

基于改进 U-Net 的遥感图像道路提取

何哲^{1,2}, 陶于祥^{1,2*}, 罗小波^{1,2}, 徐浩^{1,2}

¹重庆邮电大学计算机科学与技术学院, 重庆 400065;

²重庆邮电大学空间大数据研究中心, 重庆 400065

摘要 从遥感图像中提取的道路信息在城市规划、交通管理等领域具有极其巨大的价值,但由于背景复杂、障碍物遮挡以及大量类似的非道路区域的干扰,目前从遥感图像中提取高质量道路信息仍受到限制。为此,提出一种基于混合尺度注意力 U-Net 的遥感图像道路提取方法 HSA-UNet。首先,在编码网络使用由残差结构和注意力特征融合机制构成的注意力残差学习单元,提高对全局特征和局部特征的提取能力;其次,由于道路具有跨度大、狭窄且连续分布等特点,在桥接网络加入注意力增强的空洞空间金字塔池化模块来增强不同尺度道路特征提取能力。在 Massachusetts 道路数据集上进行的实验结果表明,在 F1、交并比等评价指标上,HSA-UNet 明显优于 D-LinkNet、DeepLabV3+ 等语义分割网络。

关键词 遥感; 遥感图像; 语义分割; 道路提取; 尺度注意力; 注意力特征融合

中图分类号 TP75

文献标志码 A

DOI: 10.3788/LOP222634

Road Extraction from Remote Sensing Image Based on an Improved U-Net

He Zhe^{1,2}, Tao Yuxiang^{1,2*}, Luo Xiaobo^{1,2}, Xu Hao^{1,2}

¹School of Computer Sciences and Technology, Chongqing University of Posts and Telecommunications,
Chongqing 400065, China;

²Spatial Big Data Research Center, Chongqing University of Posts and Telecommunications,
Chongqing 400065, China

Abstract Road information extracted from remote sensing images is of great value in urban planning, traffic management, and other fields. However, owing to the complex background, obstacles, and numerous similar nonroad areas, high-quality road information extraction from remote sensing images is still challenging. In this work, we propose HSA-UNet, a road information extraction method based on mixed-scale attention and U-Net, for high-quality remote sensing images. First, an attention residual learning unit, composed of a residual structure and an attention feature fusion mechanism, is used in the coding network to improve the extraction ability of global and local features. Second, owing to roads with the characteristics of large spans, narrowness, and continuous distribution, the attention-enhanced atrous spatial pyramid pooling module is added to the bridge network to enhance the ability of road features extraction at different scales. Experiments were performed on Massachusetts roads dataset, and the results showed that HSA-UNet significantly outperformed D-LinkNet, DeepLabV3+, and other semantic segmentation networks in terms of F1, intersection over union, and other evaluation indicators.

Key words remote sensing; remote sensing image; semantic segmentation; road extraction; scale attention; attentional feature fusion

1 引言

从遥感图像中提取道路信息一直是遥感领域的研

究热点,在许多应用中具有极其巨大的价值,如自动驾驶、城市交通规划、紧急风险管理、地理信息更新等。但由于遥感图像道路背景复杂,存在障碍物(阴影、树

收稿日期: 2022-09-26; 修回日期: 2022-10-27; 录用日期: 2022-11-24; 网络首发日期: 2023-01-04

基金项目: 国家自然科学基金(41871226)、国家重点研发计划政府间国际科技创新合作(2021YFE0194700)、重庆市教委重点合作项目(HZ2021008)、重庆市高技术产业重大产业技术研发项目(D2018-82)

通信作者: *taoyx@cqupt.edu.cn

木、建筑等)的干扰,从高分辨遥感图像中准确、完整地提取道路信息仍然极具挑战^[1-3]。

近年来,深度学习由于具有良好的性能和泛化能力,已经被越来越多地应用于从高分辨率遥感图像中提取信息^[4-5]。Mnih等^[6]首次在道路提取领域使用深度学习技术,提出一种使用受限玻耳兹曼机从高分辨率遥感图像中检测道路的方法。随后,Long等^[7]提出全卷积神经网络(FCN),通过使用标准卷积层代替全连接层,实现由简单图像分类到像素级的分类,保留了原始输入图像的空间信息,极大地改进了分割效果。此后,基于FCN的应用越来越多,尤其是在道路提取方面。Zhong等^[8]应用FCN结构从高分辨率图像中提取建筑物和道路,综合考虑学习率、输入图像大小等超参数对提取结果的影响,并确定了最佳超参数配置,显著提高了提取的准确率。尽管基于FCN的方法在道路提取中取得了较好的结果,但由于在FCN中连续的下采样操作会导致图像部分空间信息丢失,其输出结果很难恢复到输入图像的分辨率。针对这个问题,研究者提出编解码结构模型,通过在解码器部分使用跳跃连接方式连接多级特征逐渐恢复目标细节和相应空间维度。Cheng等^[9]提出一种级联的编解码网络CasNet,用于遥感图像道路和中心线的提取。受深度残差学习^[10]和U-Net^[11]的启发,Zhang等^[12]提出深度残差U-Net用于道路提取,通过残差结构简化深层网络的训练,减少参数量,同时网络中大量跳跃连接促进了信息传播,实现了更好的道路提取效果。为进一步获取详细的空间信息,Zhou等^[13]提出D-LinkNet,在网络中使用结合级联模式和并行模式的空洞卷积来扩大卷积核的感受野,保持细节信息的同时融合了多尺度特

