An Algorithm of Road Extraction from Remote Sensing Images Based on Multi-Scale Feature Fusion and Attention Mechanism*

AN Yan, DUAN Yong*
(School of Information Science and Engineering, Shenyang University of Technology, Shenyang Liaoning 110870, China)

Abstract: To address issues such as erroneous segmentation, blurred details, and feature information loss caused by substantial interference from buildings, vegetation, and other elements in remote sensing image road extraction, a road extraction algorithm that integrates multi-scale feature fusion and attention mechanisms is proposed. First, to tackle the loss of road feature information, a three-branch parallel convolution structure is introduced based on the U-Net deep learning network, enhancing the network's ability to preserve road features in remote sensing images. Second, to address the issue of blurred details in extracted roads, an efficient channel attention mechanism is incorporated to suppress interference and better focus on road detail features. Finally, a multi-scale pooling fusion module is developed and embedded at the network's bottom to emphasize multi-scale features and reduce erroneous segmentation. Experimental results show that the proposed algorithm significantly improves road extraction performance, with notable enhancements in extraction precision, completeness, and accuracy.

Key words: deep learning; remote sensing image road extraction; multi-scale features; channel attention
EEACC: 7230 **doi**: 10.3969/j.issn.1004-1699.2026.09.013

基于多尺度特征融合与注意力机制的遥感影像道路提取算法*

安 研, 段 勇*
(沈阳工业大学信息科学与工程学院, 辽宁 沈阳 110870)

摘 要: 针对遥感影像道路提取时有大量建筑物、植被等干扰信息遮挡, 导致道路出现错误分割、细节模糊、特征信息丢失等问题, 提出了一种结合多尺度特征融合和注意力机制的遥感影像道路提取算法。首先, 针对道路特征信息丢失, 以 U-Net 深度学习网络为基础, 提出了一种三路并行卷积结构, 增强了网络模型保留遥感影像中道路特征的能力; 其次, 针对遥感影像提取道路后出现细节模糊的情况, 加入高效通道注意力机制, 抑制干扰因素, 更好地关注了道路细节特征; 最后, 研究出一种多尺度池化融合模块, 嵌入到网络底部, 突出多尺度特征, 以减少错误分割情况。实验结果表明, 所提出的算法具有更好的道路提取效果, 在提取精度、完整性、准确性等方面均有显著提升。

关键词: 深度学习; 遥感影像道路提取; 多尺度特征; 通道注意力
中图分类号: TP391.4 **文献标识码**: A **文章编号**: 1004-1699(2026)09-2022-11

道路信息实时和准确的提取对城市规划和道路交通管理等应用领域具有重要的意义[1]。遥感影像中包含很多的道路特征, 且空间分辨率高, 非常适合进行道路提取任务[2]。然而, 由于遥感影像数据本身的复杂性以及人工标注耗时繁琐, 从大量遥感影像中快速而准确地提取符合需求的道路信息成为一个亟待解决的问题。

近年来, 计算机视觉和人工智能的快速发展为遥感影像道路提取方法的研究提供了新的方向。目前, 对于遥感影像道路提取方法的研究主要分为基于传统算法的道路提取方法和基于深度学习的道路提取方法。基于传统算法的道路提取方法主要局限性在于其依赖静态规则和手动调整, 难以应对复杂地形和多变环境下的道路特征提取需求, 也难以自动化处理大量数据, 并且在面对不同分辨率和光照条件下的遥感影像时表现不稳定, 给遥感影像道路提取带来了诸多困难。基于深度学习的道路提取方法目前得到广泛的关注和研究, 这类方法通常使用语义分割网络, 并结合道路样本数据进行训练, 利用训练好的参数模型, 有效地提取道路信息。这类方法可以在不被人为干涉的情况下自主学习道路特征, 从大规模的遥感影像中提取出低层到高层的语

项目来源: 辽宁省科技厅应用基础研究计划项目(101300243, 101300225)
收稿日期: 2024-05-21 **修改日期**: 2024-09-18

义信息。Zhou 等[3] 提出了以 LinkNet 为主干的网络,通过扩张卷积层的方法对道路信息进行提取,使物体边界分割效果更好。Gao 等[4] 通过向原始的 D-LinkNet 网络中加入通道注意力机制和空间注意力机制,提出了 DAD-LinkNet 网络,使模型可以捕获到更多道路特征。马天浩等[5] 提出一种基于多尺度特征融合的膨胀卷积神经网络,以获得道路信息。李代栋等[6] 通过向 ResNet-101 网络中加入由条形池化和金字塔池化构成的混合池化模块,提出了一种基于 SPUD-ResNet 的遥感影像道路提取网络,更加充分地利用全局上下文信息。何哲等[7] 在 U-Net 的基础上提出一种基于混合尺度注意力的深度学习网络 HSA-UNet,通过在不同尺度的特征上使用条带注意力来增强网络对道路特征的识别效果,同时也抑制不相关区域的干扰。林雨淮等[8] 充分考虑光学影像和 SAR 影像的结构特点,提出了一种多分支双任务的深度学习方法,充分挖掘两种模态影像的优势信息,提高了道路提取结果的精度。宦海等[9] 通过将自适应全局通道注意力模块和多特征融合模块组合,提出了一种全局指导多特征融合网络,有效解决了道路提取误分割率高的问题。李朝奎等[10] 基于 U-Net 结构构建了 DC-Net 模型,通过空洞卷积捕获大范围特征信息,有效提高了卷积效率,同时还可以提取狭长道路的不同尺度特征,有利于模型准确识别农村道路。该类方法适用于背景地物类型多样、特征相似且道路分布复杂的遥感影像,但其提取结果仍会出现道路断裂和毛刺现象,精确度有待进一步提升。

目前,基于深度学习的遥感影像道路提取方法已广泛应用于城市规划、交通管理等领域。郑宛峰[11] 将基于深度学习的遥感影像道路提取算法进行应用,设计并开发出一种面向卫星遥感影像的快速道路提取系统,该系统保证了核心算法研究工作到应用落地的完整性,且道路提取结果准确率较高。Liu 等[12] 通过设计生成对抗网络增强道路像素一致性,并将多尺度双分支分割模块和特征金字塔结构进行集成,提出了 TPEGAN 网络,该方法也被应用到实际场景中,与现有地图服务(如高德地图)所提供的真实道路网络信息进行对比和分析,结果表明,该方法提取出的道路信息真实可靠,为城市规划、道路交通管理提供了宝贵的数据支持。

