但传统图像处理方法识别精度低,泛化能力差。近年来,深度学习方法的兴起推动了地物目标识别研究的发展,成为该领域的主要方向。利用深度学习方法进行地物目标识别,已成为当前地学遥感领域热门研究课题之一。党宇等^[1]将深度学习 AlexNet 模型应用于遥感影像地表覆盖分类评价。针对青藏高原积雪识别率低的问题,阚希等^[2]提出了一种将多要素地理信息与深度学习模型特征提取相结合的方法。该方法减弱了云层对识别的干扰,提升了识别精度。Li 等^[3]在深度学习模型的基础之上进行改进,首先训练深度学习模型来对图像进行像素级分类,之后再通过线性积分卷积方法来精化道路结构。Cheng 等^[4]通过级联两个卷积神经网络来同时提取道路图像和道路中心线图像。具体而言,第 1 层级用于道路像素检测,而第 2 层级则基于第 1 层级得到的特征图来输出道路中心线。上述深度学习方法非端到端模式,处理过程较复杂,识别精度有待提高。端到端模式得到更好的发展,Zhong 等^[5]采用全卷积神经网络 FCN-4s 模型来提取道路。袁翠霞等^[6]通过 U-Net 网络来进行农田道路识别。Chen 等^[7]提出了一种用于遥感图像中道路提取的重建偏差 U-Net (RB-U-Net) 模型。该方法增加了解码分支,以从不同上采样中获取多个语义信息。李代栋等^[8]提出了 SPUD-ResNet 模型用于遥感影像道路提取。该方法通过构建空洞卷积残差网络编码器,并连接到对应解码器结构,同时优化池化模块,有效提升了道路提取的精度。胡绍凯等^[9]提出了 MLUM-Net 模型用于遥感影像的地物分类。该方法使用多尺度空洞卷积和通道注意力机制来进行编码器的特征提取,同时对解码器的上采样模块进行了优化。在训练过程中,采用混合损失函数来提高性能,并在多个数据集上进行了测试,结果显示精度有明显提升。

近两年,基于 Transformer 网络的模型被提出,Transformer 相比卷积神经网络,更能够捕捉长距离依赖关系,使得提取的特征更具有有效性。ViT 模型的提出第 1 次将 Transformer 应用于图像领域,并取得了非常好的效果,此后很多研究者,尝试将 Transformer 应用于自己的研究中。Wang 等^[10]采用编码器-解码器架构来构建语义分割模型,将 Swin Transformer 作为主干网络来提取上下文信息,并设计了密集连接的特征聚合模块的新型解码器来恢复分辨率并生成分割图。Gao 等^[11]提出了 STransFuse 模型,该模型将 Transformer 和 CNN 融合到一起。一方面,STransFuse 模型利用 Transformer 网络在遥感图像中建模全局语义信息的能力。另一方面,它利用具有预训练权重的 CNN 解决 Transformer 网络在遥感图像数据量较小时性能较差的问题。Zheng 等^[12]提出了 SETR 模型,该模型为一个纯 Transformer 结构来将图像编码为一系列图块。由于全局上下文在 Transformer 的每一层中都被建模,因此可以将这个编码器与一个简单的解码器组合起来,提供强大的分割能力。田雪伟等^[13]提出一种改进的 SegFormer 网络^[12],该网络通过级联方式将 SegFormer 编码器输出的多个尺度的特征进行合并。

总体来说,Transformer 网络更加关注目标的结构信息和全局信息,而 CNN 网络更加关注目标的纹理信息和局部信息。道路的纹理信息较弱,而呈现出狭长的连续结构,因此考虑采用基于 Transformer 的网络模型。SegFormer 模型编码器由多层次的 Transformer block 结构组成,能够输出多尺度的特征表示,是一种端到端、简洁、高效且具有较强鲁棒性的语义分割算法。因此本文考虑将其应用于遥感影像的道路识别,但卫星影像的背景复杂多变,而道路的形态也是复杂多样的。其次,道路像素只占整个影像的很小一部分,导致类别不平衡的问题。使得该模型在遥感影像道路识别的准确率有待进一步提升。为此,采用了两个主要策略来改进识别效果。第一,在 SegFormer 编码器的各个阶段的输出端添加了空间自注意力模块^[14]。这个模块有助于减弱复杂背景的干扰,同时增强对道路区域的关注。通过引入空间注意力机制,模型能够

更好地捕捉到道路的特征,从而提高识别准确性. 第二,采用了一种混合损失函数,结合了像素对比损失^[15]和交叉熵损失. 这样的损失函数能够更好地处理类别不平衡的问题.