征,提高了提取的准确性。由于D-LinkNet的成功,使用结合编解码结构和空洞卷积的网络结构来提取道路信息得到进一步研究^[14-17]。Lan等^[18]提出一种基于全局上下文信息的空洞卷积神经网络GC-DCNN,该网络通过使用残差扩张块来构建编码网络,简化深层网络的训练,进一步扩大感受野,获得具有多尺度上下文信息的高分辨率特征。

尽管基于编解码结构和空洞卷积的网络能够通过提取多尺度特征来获得全局上下文信息、提高道路提取的准确性,但仍存在一些问题。首先,目前基于编解码结构的方法在提取目标特征方面仍存在不足,导致上下文信息考虑不完整;其次,基于空洞卷积的方法也存在局限性,单个卷积层根据其卷积核的大小仅能对部分区域进行特征提取,且容易丢失小尺度目标。因此,为了有效解决以上问题、进一步提高道路提取的准确性,本文在U-Net基础上,提出一种针对遥感图像道路特点的网络,通过在编码网络使用注意力残差学习单元提升网络对道路特征的提取能力,通过在桥接网络引入注意力增强的空洞空间金字塔池化模块来提升网络对不同尺度道路信息的提取能力。最后,通过实验验证所提方法的性能并选取其他经典网络进行对比分析。

2 HSA-UNet

2.1 总体框架

以U-Net为基础并对其进行改进,提出hybrid scale attention-UNet(HSA-UNet)模型,其整体结构如图1所示,主要由编码网络、桥接网络和解码网络这3部分组成。

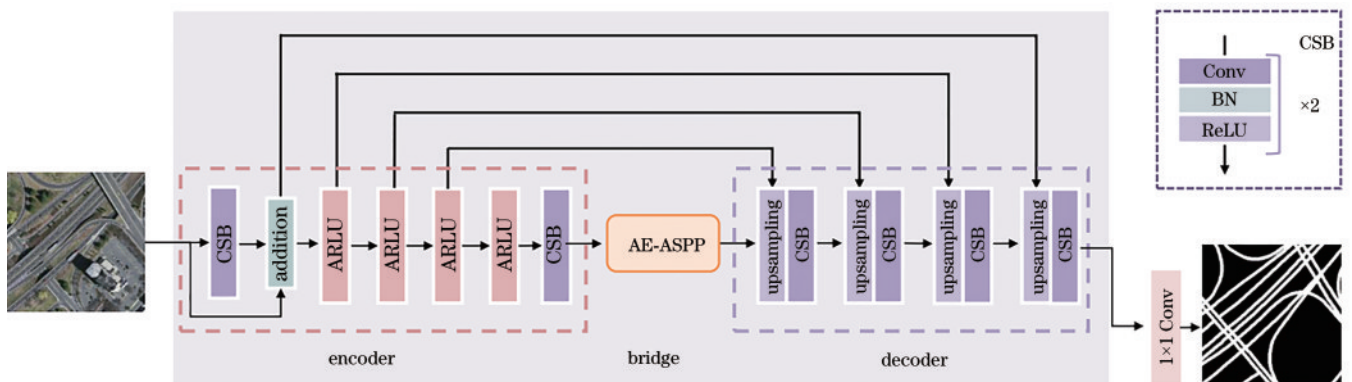


图1 HSA-UNet整体结构

Fig. 1 Overall structure of HSA-UNet

编码网络主要包括2个卷积序列块(CSB)和4个注意力残差学习单元(ARLU)。CSB的具体结构如图1右上角CSB模块所示。每个CSB包括2个 3×3 卷积层,每个卷积层后面都有1个批量归一化层(BN)和1个ReLU激活函数层。输入的RGB图像先经过第1个CSB转化为初始高维特征,然后利用4个连续的

ARLU来捕获图像的多层次多尺度特征,其中,每个ARLU包含1个池化层操作,每经过1个ARLU,特征图通道数加倍、尺寸减半。最后经过第2个CSB得到包含丰富语义信息的高维度特征。为进一步学习多尺度特征,同时缓解道路狭窄、复杂、跨度大等特性的影响,在桥接网络加入注意力增强的空洞空间金字塔池

化(AE-ASPP)模块。解码网络包括 4 个上采样块,用于将特征图恢复到输入图像大小。每个上采样块先通过转置卷积将高层特征的通道数和图像大小变为与编码网络对应低层特征一致,再将高层特征和低层特征拼接后的结果作为 CSB 的输入。每经过一个上采样块,特征图通道数减半。最后通过一个 1×1 卷积层将高维特征转化为两通道特征,得到最终预测输出。