本文针对传统算法道路提取方式复杂,且基于深度学习的道路提取方法在提取过程中容易出现错误分割、细节模糊、特征信息丢失等问题,提出了一种以 U-Net 深度学习网络[13] 为基础框架,融合三路并行卷积(Triple Parallel Convolution,TPC)结构、高效通道注意力(Efficient Channel Attention,ECA)机制以及多尺度池化融合模块(Multi-Scale Pooling Fusion Module,MSPM)的遥感影像道路提取算法。首先,将基于三路并行卷积的特征提取策略融入 U-Net 网络编码器部分,提高网络对道路完整特征的提取能力,能够解决现有的基于深度学习的遥感影像道路提取方法主要依赖单一尺度的道路特征提取,容易丢失细微特征的问题;其次,针对现有方法易忽略特征重要性分布的问题,将高效通道注意力机制加入到模型的跳跃连接中,该机制通过自适应地调整每个通道的权重,强调模型对道路特征的关注,有效抑制了无关因素的干扰,减轻细节模糊现象;最后,将多尺度池化融合模块嵌入到模型的底部,帮助模型保留更多上下文信息,减少漏检情况,有效克服了现有部分方法在面对遥感影像中不同尺度特征信息时处理能力不足的问题。

1　本文方法

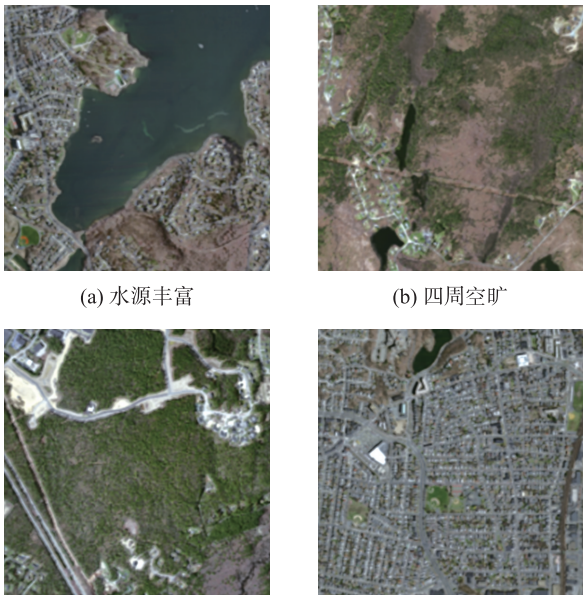
1.1　遥感影像道路提取

遥感影像道路提取任务通常是利用遥感技术获取到影像数据,并使用图像处理和分析的方法,从影像中提取出道路的位置和轮廓信息。在遥感影像中,道路具有明显的颜色、亮度特征、纹理特征以及形状特征,但由于遥感影像的分辨率不一,且具有多样性和复杂性,导致道路提取具有挑战性。遥感影像如图 1 所示,图例中包括多种道路情况,如图 1(a)代表水源丰富情况、图 1(b)代表四周空旷情况、图 1(c)代表树木茂盛情况、图 1(d)代表房屋密集情况,使用传统算法进行道路提取费时且费力,基础的深度学习模型又容易受到树木、房屋等无关因素干扰。因此,通常需要设计合适的道路提取模型对遥感影像整体进行解释和说明,以达到较好的提取效果。

1.2　总体框架

目前,U-Net 网络模型由于网络结构简洁且具有良好的分割效果,被广泛应用于医学图像分割、场景分割等任务中。U-Net 网络具有高度对称性,由编码器和解码器两个部分构成。编码器部分采用传统的卷积网络结构,包含卷积层、池化层和激活函数,用于提取遥感影像道路特征。解码器部分通过上采样层、跳跃连接和卷积层,获取更为精细的道路细节,恢复图像尺寸,从而得到道路提取结果。

但在原始 U-Net 模型的编码器模块中,随着下采样过程的进行,输入图像的尺寸逐渐缩小,一定程



(a) 水源丰富 (b) 四周空旷

(c) 树木茂盛 (d) 房屋密集

图 1 原始遥感影像

度上会丢失图像中道路的细节信息和空间信息。鉴于遥感影像本身具有分辨率高、环境复杂的特点,单一的 U-Net 模型在遥感影像道路提取任务中存在一定局限性。因此,本文以 U-Net 网络模型为主体框架,在模型底部加入提出的多尺度池化融合模块(MSPM),并在横向跳跃连接处引入高效通道注意力(ECA)机制,同时将编码器中的标准卷积替换为一种三路并行卷积(TPC),得到最终模型,其整体结构如图 2 所示。

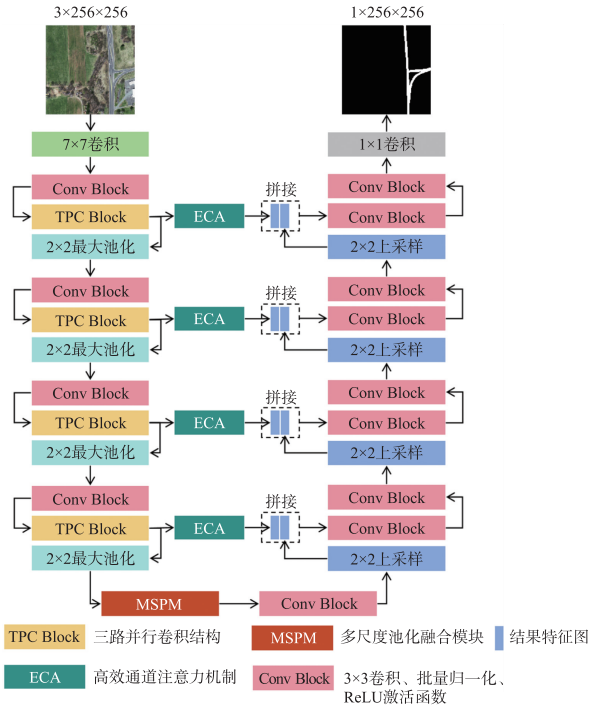


图 2 本文方法整体结构

本文模型的网络架构采用编码器-解码器结构和跳跃连接方式。网络输入图像的形状为 3×256×256,其中 3 为通道数,256×256 为输入原图大小。在编码器部分使用 Conv Block 和三路并行卷积(TPC)结构实现特征提取,其中 Conv Block 包括 3×3 卷积、批量归一化和 ReLU 激活函数。此外,将步长为 2 的最大池化(Maxpool)操作单元置于下采样层,经过一次下采样操作后,特征图大小减半。

解码器部分由 Conv Block 和上采样操作组成,在经过卷积操作和步长为 2 的上采样操作后,特征图大小增加一倍,通道数量减半,以恢复特征图的特征信息和形状。将 ECA 机制添加到编码器与解码器之间的跳跃连接(Skip Connection)中,计算单个样本元素之间的相互作用,抑制无关因素的干扰,加强模型对道路特征的关注程度。最后通过 1×1 卷积将最终结果输出。