2 改进的 SegFormer 模型

2.1 SegFormer 模型

SegFormer 模型是一种端到端、简洁、高效且具有较强鲁棒性的语义分割算法(如图1). 它采用了编

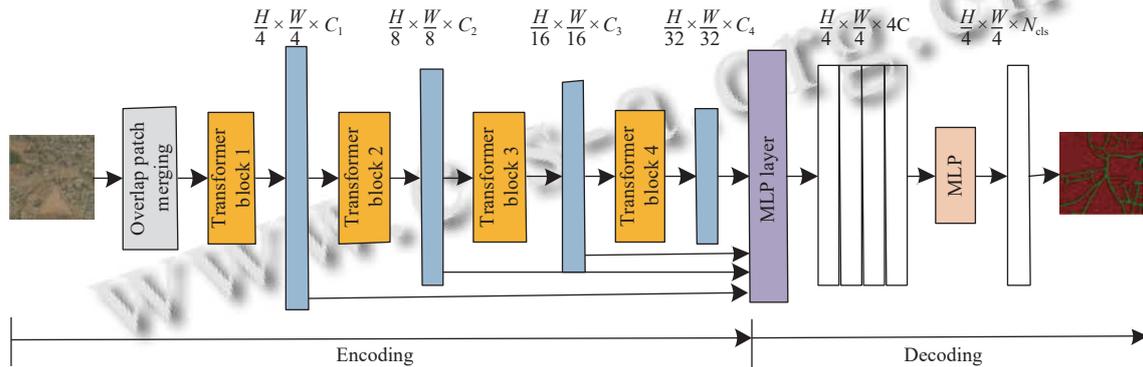


图1 SegFormer 语义分割模型

2.2 融合注意力机制的 SegFormer 模型编码器

为抑制卫星影像中复杂背景的干扰,同时提高模型对影像中道路特征的关注度,提出将卷积注意力模块(convolutional block attention module, CBAM)^[16]应用于 SegFormer 的编码器网络中,称为 MitB3-CBAM,如图2所示,以获取更加具有代表性的高级语义特征. SegFormer 编码器各个 stage 输出的特征层添加 CBAM 模块,以此来提升特征的表达能力. CBAM 模块由两部分构成:通道注意力模块和空间注意力模块.

通道注意力模块更加关注哪个通道特征对于识别道路更具有价值. 对于输入特征 F , 经过通道注意力模块之后可以得到 F' . 具体计算方式如式(1)所示:

$$F' = M_C(F) \otimes F \quad (1)$$

其中, $M_C(F)$ 代表了各个通道的注意力权重, \otimes 代表 F 中的特征通道与 $M_C(F)$ 中各个通道的注意力权重相乘, 权重更高的特征通道, 将在识别道路时发挥更重要的作用. $M_C(F)$ 计算方式如式(2)所示:

$$M_C = \sigma(\text{MLP}(\text{avgpool}(F)) + \text{MLP}(\text{maxpool}(F))) \quad (2)$$

而空间注意力模块更加关注道路本身, 这有助于抑制卫星影像复杂的背景. 对于输入特征 F' , 经过空间

码-解码的结构. 编码器部分由多层次的 Transformer block 结构组成, 能够输出多尺度的特征表示. SegFormer 模型不需要使用位置编码, 从而避免了在测试分辨率与训练不同时需要插值位置编码导致性能下降的问题. 解码器部分由多层感知器(multilayer perceptron, MLP)组成, 并且设计简单. 它通过聚合来自不同层的信息, 同时结合了局部关注和全局关注机制, 以生成强大的特征表示. 这种设计能够有效地提取图像中的语义信息, 并为语义分割任务提供精确的预测结果.