2.2 注意力残差学习单元

研究表明,更深的网络结构会取得更好的效果,但当网络增加到一定深度后,单纯增加网络深度会面临梯度消失的问题。针对这一现象,He 等^[10]提出 ResNet,其基本思想是残差映射易于优化,因此 ResNet 使用恒等映射分支来跳过卷积层,以形成残差单元,通过恒等映射分支为梯度在反向传播期间提供一条可替代路径。相比普通神经单元[图 2(a)],残差单元使深度网络易于训练,提高了训练效率,很大程度上解决了深层网络的退化问题。受残差结构启发,使用残差单元代替普通神经单元作为基本块,如图 2(b)所示。同时,由于残差结构中的恒等映射分支和残差映射分支通过元素相加方式融合,这种特征融合方式仅仅只为特征分配固定权重,没有考虑特征内容的变化,融合效率低。受 Dai 等^[19]的启发,设计了 ARLU,如图 2(c)所示,采用注意力特征融合(AFF)模块融合残差单元的两个分支,AFF 具体结构如图 3 所示。

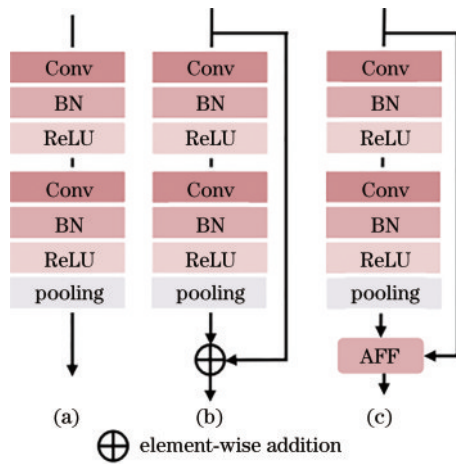


图 2 3 种结构对比。(a)普通单元;(b)残差单元;(c)ARLU

Fig. 2 Comparison of three structures. (a) Plain unit;

(b) residual unit; (c) ARLU

AFF 首先将恒等映射分支输出 $X \in \mathbb{R}^{C \times H \times W}$ 和残差映射分支输出 $Y \in \mathbb{R}^{C \times H \times W}$ 通过元素相加方式融合,把融合后的结果 A 输入多尺度通道注意力模块(MS-CAM)中,其中,一个分支 $G(A)$ 使用全局平均池化(GAP)来获取全局上下文信息,另一个分支 $L(A)$ 通过点卷积(PW)方式获取局部细节信息。通过元素相加将 $G(A)$ 和 $L(A)$ 两分支融合。令 MS-CAM 的输出为 $M(A)$,则:

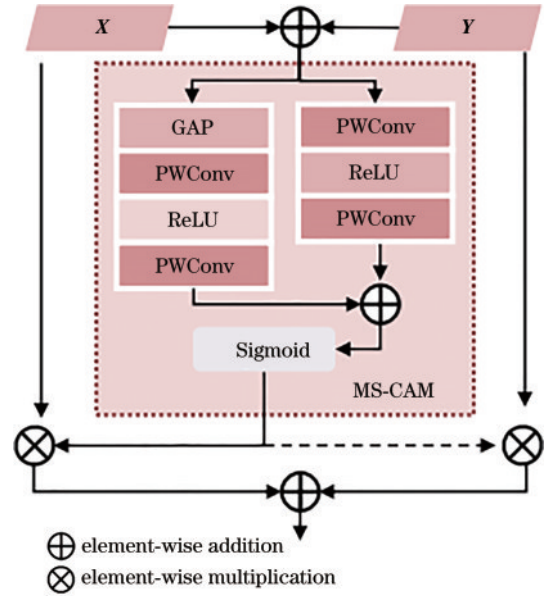


图 3 AFF 模块

Fig. 3 AFF module

$$G(A) = \text{BN} \left\{ \text{PW}_2 \left\{ \delta \left\{ \text{BN} \left[\text{PW}_1 [g(A)] \right] \right\} \right\} \right\}, \quad (1)$$

$$L(A) = \text{BN} \left\{ \text{PW}_2 \left\{ \delta \left\{ \text{BN} [\text{PW}_1 (A)] \right\} \right\} \right\}, \quad (2)$$

$$M(A) = \sigma [G(A) + L(A)], \quad (3)$$

式中: $g(\cdot)$ 表示全局平均池化操作; $\text{PW}_1(\cdot)$ 和 $\text{PW}_2(\cdot)$ 分别表示卷积核为 $C \times \frac{C}{r} \times 1 \times 1$ 和 $\frac{C}{r} \times C \times 1 \times 1$ 的点卷积操作, r 表示通道缩减率; $\text{BN}(\cdot)$ 表示批量归一化操作; $\delta(\cdot)$ 表示 ReLU 激活函数; $\sigma(\cdot)$ 表示 Sigmoid 函数。因此, AFF 的输出 $F(X, Y)$ 为

$$F(X, Y) = M(X + Y) \otimes X + [1 - M(X + Y)] \otimes Y, \quad (4)$$

式中: \otimes 表示元素乘法。

MS-CAM 通过在注意力模块内使用点卷积将局部信息添加到全局信息中,让网络可以沿通道维度对特征图进行多个尺度上的信息提取,同时保持网络的轻量化。ARLU 利用 AFF 融合残差结构的两个分支,以一种尺度感知的方式动态地、自适应地融合接收到的特征,弥补了不同分支特征之间存在的语义差距,增强了网络对遥感图像全局和局部信息的学习能力,提升了网络训练效率和道路识别准确度。