1.2.1 三路并行卷积结构

为了提高模型的遥感影像道路特征提取能力,缓解道路特征信息丢失问题,受 GoogleNet[14] 中 Inception 结构的启发,本文设计了一种三路并行卷积(TPC)结构进行道路特征提取。在遥感影像道路提取过程中,普通的卷积操作能有效地捕捉到道路局部特征,但由于其卷积核大小固定,易忽略掉道路的整体形态信息,甚至造成一些重要的道路特征丢失。空洞卷积[15]通过在卷积核中引入间隔点(即空洞),使得卷积核在不增加参数量的前提下可以覆盖更广泛的区域,但将导致在道路细节丰富的区域引入伪影。深度可分离卷积[16]通过逐通道卷积和逐点卷积,显著减少模型参数量,并在一定程度上保留输入数据的空间结构,但逐通道卷积会导致一些低级别的道路特征信息在卷积过程中丢失。相比于上述三种卷积结构,三路并行卷积通过有效整合不同尺度的感受野,能够在遥感影像道路提取中更全面地捕捉道路特征,避免单一卷积类型带来的特征信息丢失或伪影问题,从而提升算法在复杂场景下的表现和实用性。在三路并行卷积结构中,三路分支的感受野分别为 1、3、5,其中 5×5 卷积由两个 3×3 卷积堆叠而成。具体结构如图 3 所示。其中,小感受野可以保留更细致的特征信息,大感受野可以保留像素点之间的相对位置关系,提取更多的空间信息,从而达到缓解特征信息丢失的目的。此外,该结构使用两次 3×3 卷积堆叠的形式替换 5×5 卷积,这种方式可以在具有相同感受野的前提下,一定程度上减少模型的参数量,并提供更多的激活函数以增加模型的非线性,降低模型训练时过拟合的风险。

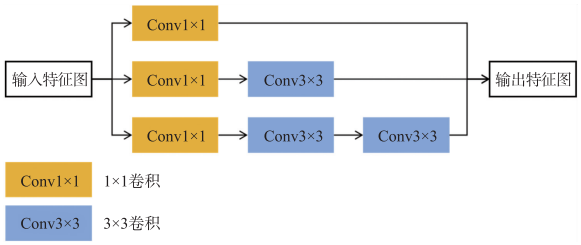
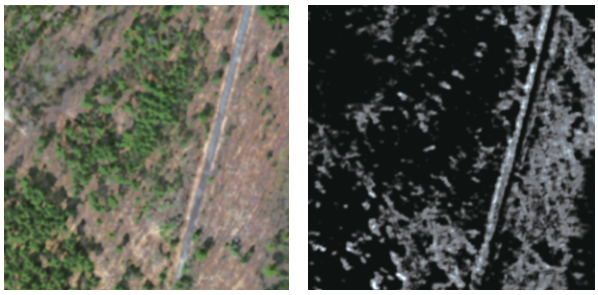


图 3　三路并行卷积结构

计算过程如式(1)所示:

$$F(x)=\delta\{\mathrm{BN}[f_1(x)+f_3(f_1(x))+f_3(f_3(f_1(x)))]\} \tag{1}$$

式中:x 为输入特征图,$F(x)$ 为输出特征图;$f_1(\cdot)$ 表示进行 1×1 卷积操作,$f_3(\cdot)$ 表示进行 3×3 卷积操作;$\mathrm{BN}[\cdot]$ 表示进行批量归一化操作;$\delta\{\cdot\}$ 表示使用 ReLU 激活函数。模型中第一次三路并行卷积后特征图的可视化结果如图 4 所示。其中,图 4(a)为包含道路的原始遥感影像,图 4(b)为 TPC 输出的特征图,图中除明显的道路特征外,还提取到树木、土地等地物信息,能够帮助模型更好地理解和解释所学习的遥感影像。



(a) 原始遥感影像　　(b) TPC输出特征图

图 4　网络中间层特征图可视化

1.2.2　高效通道注意力机制

针对基本 U-Net 网络无法有效克服遥感影像中复杂多样的背景因素对道路特征提取的干扰,本文算法对编码器部分每一层输出特征图使用高效通道注意力(ECA)机制[17],对处理后的特征图和解码器部分对应特征图进行拼接,增强模型对道路特征的表示能力,抑制无关因素的干扰。高效通道注意力机制结构如图 5 所示。

高效通道注意力机制在本文算法中的具体实现过程如下:首先,将 U-Net 模型编码器每一层的输出特征图单独作为高效通道注意力机制的输入特征图,将输入特征图进行不降低维数的全局平均池化(Global Average Pooling,GAP)以获取聚合特征。接着,利用可共享权重的一维卷积对聚合特征进行注意力权重学习。其中,一维卷积的卷积核大小作为该机制的核心参数,表示局部跨通道交互的覆盖范围,该参数由输入特征图的通道数(C)自适应确定。

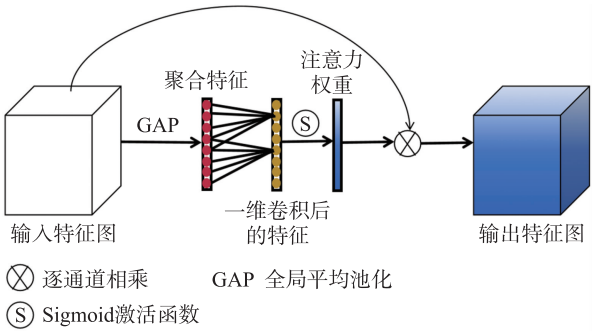


图 5　高效通道注意力机制结构

经过一维卷积处理后的结果通过 Sigmoid 激活函数进行权值重分配,最终将形状为 1×1×C 的权重特征图与原始输入特征图进行逐通道相乘,得到最终的输出特征图。此特征图随后与解码器对应层的结果进行后续处理。具体实现过程如式(2)、式(3)所示,其中式(2)表示注意力权重的生成,式(3)表示核心参数的计算。

$$W=\sigma\{C_k[P(x)]\} \tag{2}$$

式中:x 代表输入特征图,W 代表生成的注意力权重;$\sigma(\cdot)$代表学习注意力权重的 Sigmoid 激活函数;$P(\cdot)$代表进行全局平均池化(GAP)操作;$C_k(\cdot)$代表进行卷积核大小为 k,维度为 1 的卷积操作。其中,卷积核大小 k 由输入特征图 x 的通道数自适应确定,确定参数 k 的过程如式(3)所示:

$$k=\left|\frac{\log_2 C}{\gamma}+\frac{b}{\gamma}\right|_{\mathrm{odd}} \tag{3}$$