注意力模块之后可以得到 F'' . 具体计算方式如式(3):

$$F'' = M_S(F') \otimes F' \quad (3)$$

其中, $M_S(F')$ 代表了通道中各个位置元素的注意力权重, \otimes 代表 F'' 中的各个位置上的特征元素与 $M_S(F')$ 中各位置上的特征元素的注意力权重相乘, 权重更高的位置特征, 将更有利于定位卫星影像中的道路位置. $M_S(F')$ 计算方式如式(4)所示:

$$M_S = \sigma(f([\text{avgpool}(F'); \text{maxpool}(F')])) \quad (4)$$

2.3 混合损失函数

SegFormer 采用标准的交叉熵损失函数. 然而, 在卫星影像中, 道路像素仅占整张图像的很小比例, 导致类别严重不平衡. 由于这种情况, 仅使用标准交叉熵损失函数难以有效解决问题. 因此, 在训练过程中, 本文引入了像素对比损失函数作为辅助函数, 以优化类别不平衡问题. 在解码器部分添加了一个像素对比损失函数分支, 该分支只在训练过程中使用. 网络结构如图3所示. 该分支包含了两个 1×1 卷积层和一个 ReLU 层.

交叉熵损失函数如式(5)所示:

$$L_i^{ce} = -\log \frac{\exp(y_c)}{\sum_{c'=1}^{|C|} \exp(y_{c'})} \quad (5)$$

像素对比损失函数如式(6)所示:

$$L_i^{nce} = \frac{1}{|P_i|} \sum_{i^+ \in P_i} -\log \frac{\exp(i \cdot i^+ / \tau)}{\exp(i \cdot i^+ / \tau) + \sum_{i^- \in N_i} \exp(i \cdot i^- / \tau)} \quad (6)$$

其中, i 表示每个像素归一化的特征向量, P_i 和 N_i 分别

表示像素 i 的正样本和负样本集合. τ 为超参数, 本实验设定 $\tau=0.1$, i^+ 为正样本特征向量, i^- 为负样本特征向量.

本文的训练损失函数如式(7)所示:

$$L = \sum_i (L_i^{ce} + \lambda L_i^{nce}) \quad (7)$$

其中, 本文 $\lambda = 1$.