2.3 注意力增强的空洞空间金字塔池化

ASPP^[20] 中并行的空洞卷积层相当于多个不同的接收域,可以有效地在多个尺度上并行采样。但是,为避免带来太多额外计算量,其卷积核内除 9 个有效点之外,其余所有点都用零填充,这些零填充位置相对应的局部细节特征无法收集^[21]。其次,ASPP 中的卷积层均使用方形卷积核进行特征学习,但是遥感图像中的道路细长、复杂,且只占据整幅图像的小部分,这意味着 ASPP 对道路的提取能力有限。针对以上问题,设计了注意力增强的 ASPP,具体结构如图 4 所示。

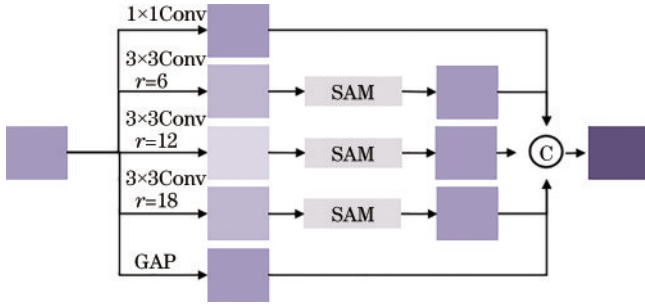


图4 AE-ASPP模块

Fig. 4 AE-ASPP module

AE-ASPP结合ASPP和空间注意力机制^[22-24],在ASPP基础上将空间注意力机制分别应用于不同尺度特征上,以达到密集特征采样效果。在编码网络最终的输出上设置4个并行卷积分支和1个全局平均池化分支以学习高层多尺度特征,4个并行卷积分支包括1个 1×1 卷积分支和3个 3×3 的空洞卷积分支。空洞卷积层的扩张率分别设置为6、12、18,在3个空洞卷积分支上分别应用条带注意力模块(SAM)^[25]。SAM具体结构如图5所示。

SAM使用条带池化在垂直方向上部署一个长条形状的池化核对全局信息进行编码,并沿水平方向进

一步收集每个像素和条带池化核之间的关联程度信息。通过SAM,特征图中每个像素都与不同列空间中的像素关联起来,有效解决了ASPP的“棋盘效应”问题,在一定程度上缓解了道路自然特性带来的影响。令 $F_{in} \in \mathbb{R}^{C \times H \times W}$ 和 $F_{out} \in \mathbb{R}^{C \times H \times W}$ 分别代表SAM的输入和输出,则特征图A可表示为

$$A = \text{Softmax} \left\{ R_1 \left[\text{Conv} \left(F_{in} \right) \right]^T \times R_2 \left\{ S \left[\text{Conv} \left(F_{in} \right) \right] \right\} \right\}, \quad (5)$$

式中: $\text{Softmax}(\cdot)$ 表示激活函数; $R_1(\cdot)$ 表示将特征重塑为 $\mathbb{R}^{C \times N}$, $N = H \times W$; $R_2(\cdot)$ 表示将特征重塑为 $\mathbb{R}^{C \times M}$, $M = 1 \times W$; $S(\cdot)$ 表示条带池化操作,将特征重塑为 $\mathbb{R}^{C \times 1 \times W}$; $\text{Conv}(\cdot)$ 表示 1×1 卷积操作; T 表示特征转置。将A与 F_{in} 结合以生成注意力增强的特征图,最后与输入特征 F_{in} 进行元素相加操作得到最终输出 F_{out} ,可表示为

$$F_{out} = F_{in} + R_3 \left\{ R_2 \left\{ S \left[\text{Conv} \left(F_{in} \right) \right] \right\} \times A^T \right\}, \quad (6)$$

式中: $R_3(\cdot)$ 表示将特征重塑为 $\mathbb{R}^{C \times H \times W}$ 。由于 1×1 卷积分支和全局平均池化分支相当于分别使用最小和最大的感受野来保持输入图像的固有特征,因此在AE-ASPP中完整保留这2个分支。最后,将5个分支的输出拼接融合,作为解码网络输入。