式中:$|\cdot|_{\mathrm{odd}}$表示最接近该值的奇数,在计算时若该值为奇数则直接使用该值作为卷积核大小,若为偶数则该值加 1 后再作为卷积核大小,C 为输入特征图的通道数,$\gamma=2$ 和 $b=1$ 为超参数。确定参数 k 后,将经过式(2)计算得到的注意力权重图与输入特征图 x 逐通道相乘,从而得到新的输出特征图,如式(4)所示:

$$G(x)=W\otimes x \tag{4}$$

式中:x 为输入特征图,W 为式(2)运算得到的注意力权重图,$\otimes$ 表示逐通道相乘,$G(x)$ 为最终输出结果。

对于一些常见的通道注意力机制,如 SENet[18]中的通道注意力模块,通过两次全连接操作学习特征权重,不依据特征通道数值的分配进行判断。这种方法不仅增加了模型的复杂性,而且降维操作对通道注意力的预测不利。本文采用的高效通道注意力机制通过不降维度的局部跨通道交互策略,减轻了降维操作对注意力学习的负面影响,使模型更好地关注道路区域。此外,高效通道注意力模块参数

量较少,有助于降低模型复杂度,缓解因无关因素影响而导致的道路细节模糊问题。

1.2.3 多尺度池化融合模块

针对遥感影像道路提取中出现的错误分割、过度分割情况,本文在 U-Net 网络结构的底部嵌入了一种多尺度池化融合模块(MSPM)。该模块通过应用不同大小、方式的池化操作,实现多尺度特征的融合,从而捕获更多全局上下文信息。通过将多尺度策略与池化操作整合,该模块提高了语义分割模型对不同尺度物体的识别能力,并有效减少了预测结果中的错误分割情况。该模块具体结构如图 6 所示。

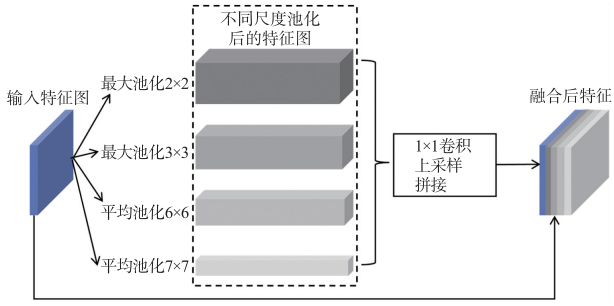


图 6 多尺度池化融合模块

输入特征图分别经过 2×2、3×3 的最大池化操作以及 6×6、7×7 的平均池化操作,对全局上下文信息进行处理。接着,四个分支的输出分别经过一维卷积,将通道数调整至原始特征图维度的 1/N,其中 N 为原始特征图的通道数。随后,通过双线性插值对四个输出进行上采样,最后将四个结果和原始特征图进行拼接,从而获得最终输出结果。

上述结构采用了多尺度池化融合策略,由于深层特征图的特征信息均由编码器提取,而单一的池化方式可能无法充分利用全局语义信息,从而导致一定程度的错误分割,因此,本文引入多尺度池化融合模块。在该模块中,2×2 和 3×3 两个尺度采用最大池化方式,而 6×6 和 7×7 两个尺度采用平均池化方式,以编码上下文特征信息。这种方法有助于模型对不同尺度的道路特征进行有效感知,从而减少道路的错误分割情况。

1.3 损失函数

遥感影像道路提取任务属于语义分割领域,实质上是对遥感影像中像素点进行分类的过程。常用的损失函数是二元交叉熵(Binary Cross-Entropy, BCE)损失函数,然而,在分割过程中,如果图像存在类别不平衡问题,例如道路提取任务中道路是一段细长狭窄的区域,仅占整幅图像的小部分,道路和背景之间存在样本类别高度不平衡的问题,会导致模型倾向于学习背景信息。骰子系数(Dice Coefficient, Dice)是一种关于集合相似度的度量函数,其对应的骰子损失函数(Dice Loss)可以把某一类别的所有像素当作一个整体,并计算类别间交集部分在整体中所占比例,这种方法不受背景像素数量的影响,在样本不平衡情况下表现更佳。但骰子损失函数对噪声敏感,可能忽略边界信息导致边界分割不佳。因此,本文将二元交叉熵损失函数与骰子损失函数相结合,设计了一种复合型损失函数。

二元交叉熵损失函数与骰子损失函数的计算公式如(5)、(6)所示:

$$L_{\mathrm{BCE}}=-\frac{1}{m}\sum_{i=1}^{m}\left[y_i\log_2\bar{y}+(1-y_i)\log_2(1-\bar{y})\right] \tag{5}$$

式中:y_i 为像素真实值,$\bar{y}$ 为像素预测值,m 为类别数。

$$L_D=1-\frac{2|X\cap Y|}{|X|+|Y|} \tag{6}$$

式中:$|X\cap Y|$ 表示 X 和 Y 之间交集元素的个数,$|X|$ 和 $|Y|$ 分别表示 X、Y 中元素的个数。

最终的复合型损失函数定义如式(7)所示:

$$L=L_{\mathrm{BCE}}+\alpha L_D \tag{7}$$

式中:α 代表权重系数。本文将使用该复合型损失函数用于模型损失的计算,以缓解道路提取任务中类别不平衡现象。

2 实验结果及分析

为了验证本文提出方法在遥感影像道路提取任务上的有效性,本文在 Massachusetts 道路数据集上进行测试,并设置了消融实验,以评估本文提出的各个模块对遥感影像道路提取任务的作用。此外,本文还进行了对比实验,将本文所提方法与几种主要语义分割模型的道路提取结果进行对比和分析。

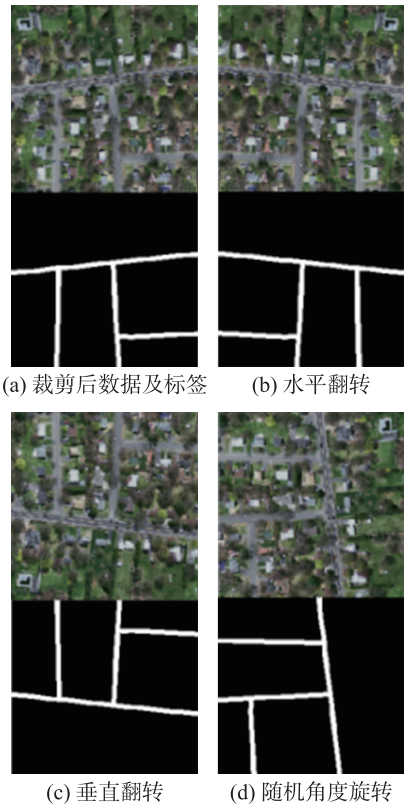
2.1 实验数据集及数据预处理

本文所使用的数据集是 Massachusetts 道路数据集,该数据集是世界上最大的公开道路数据集,共包含 1 171 张图片,每张图片大小为 1 500 像素×1 500 像素,空间分辨率为 1.2 m。该数据集覆盖了美国马萨诸塞州大约 2 600 km^2 范围内所有地物信息的遥感影像数据,道路信息主要包括城市道路、郊区道路、农村道路和山区道路,地物信息包括道路、河流、各种建筑物、植被、桥梁等。该数据集广泛的覆盖范围和多样化的地物信息,使遥感影像道路提取任务具有挑战性和复杂性,且可以帮助模型学习到不同环境条件下的道路特征,有助于验证本文算法的泛化能力,充分体现出本文研究的遥感影像道路提取问题具有实际