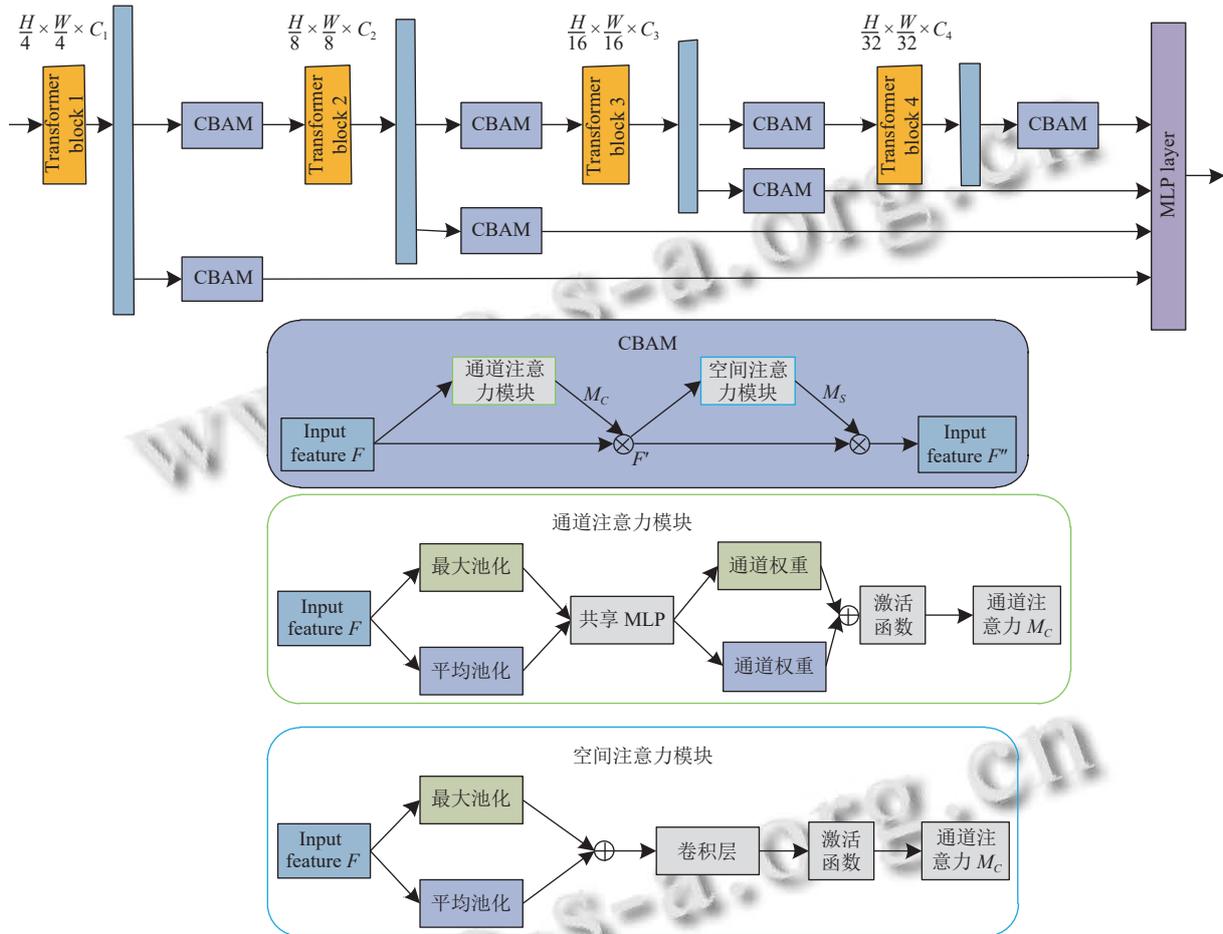


图2 融合注意力机制的 SegFormer 模型编码器

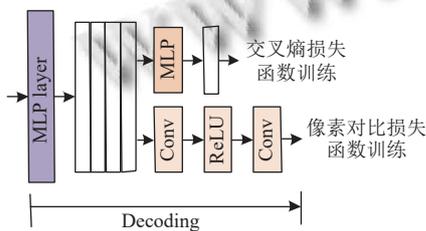


图3 混合损失函数结构

3 实验分析

3.1 道路数据集

所采集的道路数据来自于国产卫星高分一号和高

分二号. 所采集影像的分辨率大小如表1所示.

所采集的详细数据如表2所示. 其中训练集和测试的比例大概为 2:1, 图像大小在 256×256–1024×1024 之间.

表1 原始卫星影像来源及分辨率

卫星	分辨率 (m)
高分一号	2
高分二号	0.8

表2 道路数据集

数据集类别	训练集	测试集	总共
数量	3308	827	4135

数据集中的部分图像如图4所示,分别展示了不同地貌特征下的道路.由此可见卫星影像背景复杂多变,道路形态各异,这给识别精度因拍摄方向的变换杆塔呈现不同的形态特征,同时背景复杂多变,影像分辨率低进一步加大了检测的难度.



图4 数据集样例

3.2 硬件参数和模型参数设定

硬件环境如下:运行在 Windows 10 操作系统上的

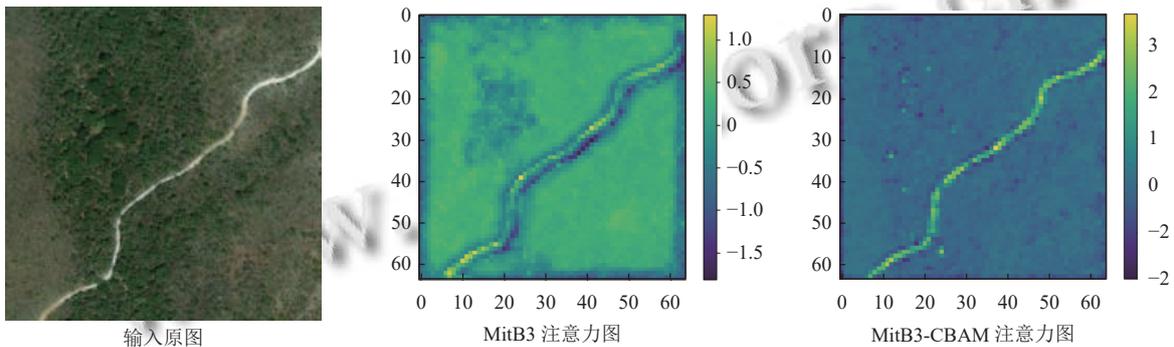


图5 SegFormer 编码器改进前后注意力图对比

3.4 实验对比分析

为了更好地评估改进后的 SegFormer 模型的检测效果,本文与 FCN^[5], SegNet^[17], DeepLabv3+^[18], Swin-UNet^[19], SETR^[12]等模型进行对比,如表4所示.