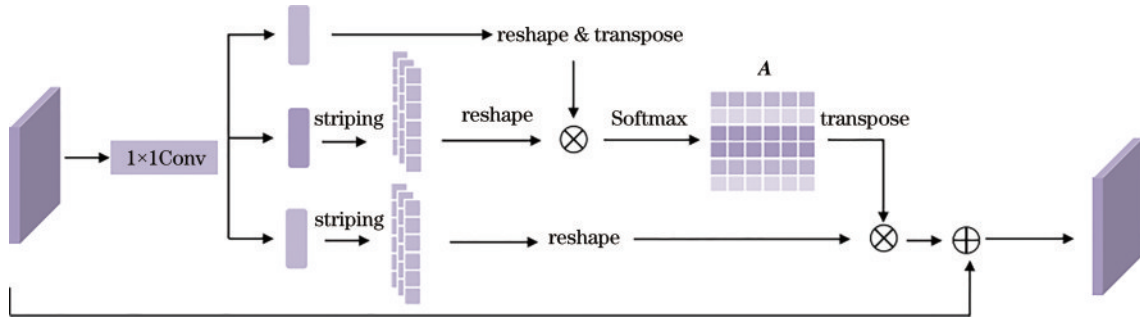


图5 SAM

Fig. 5 SAM

2.4 损失函数

在遥感图像中,道路是一段细长狭窄的区域,且只占整幅图像的小部分,道路和背景之间存在样本类别的高度不平衡问题。在图像分割领域,交叉熵(cross-entropy)损失函数对分割结果的每个像素进行评估,如果图像存在类别不平衡问题,会导致较多类别的样本出现过度拟合。因此,从遥感图像提取道路时,网络会偏向背景学习,这会降低网络提取道路目标的能力。Dice损失函数把一个类别的所有像素作为一个整体,并计算两个类别的交集在整体的比例,不受大量背景像素的影响,在样本不平衡情况下可达到更好的效果。但Dice损失函数对噪声敏感,可能会忽略边界信息,导致边界分割不佳。因此,结合二元交叉熵(BCE)损失函数与Dice损失函数,设计了复合型损失函数对模型损失进行计算。BCE损失函数和Dice损失函数计

算公式分别如下:

$$L_{BCE} = -\frac{1}{N} \sum_{i=1}^N \left[g_i \log(p_i) + (1 - g_i) \log(1 - p_i) \right], \quad (7)$$

$$L_D = 1 - \frac{\sum_{i=1}^N p_i g_i}{\sum_{i=1}^N p_i^2 + \sum_{i=1}^N g_i^2}, \quad (8)$$

式中: N 表示像素总数; g_i 表示像素 i 的真实标签值; p_i 表示像素 i 预测值。最终的复合损失函数定义如下:

$$L = L_{BCE} + L_D. \quad (9)$$

3 实验结果与分析

3.1 数据集

实验使用Massachusetts道路数据集^[26],该数据集是世界上最大的公开道路数据集。该数据集包含

1171 张图片, 每张图片大小为 $1500 \text{ pixel} \times 1500 \text{ pixel}$, 空间分辨率为 1.2 m 。其中, 包括训练图像 1108 张、验证图像 14 张、测试图像 49 张。由于数据集中图片数量较少, 无法支撑模型训练, 对数据集进行剪裁以增加样

本数量。设置一个 256×256 区域对原始数据集所有图像以及相应标签进行剪裁, 最终得到 22707 张道路图像以及对应标签图像, 如图 6 所示, 并将其按照 8:1:1 的比例划分为训练集、验证集和测试集。



图 6 训练样本与标签

Fig. 6 Training samples and labels

3.2 实验设置

所有实验基于 Windows 10 操作系统, GPU 为 Tesla-V100-SXM2 (32 GB), 使用 PyTorch 深度学习框架搭建网络模型。采用 Adam 优化器优化权重更新, 初始学习率设置为 0.001, 模型迭代次数为 80, batch size 大小为 16。

3.3 评价指标

为检验所提方法的有效性, 使用准确率 ($R_{\text{precision}}$)、召回率 (R_{recall})、F1 分数 (S_{F1}) 和交并比 (R_{IoU}) 来对网络性能进行评估, 计算公式如下:

$$R_{\text{precision}} = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FP}}}, \quad (10)$$

$$R_{\text{recall}} = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FN}}}, \quad (11)$$

$$S_{\text{F1}} = \frac{2 \times R_{\text{precision}} \times R_{\text{recall}}}{R_{\text{precision}} + R_{\text{recall}}}, \quad (12)$$

$$R_{\text{IoU}} = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FP}} + N_{\text{FN}}}, \quad (13)$$

式中: $R_{\text{precision}}$ 表示正确预测为道路的像素占所有预测为道路的像素的比值; R_{recall} 表示正确预测为道路的像素占道路总像素的比值; S_{F1} 表示 F1 分数; R_{IoU} 表示交并比; N_{TP} 表示正确预测为道路的像素数量; N_{FP} 表示将背景错误预测为道路的像素数量; N_{FN} 表示将道路错误预测为背景的像素数量。准确率和召回率分别表示道路提取的准确性和完整性; F1 分数表示准确率和召回率的调和均值; 交并比表示预测值和真实值两个集合的交集与并集之比。