研究意义,因此本文选用 Massachusetts 道路数据集进行实验。

由于该数据集中存在严重的图像及标签数据丢失现象,故需要删除缺失、损坏的图像及对应标签,并将训练集、测试集和验证集合并,后续重新按比例划分。由于删除后数据集中图像数量较少,无法支撑模型训练,故将数据集中原始遥感图像和标签图像以 256 为步长进行零重叠率裁剪,裁剪共得到 25 张切片图像,裁剪过后的切片图像大小为 256 像素×256 像素。

此外,本文还采用了水平翻转、垂直翻转和随机角度旋转(顺时针 90°、180°、270°)的数据增强方式扩充数据集。经上述处理,最终得到 18 007 张道路图像以及对应标签图像,并将其按照 8∶1∶1 的比例划分为训练集、验证集和测试集。部分样本如图 7 所示,图 7(a)所示为裁剪后数据集中遥感影像及标签;对图 7(a)进行水平翻转得到图 7(b);对图 7(a)进行垂直翻转得到图 7(c);对图 7(a)进行随机角度旋转得到图 7(d)。



(a) 裁剪后数据及标签　(b) 水平翻转

(c) 垂直翻转　(d) 随机角度旋转

图 7　扩充数据集

本文所有实验均采用 Adam 优化器优化权重更新,初始学习率设置为 0.000 3,模型迭代次数为 80,Batch Size 大小为 12。

2.2　评价指标

为检验所提方法的有效性,本文使用准确率 Accuracy、精准率 Precision、召回率 Recall 和 F1-Score 对网络性能进行评估,计算公式如下:

$$\text{Accuracy}=\frac{(\text{TP}+\text{TN})}{(\text{TP}+\text{TN}+\text{FP}+\text{FN})} \tag{8}$$

$$\text{Precision}=\frac{\text{TP}}{(\text{TP}+\text{FP})} \tag{9}$$

$$\text{Recall}=\frac{\text{TP}}{(\text{TP}+\text{FN})} \tag{10}$$

$$\text{F1-Score}=\frac{2\times\text{Precision}\times\text{Recall}}{(\text{Precision}+\text{Recall})} \tag{11}$$

式中:TP 表示预测和标签都是正样本;TN 表示预测和标签都是负样本;FP 表示预测为正样本,标签为负样本;FN 表示预测是负样本,标签为正样本。在计算时,通常在每个指标的分母处加一个极小的常数,避免出现除零错误,本文该值取 1e-6。

2.3　消融实验

2.3.1　复合型损失函数有效性

首先讨论在复合型损失函数中,使用不同权重系数 α 对本文方法道路提取能力的影响,从而确定 α 的最优值,并验证该损失函数的有效性。使用本文方法在 Massachusetts 道路数据集上共做了五组实验,逐渐增加骰子系数(Dice)损失函数所占比重,观察各评价指标。实验结果如表 1 所示。当 α 为 0 时,表示仅单独使用二元交叉熵(BCE)损失函数进行训练,而当 α 取 1 时,得到的准确率(A)、精准率(P)、召回率(R)和 F1 分数(F1)均高于其他组合结果,因此本文取 α 为 1。根据上述实验可以得出结论,使用本文提出的复合型损失函数对于处理样本不平衡问题的效果,比单独使用二元交叉熵损失函数更好。

表 1　不同权重 Dice loss 性能结果

参数设置 α	评价指标			
	A/%	P/%	R/%	F1/%
0	96.20	68.95	63.50	62.97
0.25	96.55	69.32	63.83	63.88
0.5	96.98	69.94	64.82	65.03
0.75	97.18	70.92	65.58	65.13
1	**97.22**	**73.36**	**68.84**	**68.62**

2.3.2　卷积结构对比

为了验证本文提出的三路并行卷积结构相比于其他卷积结构在遥感影像道路提取任务中具有一定优势,本文在保证超参数等其他因素不变的情况下,选择了普通卷积、空洞卷积(扩张率为 2)、深度可分离卷积以及三路并行卷积结构进行对比实验,并将 U-Net 模型作为实验的基准模型(Baseline),在测试

集上得到各指标结果的平均值,如表 2 所示。结果表明,使用空洞卷积的模型在各项评价指标上均高于使用普通卷积和深度可分离卷积的模型,而使用三路并行卷积的模型相比于使用空洞卷积的模型在准确率、精准率、召回率和 F1 分数分别提高了 0.34%、1.93%、8.56%、6.24%。根据上述实验可得出结论,使用本文提出的三路并行卷积结构相较于使用普通卷积、空洞卷积或深度可分离卷积结构,在准确率、精准率、召回率和 F1 分数上可以取得更好的表现。

表 2 不同卷积结构性能结果

模型	评价指标			
	A/%	*P*/%	*R*/%	F1/%
Baseline+ECA+MSPM(普通卷积)	96.71	70.93	56.28	58.73
Baseline+AC+ECA+MSPM(空洞卷积)	96.88	71.43	60.28	62.38
Baseline+DW+ECA+MSPM(深度可分离卷积)	96.58	70.55	55.73	57.71
Baseline+TPC+ECA+MSPM(本文方法)	**97.22**	**73.36**	**68.84**	**68.62**

2.3.3 功能模块有效性

为验证本文方法提出各模块的有效性,本文对三路并行卷积(TPC)结构、高效通道注意力(ECA)机制和多尺度池化融合模块(MSPM)进行消融实验,在测试集上得到结果的平均值如表 3 所示。从评价指标可以看出,本文方法在各项指标上都高于基准模型 U-Net,且准确率、精准率、召回率和 F1 分数分别提高了 0.68%、2.74%、15.51%、13.64%。此外,向 U-Net 模型中加入 TPC 模块和向 U-Net 模型中加入 TPC 以及 ECA 模块,除前者的精准率略有下降外,其他指标均有明显提高,说明两个功能模块对遥感影像道路提取起到促进作用。最后,本文提出模型与向 U-Net 模型中加入 TPC 以及 ECA 模块的模型从评价指标上看,四项指标均高于后者,说明 MSPM 有效扩大了网络感受野范围,提高了模型提取全局特征的能力。