从表4可见,FCN与SegNet的识别效果最差,主要是因为采用传统的全卷积神经网络. DeepLabv3+相比于上述两种方法,识别效果有所提升,主要是因为

计算机,配备了酷睿 i7 系列的 CPU、32 GB 内存以及一块 TITAN Xp 显卡,显存大小为 12 GB. 该模型的开发基于国产的 PaddlePaddle 1.4.1 深度学习框架.

SegFormer 优化器选择 AdamW,使用两个动量参数, $\beta_1=0.9$ 和 $\beta_2=0.999$,分别用于控制一阶和二阶动量. 权重衰减值 $\text{weight_decay}=0.01$. 网络输入大小为 512×512 . 本文使用多项式衰减策略来调整学习率,初始学习率为 0.000 06. batch size 为 2, epoch 为 256.

3.3 消融实验对比分析

为了验证融合注意力机制的编码器和混合损失函数对 SegFormer 模型精度的影响,本文基于上述数据集进行了测试验证,测试结果如表3所示,在采用 MitB3-CBAM 模块后,相比于基础模型, mIoU 提升约 1.5%. 同时在采用混合损失函数 CE 和 NCE 相结合进行训练优化的情况下, mIoU 进一步提升约 1.8%.

表3 消融实验对比测试结果

Encoder	Loss	Precision	Recall	mIoU
MitB3	CE	0.8875	0.8111	0.7585
MitB3-CBAM	CE	0.8864	0.8321	0.7836
MitB3-CBAM	CE+NCE	0.8912	0.8534	0.7918

从特征图中进行对比分析,如图5, MitB3-CBAM 注意力图相比于 Mit 注意力图有更高的对比度差异,说明更加关注道路区域,从而抑制背景干扰.

有更巧妙复杂的解码器结构. Swin-UNet 和 SETR 采用 Transformer 结构的特征提取网络,在识别效果上更好一些,因为自注意力机制有更大的感受野,可以提取到更有效特征信息. 本文方法采用 Transformer 结构,并同时融合空间注意力和通道注意力机制,效果在上述方法中得到最优的结果. 图6展示了上述算法的部分测试结果. 如图6中第1行中蓝色框中所示,针对这种

对比度较低的 S 型道路, 本文模型相比于其他算法有更完整的分割结果. 如图 6 中第 2 行中黑色框所示, 道路因为复杂背景干扰, 出现了中断的情况, 本文算法可以有效分割出中间的中断区域, 而其他算法并没有分割出来. 如图 6 中第 3 行中棕色框所示, 在城市道路的复杂交叉口处, 本文算法能够分割出较完整的交叉口结构, 而其他算法容易出现缺口. 如图 6 中, 第 4 行黄色框中所示, 针对裸地中比较细窄的道路, 本文算法能

够分割出较完整的结构, 而其他算法在交叉口处易出现缺口的问题.

表 4 不同模型的分割结果对比

方法	Encoder	Precision	Recall	mIoU
FCN	ResNet-50	0.8475	0.7611	0.7085
SegNet	ResNet-50	0.8364	0.7721	0.7036
DeepLabv3+	ResNet-50	0.8774	0.8412	0.7463
Swin-Unet	ViT-Base	0.8966	0.8293	0.7619
SETR	ViT-Base	0.8980	0.8291	0.7626
Ours	MitB3-CBAM	0.8912	0.8534	0.7918