3.4 消融实验

为验证 HSA-UNet 模型中改进模块的有效性, 分别对 ARLU 和 AE-ASPP 进行消融实验, 结果如表 1

所示。选取 U-Net 模型作为实验的基准模型 (No. 1), 其 R_{IoU} 值为 62.41%。将 ARLU (No. 2) 和 AE-ASPP (No. 3) 分别加入基准模型后, R_{IoU} 值分别增加到 64.65% 和 64.24%。同时加入 ARLU 和 AE-ASPP (No. 4) 后, R_{IoU} 值达到 65.60%, 相比基准模型提高了 3.19 个百分点。从评价指标 R_{IoU} 可以看出, 所提两个模块对遥感图像道路提取精度均具有较好的提升。

表 1 消融实验评价结果对比

Table 1 Comparison of evaluation result of ablation experiment

| No. | BASE | ARLU | AE-ASPP | $R_{\text{IoU}} / \%$ |
|-----|------|------|---------|-----------------------|
| 1 | ✓ | | | 62.41 |
| 2 | ✓ | ✓ | | 64.65 |
| 3 | ✓ | | ✓ | 64.24 |
| 4 | ✓ | ✓ | ✓ | 65.60 |

3.5 所提网络与其他网络对比实验

为验证所提 HSA-UNet 的有效性, 与 U-Net、SegNet、ResUnet、DeepLabV3+、D-LinkNet、NL-LinkNet^[27] 和 GC-DCNN^[18] 进行对比。表 2 列出了不同网络模型在 Massachusetts 道路数据集上的评价指标数据。与 $R_{\text{precision}}$ 、 R_{recall} 两个指标相比, R_{IoU} 是一个总体度量指标, 考虑道路像素错分为背景的现象, 评估更加全面。表 2 中, 粗体显示的值是最好的。由表 2 可以看出: 尽管 GC-DCNN 和 NL-LinkNet 分别在 $R_{\text{precision}}$ 指标和 R_{recall} 指标上取得最好的效果; 但在 R_{IoU} 指标上的结果比所提 HSA-UNet 分别低 1.34 个百分点和 2.09 个百分点; 在 S_{F1} 指标上相较于所提模型也分别低 0.99 个百分点和 1.56 个百分点。所提 HSA-UNet 在 $R_{\text{precision}}$ 指标和 R_{recall} 指标上都取得了相对令人满意的效果, 而且在 S_{F1} 和 R_{IoU} 指标上结果最好, 较 U-Net 分别提高

表 2 不同网络道路提取结果对比

Table 2 Comparison of road extraction result of different networks

| Model | $R_{Precision}$ | R_{Recall} | S_{F1} | R_{IoU} |
|------------|-----------------|--------------|--------------|--------------|
| U-Net | 80.71 | 73.34 | 76.81 | 62.41 |
| SegNet | 75.82 | 74.09 | 74.95 | 59.93 |
| ResUNet | 81.42 | 72.53 | 76.72 | 62.23 |
| DeepLabV3+ | 80.01 | 70.46 | 74.94 | 59.92 |
| D-LinkNet | 75.35 | 79.67 | 77.45 | 63.20 |
| NL-LinkNet | 75.27 | 80.26 | 77.67 | 63.51 |
| GC-DCNN | 82.64 | 74.28 | 78.24 | 64.26 |
| HSA-UNet | 80.50 | 77.99 | 79.23 | 65.60 |

2.42个百分点和3.19个百分点。因此,HSA-UNet在整体上要优于其他网络模型。

为了更直观地对比不同模型的道路提取效果,选

取实验指标最好的D-LinkNet、NL-LinkNet以及GC-DCNN这3个网络的道路提取结果,如图7所示。

由图7标注的矩形框区域可以看出,D-LinkNet、NL-LinkNet和GC-DCNN的预测结果较为粗糙,预测图中出现较多的孤立点,道路的断裂现象明显,在提取树木遮挡的道路上效果不佳,其中,GC-DCNN在提取此类道路方面表现最差、丢失程度最高。所提模型道路预测结果更加平滑,没有出现孤立点。在道路被树木、建筑等障碍物部分或完全遮挡的情况下,提取结果要更准确、完整。通过对D-LinkNet、NL-LinkNet以及GC-DCNN这3种网络提取结果的分析,可以发现所提HSA-UNet可以更有效、更全面地提取道路,能够准确地分割道路边缘,并且有效解决树木、建筑物等背景特征带来的干扰,最终提取的道路目标完整度更高,与标签影像有较高的相似度。