表 3 消融实验结果对比

模型	评价指标			
	A/%	*P*/%	*R*/%	F1/%
Baseline	96.54	70.82	53.33	54.98
Baseline+TPC	96.61	70.62	54.73	55.20
Baseline+TPC+ECA	97.01	72.25	62.92	64.22
Baseline+TPC+ECA+MSPM(本文方法)	**97.22**	**73.36**	**68.84**	**68.62**

2.4 对比实验

为验证本文所提方法的有效性,将其分别与其他现有模型(FCN[19]、PSPNet[20]、U-Net、ResU-Net[21]、Deeplabv3+[22]、CE-Net[23])在选取的 Massachusetts 道路数据集上进行对比。利用准确率、精准率等评价指标对测试集结果的平均值进行评估,具体结果如表 4 所示。

表 4 Massachusetts 数据集上不同模型结果对比

模型	评价指标			
	A/%	*P*/%	*R*/%	F1/%
FCN	96.43	68.73	45.88	51.50
PSPNet	96.50	68.88	47.82	51.83
U-Net	96.54	72.02	53.33	54.98
ResU-Net	96.58	72.57	53.59	56.58
Deeplabv3+	96.80	73.14	55.27	54.21
CE-Net	96.84	**76.75**	51.48	58.52
本文方法	**97.22**	73.36	**68.84**	**68.62**

不同模型各评价指标对比如图 8 所示。从图中可以看出,原始 U-Net 模型在准确率、精准率、召回率和 F1 分数均高于 FCN 和 PSPNet 模型,说明 U-Net 模型由于其编码器-解码器的特殊结构,使得它相对于 FCN、PSPNet 等模型在语义分割任务中的性能更好,但整体指标较低;ResU-Net 模型由于加入残差结构,建立了更深层次的网络,促进道路特征信息传递;Deeplabv3+模型通过引入空洞空间金字

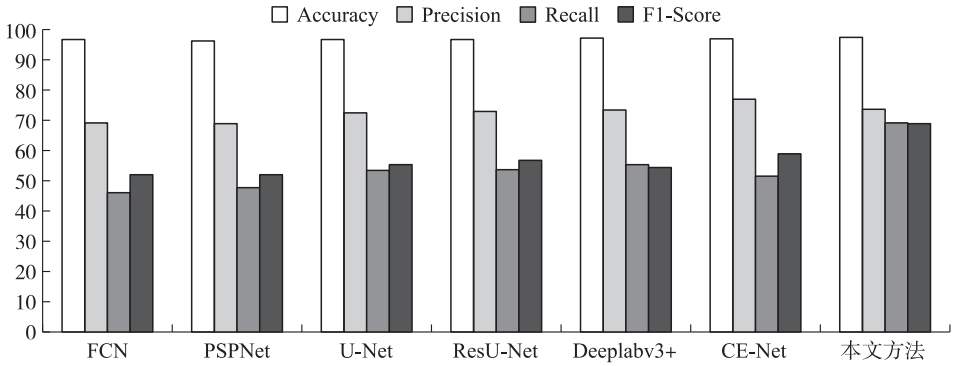


图 8 Massachusetts 数据集上不同指标评估模型性能

塔池化模块,扩大感受野的同时可以有效地增强模型对遥感影像中道路特征的提取能力。上述两种模型由于各自结构的特点,评价指标都不同程度地高于 U-Net 模型,但它们无法充分利用上下文信息,网络预测结果容易出现错误分割情况。CE-Net 在其模型底部嵌入上下文提取网络,提取到多尺度的道路特征信息,并联系上下文,提高模型对道路的预测能力,但由于模型提取到的特征信息过多,无法关注到重点区域,容易受到无关因素的干扰,因此在像素级别的边界分割和召回上不如 U-Net 等模型。本文模型的准确率、召回率和 F1 分数指标值均高于其余六个对比网络,尤其在召回率和 F1 分数上效果显著,相较于原始 U-Net 模型,召回率提高了 15.51%,F1 分数提高了 13.64%;相比于表现较好的 ResU-Net 和 Deeplabv3+模型,在召回率上高出 ResU-Net 模型 15.25%、高出 Deeplabv3+模型 13.57%,在 F1-分数上高出 ResU-Net 模型 12.04%、高出 Deeplabv3+模型 14.41%,和 CE-Net 模型对比,仅在精准率方面低于 CE-Net 模型。对比分析表明,本文算法的道路提取性能更好,具有明显优势。

为进一步证明本文方法在其他数据集上具有有效性和普遍适用性,本文进行了两个补充实验,实验数据来自于 CHN6-CUG 数据集和 DRIVE 数据集。CHN6-CUG 道路数据集是来自于中国代表性城市的新型大规模卫星影像数据集,其遥感影像底图来自谷歌地球,该数据集选取了 6 个城市规模、发展程度不同的城市,包括北京、武汉等,标示道路包括铁路、公路、城市道路和农村道路等。该数据集可以使不同模型在道路提取时提取到更精细的边缘和纹理特征,为后续结果对比和分析提供重要的数据支持。

另外一个数据集选择的是 DRIVE 数据集。DRIVE 数据集源自于荷兰一项专注于糖尿病视网膜病变筛查的研究项目,训练集每幅图像都有相应的专家手工分割结果和掩膜。该数据集提供了高分辨率的视网膜血管影像,这些影像中包含的血管结构具有明显的边缘和细节特征,与道路提取任务相似,通过在不同类型数据上验证,能够更全面地评估算法的普遍适用性,确保其不仅在特定领域数据集上有效,也能适用于其他领域类似任务。

为了更好地进行模型训练,将两个数据集中的原始训练数据进行零重叠率裁剪,裁剪后去掉不包含标签的数据,并采用和 Massachusetts 道路数据集相同的数据增强方式扩充数据集,最终将处理后的两个数据集均按照 8∶1∶1 的比例划分为训练集、验证集和测试集。

不同模型在 CHN6-CUG 数据集上的评价指标结果如表 5 所示。

表 5　CHN6-CUG 数据集上不同模型结果对比

模型	评价指标			
	A/%	*P*/%	*R*/%	F1/%
FCN	81.80	49.50	53.15	46.68
PSPNet	82.34	51.33	53.24	48.61
U-Net	90.64	55.04	59.07	52.18
ResU-Net	91.58	58.62	59.18	56.82
Deeplabv3+	92.78	62.62	61.78	55.19
CE-Net	92.96	63.91	57.48	57.03
本文方法	**93.22**	**64.42**	**72.69**	**58.09**