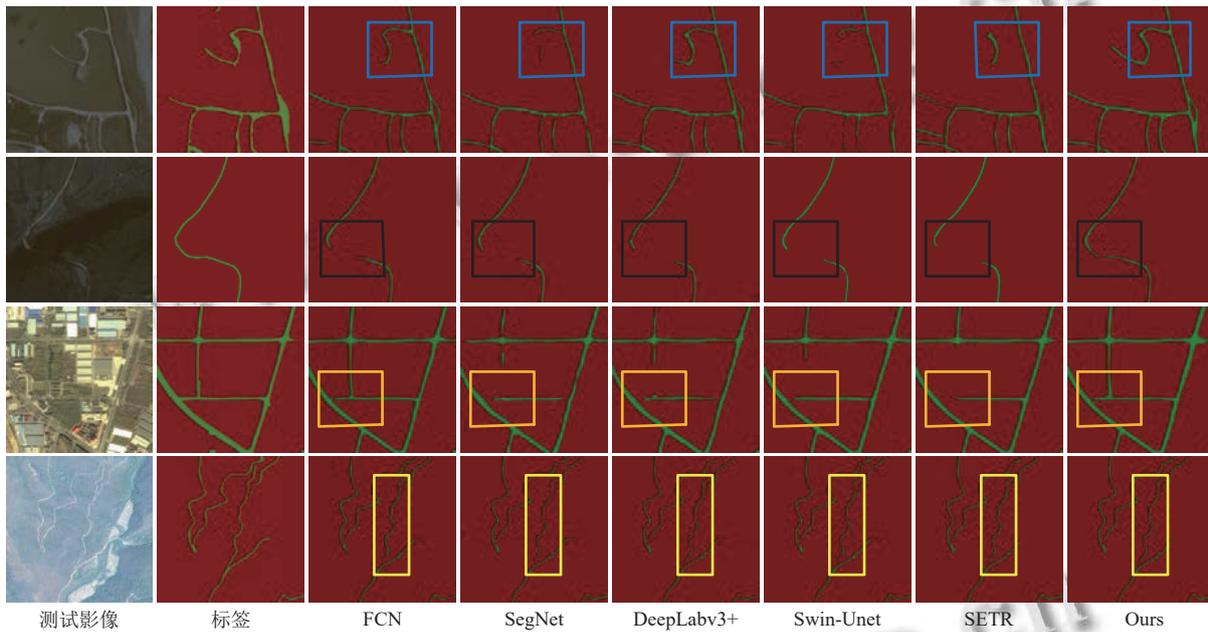


图 6 不同模型的测试分割结果对比

3.5 不同道路识别方法的实验对比分析

本节通过与其他不同道路识别方法进行对比, 来更好地评估改进后的 SegFormer 模型的检测效果. 对比方法有以下 4 个模型, 分别是 FCN-4s^[5], RB-U-Net^[7], SPUD-ResNet^[8], MLUM-Net^[9]等. 实验测试对比如表 5 所示.

表 5 不同道路识别模型的分割结果对比

方法	Encoder	Precision	Recall	mIoU
FCN-4s	VGG-19	0.7465	0.6671	0.6285
RB-U-Net	Mult-CNN-block	0.7364	0.7621	0.7236
SPUD-ResNet	UD-ResNet101	0.8674	0.8312	0.7563
MLUM-Net	MDSPA	0.8456	0.7893	0.7419
Ours	MitB3-CBAM	0.8912	0.8534	0.7918

如表 5 所示, FCN-4s 与 RB-U-Net 的识别效果最差, 主要还是因为采用传统的全卷积神经网络. SPUD-

ResNet 相比于上述两种方法, 识别效果有所提升, 主要是因为设计了更好的骨干网络. MLUM-Net 方法也具有较好的效果, 主要是因为使用多尺度空洞卷积和通道注意力机制来进行编码器的特征提取, 同时对解码器的上采样模块进行优化. 本文方法采用 Transformer 结构, 并同时融合空间注意力和通道注意力机制, 效果在上述方法中得到最优结果. 图 7 展示了上述算法的部分测试结果.

3.6 宽幅影像处理应用

本节将本文方法应用于宽幅卫星影像中, 采用滑动窗口的方式对大尺度影像进行处理. 具体来说, 为了提取大场景中每个区域的目标, 本文设置一个固定大小的窗口, 并沿着从左到右、从上到下的方向对图像进行滑动切分. 这样可以确保获取图像中的每个区域. 然而, 由于切分可能导致目标被分割, 我们需要设置重

叠率,使得相邻切片分子图之间有一定的重叠部分.通过设置重叠率,我们可以更好地解决目标被分割的情况.