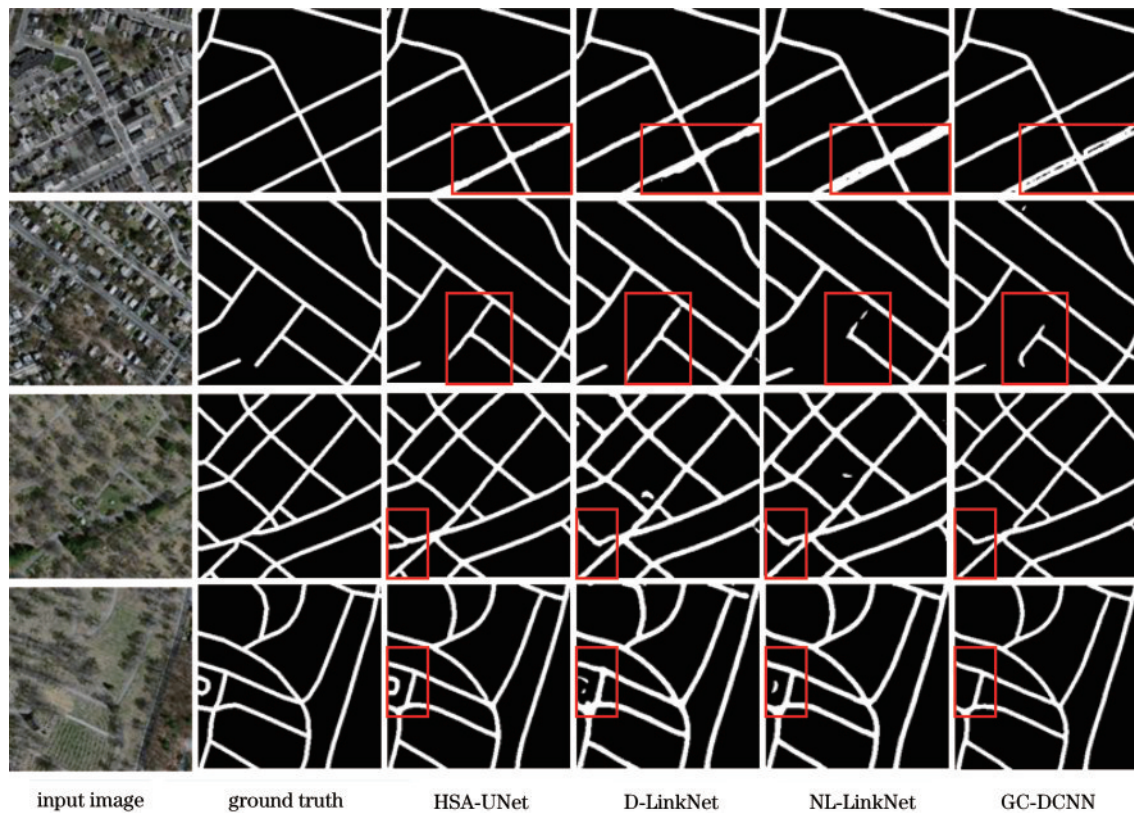


图7 所提网络与其他网络的道路预测图

Fig. 7 Road prediction diagrams of proposed network and other networks

4 结 论

对遥感图像道路目标提取进行研究,旨在提高道路提取的准确性和完整性。在U-Net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。首先,设计ARLU构建编码网络:一方面,通过残差结构促进信息在高低层之间流动,简化网络训练;另一方面,利用注意力特征融合模块融合残差结构的两个分支,通过在注意力模块内使用点卷积来聚合多尺度的特征,克服不同层次特征之间语义和尺度不一致问题。

其次,针对遥感图像中道路尺度的多样性和自然属性的特殊性,设计AE-ASPP构建桥接网络,通过在不同尺度的特征上使用条带注意力模块来增强网络对道路特征的识别效果,同时防止不相关区域的干扰。实验结果表明,所提方法具有更好的性能,所提取的道路结构更加完整。但是所提模型仍需做进一步的改进,提升其在不同场景下更为复杂的背景信息中提取道路的能力,在保证准确率的前提下进一步优化模型结构,下一步可结合更为前沿完善的深度学习方法进行研究。