表中可以看出,FCN 和 PSPNet 各项指标较低且相差不大,而 U-Net 的表现优于这两者,这归因于其对称的编码-解码结构,有效地保留了细节信息;ResU-Net 的性能在 U-Net 之上,由于引入了残差连接,增强了模型对道路特征的提取能力;Deeplabv3+模型引入了空洞空间金字塔池化模块,使得整体预测精准率和召回率得到有效提高;CE-Net 模型面对该数据集并没有表现出较高的性能,召回率值偏低,这可能由于模型没有很好地适应该数据集的分辨率,导致模型受到无关地物因素的干扰;本文方法在该数据集上取得了较高的性能,各项指标均高于对比模型,其中召回率高出 Deeplabv3+约 10.91%、高出 CE-Net 约 15.21%,说明本文提出的各模块适用于不同分辨率的遥感影像道路提取任务。上述对比可见,本文算法相比于其他对比模型在 CHN6-CUG 数据集上的道路提取性能更佳。

不同模型在 DRIVE 数据集上的评价指标结果如表 6 所示。

表 6　DRIVE 数据集上不同模型结果对比

模型	评价指标			
	A/%	*P*/%	*R*/%	F1/%
FCN	96.90	74.93	80.48	77.42
PSPNet	96.55	75.69	78.70	76.83
U-Net	96.91	78.79	81.24	79.17
ResU-Net	97.17	75.69	84.70	81.11
Deeplabv3+	97.27	78.95	85.13	81.55
CE-Net	97.42	79.99	86.55	82.96
本文方法	**97.89**	**83.52**	**91.76**	**84.15**

在 DRIVE 数据集上不同模型的评价指标揭示了不同模型在血管分割任务中的表现差异,首先,FCN 和 PSPNet 的性能较低,这可能由于模型缺乏足够的特征细节提取能力;U-Net 通过其跳跃连接结构,使得性能有所提升,但对于血管细节部分的准

确提取仍存在改进空间;ResU-Net 引入的残差结构进一步优化了血管特征提取,但其精准率较低;Deeplabv3+采用的金字塔技术在多尺度特征提取方面表现优异,使得模型的整体性能优于之前几种模型,但仍无法超越 CE-Net;CE-Net 特殊的上下文提取结构使其在准确率和召回率上表现良好,适应了眼底视网膜血管的复杂结构;最后,本文算法在所有对比模型中表现最优,本文提出的三路并行卷积结构和多尺度池化融合模块可以结合上下文信息,加强对细微血管特征的捕捉。同时,高效通道注意力机制也抑制了无关因素对血管特征提取造成的干扰。通过上述实验分析,本文方法在眼底视网膜血管分割任务上具有一定优势。

除上述对评价指标结果的对比分析外,本文还选取了三个数据集中部分测试影像的提取结果进行对比分析,如图 9 所示。

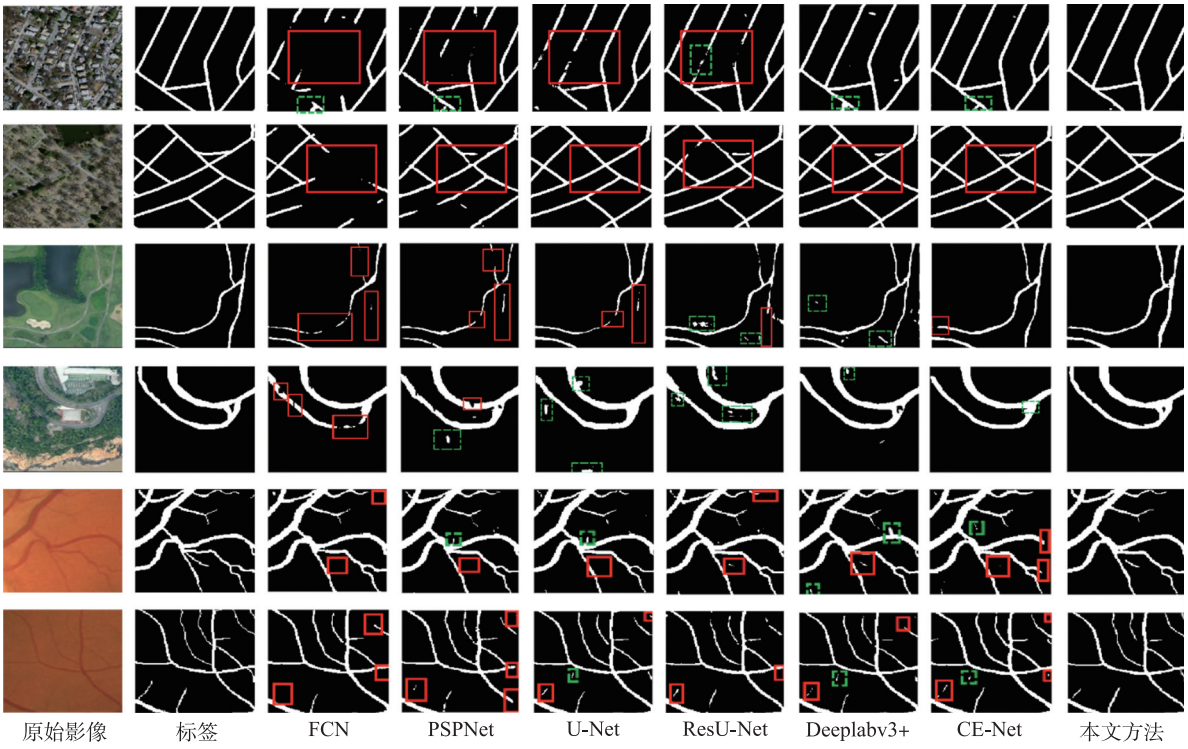


图 9 各数据集提取结果对比

图 9 中,第一、二行选自 Massachusetts 数据集,第三、四行选自 CHN6-CUG 数据集,第五、六行选自 DRIVE 数据集。第二、三行分别代表四周空旷和水源丰富两种道路情况。对于两个道路数据集中背景单一、干扰信息较少的四周空旷遥感影像和水源丰富遥感影像,FCN 模型和 PSPNet 模型表现一般,预测结果中出现大范围的道路缺失现象,如图中实线矩形区域所示;相比上述两种模型,U-Net 模型更适应这两类情况,预测结果中道路提取较好,但 FCN、PSPNet 和 U-Net 模型的预测结果与标签图像相比,漏检情况仍较多,主要表现为道路断裂、缺失;ResU-Net、Deeplabv3+和 CE-Net 模型由于结构更丰富,故而预测出的道路连续且完整,但对细节处理不足,出现了过度提取现象,如图中虚线矩形区域所示。本文方法通过在模型底部嵌入 MSPM,保留了更多上下文信息,可以有效改善道路错误分割、过度提取现象。此外,本文使用 TPC 结构,通过大小感受野的作用,将像素点间相对位置进行联系,提取到更细致的道路信息,缓解了由于道路信息丢失所造成的道路断裂现象。第一、四行分别代表房屋密集和树木茂盛两种道路情况,对于两个道路数据集中背景复杂、干扰因素较多的房屋密集影像和树木茂盛影像,由于受到干扰,FCN、PSPNet 和 U-Net 模型道路提取效果较差;ResU-Net、Deeplabv3+和 CE-Net 模型可以提取出较为全面的道路信息,并且预测结果中大部分道路整体连续,但由于上述几种模型结构各异,且都没有针对干扰项进行处理,使得预测结果中出现道路连续性差、边缘模糊现象,如图中实线矩形区域和虚线矩形区域所示。本文方法通过引入 ECA 机制,有效抑制了无关地物信息的干扰,预测结果与标签图像的重合程度更高,相比其他模型具有一定优势。