本文采用滑动窗口的大小为 512×512 像素,重叠率设置为 10%.应用测试示例如图 8 所示.

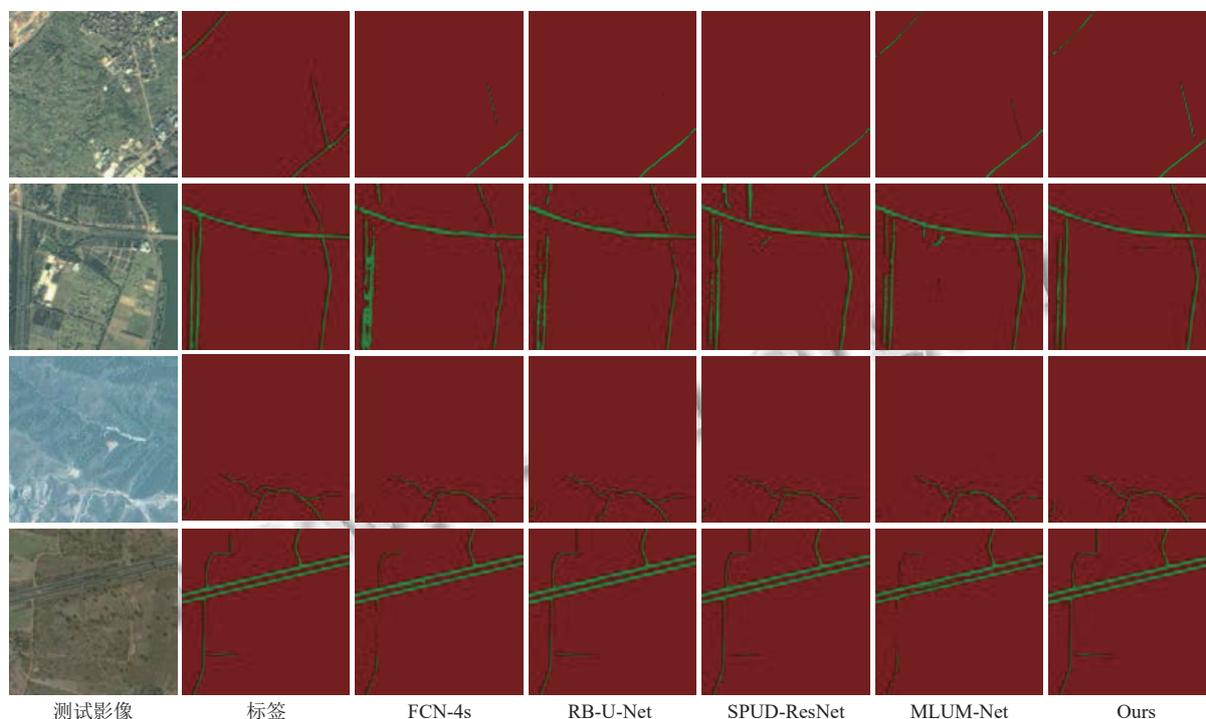


图 7 不同模型的测试分割结果对比



图 8 高分一号影像处理结果

4 结论与展望

卫星影像道路自动识别面临着复杂多变的背景和道路形态多样化的挑战,同时道路像素仅占整个影像的一小部分,导致类别不平衡问题的存在.针对这些问题,本文提出了一种基于改进的 SegFormer 模型的道路自动识别算法.通过两项关键策略的应用,我们取得了显著的改进效果.首先,在 SegFormer 编码器的各个

阶段输出端引入了空间注意力模块,以减弱复杂背景干扰,增强对道路区域的关注.其次,采用了像素对比损失和交叉熵损失相结合的混合损失函数,有效应对了类别不平衡问题.经过优化训练,改进后的模型在测试集上的 mIoU 指标提升了约 3.3%.因此,通过本文所提出的改进算法,我们实现了对道路特征更准确地捕捉和更精确地识别,为卫星影像道路自动识别领域的