参 考 文 献

- [1] Qi J, Tao C, Wang H, et al. Spatial Information Inference Net: Road Extraction Using Road-Specific Contextual Information[C]//2019 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), July 28-August 2, Yokohama, Japan. New York: IEEE Press, 2019: 9748-9481.
- [2] 戴激光, 王杨, 杜阳, 等. 光学遥感影像道路提取的方法综述[J]. 遥感学报, 2020, 24(7): 804-823.
Dai J G, Wang Y, Du Y, et al. Development and prospect of road extraction method for optical remote sensing image[J]. Journal of Remote Sensing, 2020, 24(7): 804-823.
- [3] 吴强强, 王帅, 王彪, 等. 空间信息感知语义分割模型的高分辨率遥感影像道路提取[J]. 遥感学报, 2022, 26(9): 1872-1885.
Wu Q Q, Wang S, Wang B, et al. Road extraction method of high-resolution remote sensing image on the basis of the spatial information perception semantic segmentation model[J]. National Remote Sensing Bulletin, 2022, 26(9): 1872-1885.
- [4] 高慧, 阎晓东, 张衡, 等. 基于 Res2Net 的多尺度遥感影像海陆分割方法[J]. 光学学报, 2022, 42(18): 1828004.
Gao H, Yan X D, Zhang H, et al. Multi-scale Sea-land segmentation of remote sensing image method based on Res2Net[J]. Acta Optica Sinica, 2022, 42(18): 1828004.
- [5] 刘金香, 班伟, 陈宇, 等. 融合多维度 CNN 的高光谱遥感图像分类算法[J]. 中国激光, 2021, 48(16): 1610003.
Liu J X, Ban W, Chen Y, et al. Multi-dimensional CNN fused algorithm for hyperspectral remote sensing image classification[J]. Chinese Journal of Lasers, 2021, 48(16): 1610003.
- [6] Mnih V, Hinton G E. Learning to detect roads in high-resolution aerial images[M]//Daniilidis K, Maragos P, Paragios N. Computer vision-ECCV 2010. Lecture notes in computer science. Heidelberg: Springer, 2010, 6316: 210-223.
- [7] Long J, Shelhamer E, Darrell T. Fully Convolutional Networks for Semantic Segmentation[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA, USA. New York: 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, July 10-15, 2016, Beijing, China. New York: IEEE Press, 2016: 1591-1594.
- [9] Cheng G L, Wang Y, Xu S B, et al. Automatic Road detection and centerline extraction via cascaded end-to-end convolutional neural network[J]. IEEE Transactions on Geoscience and Remote Sensing, 2017, 55(6): 3322-3337.
- [10] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, July 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 770-778.
- [11] Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation[M]//Navab N, Hornegger J, Wells W M, et al. Medical image computing and computer-assisted intervention-MICCAI 2015. Lecture notes in computer science. Cham: Springer, 2015, 9351: 234-241.
- [12] Zhang Z X, Liu Q J, Wang Y H. Road extraction by deep residual U-net[J]. IEEE Geoscience and Remote Sensing Letters, 2018, 15(5): 749-753.
- [13] Zhou L C, 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 Workshops (CVPRW), June 18-22, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 192-1924.
- [14] Wu Q Q, Luo F, Wu P H, et al. Automatic Road extraction from high-resolution remote sensing images using a method based on densely connected spatial feature-enhanced pyramid[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2020, 14: 3-17.
- [15] 顾清华, 薛步青, 卢才武, 等. 基于 D-LinkNet 网络的露天矿道路智能识别与路网建模[J]. 煤炭学报, 2020, 45(S2): 1100-1108.
Gu Q H, Xue B Q, Lu C W, et al. Road intelligent recognition and road network modeling of open pit based on D-LinkNet network[J]. Journal of China Coal Society, 2020, 45(S2): 1100-1108.
- [16] He H, Yang D F, Wang S C, et al. Road extraction by using atrous spatial pyramid pooling integrated encoder-decoder network and structural similarity loss[J]. Remote Sensing, 2019, 11(9): 1015.
- [17] Mei J, Li R J, Gao W, et al. CoANet: connectivity attention network for road extraction from satellite imagery[J]. IEEE Transactions on Image Processing, 2021, 30: 8540-8552.
- [18] Lan M, Zhang Y P, Zhang L F, et al. Global context based automatic road segmentation via dilated convolutional neural network[J]. Information Sciences, 2020, 535: 156-171.
- [19] Dai Y M, Gieseke F, Oehmcke S, et al. Attentional feature fusion[C]//2021 IEEE Winter Conference on Applications of Computer Vision, January 3-8, 2021, Waikoloa, HI, USA. New York: IEEE Press, 2021: 3559-3568.
- [20] Chen L C, Papandreou G, Kokkinos I, et al. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(4): 834-848.
- [21] Lian X H, Pang Y W, Han J G, et al. Cascaded hierarchical atrous spatial pyramid pooling module for semantic segmentation[J]. Pattern Recognition, 2021, 110: 107622.
- [22] Fu J, Liu J, Tian H J, et al. Dual attention network for scene segmentation[C]//2019 IEEE/CVF Conference on

- Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 3141-3149.
- [23] Hou Q B, Zhang L, Cheng M M, et al. Strip pooling: rethinking spatial pooling for scene parsing[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 14-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 4002-4011.
- [24] Zhao X Q, Zhang L H, Pang Y W, et al. A single stream network for robust and real-time RGB-D salient object detection[EB/OL]. (2020-07-14) [2022-05-06]. <https://arxiv.org/abs/2007.06811>.
- [25] Song Q, Mei K F, Huang R. AttaNet: attention-augmented network for fast and accurate scene parsing[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2021, 35(3): 2567-2575.
- [26] Mnih V. Machine learning for aerial image labeling[D]. Toronto: University of Toronto, 2013.
- [27] Wang Y, Seo J, Jeon T. NL-LinkNet: toward lighter but more accurate road extraction with nonlocal operations[J]. IEEE Geoscience and Remote Sensing Letters, 2022, 19: 1-5.