对于第五、六行的视网膜血管影像,FCN、PSPNet、U-Net、ResU-Net 模型都能基本分割出视网

膜血管轮廓,但仍存在细微血管特征提取较少的问题,从而导致细微血管未完全分割、血管末端消失,如图中实线矩形区域所示。而 Deeplabv3+、CE-Net 模型对于视网膜血管的分割较为完整,一定程度上减少了血管末端漏检的现象,但也由于重复捕捉视网膜血管特征,出现了一定的过度分割,如图中虚线矩形区域所示。本文方法通过三个模块间协调作用,最大化采集各种形状血管的特征,从而实现对视网膜主血管与细微血管的精准提取。

3　结论

在遥感影像道路提取过程中,深度学习模型常常会受到植被和建筑物等干扰信息的影响,从而导致分割错误、细节模糊以及特征信息丢失等问题。本文提出了一种基于多尺度特征融合与注意力机制的遥感影像道路提取算法,并在 Massachusetts 道路数据集、CHN6-CUG 道路数据集和 DRIVE 视网膜血管数据集上进行实验。通过设置消融实验、对比实验,验证了本文方法中提出的各个模块的有效性。此外,还设置了补充实验,保证本文方法对其他数据集具有普遍适用性。实验结果显示,本文方法可以有效提取遥感影像中的道路特征,预测出的结果道路连续且完整,边界清晰,在准确率、精准率、召回率和 F1 分数四种评价指标上均具有一定优势。

参考文献:

[1] Wang W, Yang N, Zhang Y, et al. A Review of Road Extraction from Remote Sensing Images [J]. Journal of Traffic and Transportation Engineering 2016, 3 (3): 271-282.

[2] 刘志恒,岳子腾,周绥平,等. 融合空洞空间金字塔池化和注意力的轻量化遥感影像道路提取[J]. 航天返回与遥感,2024,45(1):111-122.

[3] Zhou L, Zhang C, Wu M. D-LinkNet: LinkNet with Pretrained Encoder and Dilated Convolution for High Resolution Satellite Imagery Road Extraction[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, IEEE, 2018: 182-186.

[4] Gao L, Wang J, Wang Q, et al. Road Extraction Using a Dual Attention Dilated-linknet Based on Satellite Images and Floating Vehicle Trajectory Data[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2021, 14(1): 10428-10438.

[5] 马天浩,谭海,李天琪,等. 多尺度特征融合的膨胀卷积残差网络高分一号影像道路提取[J]. 激光与光电子学进展,2021,58(2):341-348.

[6] 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[J]. 计算机工程与应用,2021,57(23):240-247.

[7] 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):385-392.

[8] 林雨准,金飞,王淑香,等. 多分支双任务的多模态遥感影像道路提取方法[J]. 地球信息科学学报,2024,26(6):1547-1561.

[9] 宦海,盛宇,顾晨曦. 基于遥感图像道路提取的全局指导多特征融合网络[J]. 浙江大学学报(工学版),2024,58(4):696-707.

[10] 李朝奎,曾强国,方军,等. 改进全卷积网络方法的高分二号影像农村道路提取[J]. 遥感学报,2021,25(9):1978-1988.

[11] 郑宛峰. 卫星图像道路快速提取的技术研究与实现[D]. 北京:北京邮电大学,2023.

[12] Liu R Q, Li F T, Jiang W P, et al. Generating Pixel Enhancement for Road Extraction in High-Resolution Aerial Images[J]. IEEE Transactions on Intelligent Vehicles, 2024, 9(10): 6313-6325.

[13] Ronneberger O, Fischer P, Brox T. U-net: Convolutional Networks for Biomedical Image Segmentation [C]//The 18th International Conference of Medical Image Computing and Computer-assisted Intervention, Munich, Germany, 2015: 234-241.

[14] Szegedy C, Liu W, Jia Y, et al. Going Deeper with Convolutions[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, IEEE, 2015: 1-9.

[15] 韩玲,杨朝辉,李良志,等. 利用 Deeplabv3 提取高分辨率遥感影像道路[J]. 遥感信息,2021,36(1):22-28.

[16] Li W, Chen H, Liu Q, et al. Attention Mechanism and Depthwise Separable Convolution Aided 3DCNN for Hyperspectral Remote Sensing Image Classification[J]. Remote Sensing, 2022, 14(9): 2215-2240.

[17] Wang Q, Wu B, Zhu P, et al. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, IEEE, 2020: 11534-11542.

[18] Hu J, Shen L, Sun G. Squeeze-and-Excitation Networks [C]//2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, IEEE, 2018: 7132-7141.

[19] Long J, Shelhamer E, Darrell T. Fully Convolutional Networks for Semantic Segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 39(4): 640-651.

[20] Zhao H, Shi J, Qi X, et al. Pyramid Scene Parsing Network [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, IEEE, 2017: 2881-2890.

[21] 胡煦航,程小龙,朱滨,等. 基于改进 ResUnet 的高分辨率遥感影像道路信息提取[J]. 遥感信息, 2022, 37(4): 87-93.

[22] Chen L C, Zhu Y, Papandreou G, et al. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation[C]//2018 European Conference on Computer Vision (ECCV), Munich, Germany, 2018: 801-818.

[23] Gu Z, Cheng J, Fu H, et al. CE-Net: Context Encoder Network for 2d Medical Image Segmentation[J]. IEEE Transactions on Medical Imaging, 2019, 38(10): 2281-2292.

安 研(2000—),男,硕士研究生。主要研究方向为深度学习、计算机视觉等,anyan0611@163.com;

段 勇(1978—),男,教授,博士生导师,通讯作者。主要研究方向为机器学习、计算机视觉、自主机器人等,duanyong0607@163.com。