研究和应用贡献了新的思路和方法。

参考文献

- 1 党宇, 张继贤, 邓喀中, 等. 基于深度学习 AlexNet 的遥感影像地表覆盖分类评价研究. 地球信息科学学报, 2017, 19(11): 1530–1537.
- 2 阚希, 张永宏, 曹庭, 等. 利用多光谱卫星遥感和深度学习方法进行青藏高原积雪判识. 测绘学报, 2016, 45(10): 1210–1221.
- 3 Li Y, Zhang R, Wu YF. Road network extraction in high-resolution SAR images based CNN features. Proceedings of the 2017 IEEE International Geoscience and Remote Sensing Symposium. Fort Worth: IEEE, 2017. 1664–1667.
- 4 Cheng GL, Wang Y, Xu SB, *et al.* Automatic road detection and centerline extraction via cascaded end-to-end convolutional neural network. IEEE Transactions on Geoscience and Remote Sensing, 2017, 55(6): 3322–3337. [doi: [10.1109/TGRS.2017.2669341](https://doi.org/10.1109/TGRS.2017.2669341)]
- 5 Zhong ZL, Li J, Cui WH, *et al.* Fully convolutional networks for building and road extraction: Preliminary results. Proceedings of the 2016 IEEE International Geoscience and Remote Sensing Symposium. Beijing: IEEE, 2016. 1591–1594.
- 6 袁翠霞, 任艳敏, 韩志强, 等. 高分辨率遥感影像的田间道路识别方法. 测绘通报, 2023(1): 145–148.
- 7 Chen ZY, Wang C, Li J, *et al.* Reconstruction bias U-Net for road extraction from optical remote sensing images. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2021, 14: 2284–2294. [doi: [10.1109/JSTARS.2021.3053603](https://doi.org/10.1109/JSTARS.2021.3053603)]
- 8 李代栋, 赫晓慧, 李盼乐, 等. 基于 SPUD-ResNet 的遥感影像道路提取网络. 计算机工程与应用, 2021, 57(23): 240–247.
- 9 胡绍凯, 赫晓慧, 田智慧. 基于 MLUM-Net 的高分遥感影像土地利用多分类方法. 计算机科学, 2023, 50(5): 161–169.
- 10 Wang LB, Li R, Duan CX, *et al.* A novel Transformer based semantic segmentation scheme for fine-resolution remote sensing images. IEEE Geoscience and Remote Sensing Letters, 2022, 19: 6506105.
- 11 Gao L, Liu H, Yang MH, *et al.* STransFuse: Fusing Swin Transformer and convolutional neural network for remote sensing image semantic segmentation. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2021, 14: 10990–11003. [doi: [10.1109/JSTARS.2021.3119654](https://doi.org/10.1109/JSTARS.2021.3119654)]
- 12 Zheng SX, Lu JC, Zhao HS, *et al.* Rethinking semantic segmentation from a sequence-to-sequence perspective with Transformers. Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021. 6877–6886.
- 13 田雪伟, 汪佳丽, 陈明, 等. 改进 SegFormer 网络的遥感图像语义分割方法. 计算机工程与应用, 2023, 59(8): 217–226.
- 14 Woo S, Park J, Lee JY, *et al.* CBAM: Convolutional block attention module. Proceedings of the 15th European Conference on Computer Vision. Munich: Springer, 2018. 3–19.
- 15 Wang WG, Zhou TF, Yu F, *et al.* Exploring cross-image pixel contrast for semantic segmentation. Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision. Montreal: IEEE, 2021. 7283–7293.
- 16 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.
- 17 Badrinarayanan V, Kendall A, Cipolla R. SegNet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(12): 2481–2495. [doi: [10.1109/TPAMI.2016.2644615](https://doi.org/10.1109/TPAMI.2016.2644615)]
- 18 Chen LC, Zhu YK, Papandreou G, *et al.* Encoder-decoder with atrous separable convolution for semantic image segmentation. Proceedings of the 15th European Conference on Computer Vision. Munich: Springer, 2018. 833–851.
- 19 Cao H, Wang YY, Chen J, *et al.* Swin-Unet: Unet-like pure transformer for medical image segmentation. Proceedings of the 2022 European Conference on Computer Vision. Tel Aviv: Springer, 2022. 205–218.

(校对责编: 孙君艳)