

文章编号 1004-924X(2023)22-3371-12

面向遥感建筑物提取的轻型多尺度差异网络

李国燕¹, 武海苗¹, 董春华^{2*}, 刘毅¹

(1. 天津城建大学 计算机与信息工程学院, 天津 300384;

2. 天津城建大学 地质与测绘学院, 天津 300384)

摘要:针对高分辨率遥感影像中建筑物形状多样、大小不一引起建筑物提取精度低及传统分割模型存在参数量大等问题,提出一种基于编码-解码的轻型多尺度差异网络 LMD-Net(Lightweight Multi-scale Difference Network)。首先,为了避免单一的特征处理单元堆叠使得模型性能弱化而产生无效参数,通过融合编解码结构的功能差异性,设计出一种轻型差异模型优化性能。其次,引入一种多尺度膨胀感知模块(Multi-Scale Dilation Perception, MSDP)来增强网络捕捉多尺度目标特征的能力。最后,通过双融合机制有效聚合深层跳跃连接和深层解码器两组的特征信息,从而实现增强解码器的特征恢复能力。为验证轻型多尺度差异网络 LMD-Net 的有效性和适用性,以开源 WHU building dataset 数据集作为数据源,对 LMD-Net 网络与常用语义分割网络及近年相关文献研究成果进行了精度、效率方面的评估实验。结果表明:LMD-Net 网络在效率与精度两方面均表现出明显优势,不仅很大程度上减少模型的参数量和计算量,而且交并比、准确率分别提高了 3.23%, 2.57%。表明在基于高分辨率遥感影像建筑物提取领域中,该模型所表现的优势具有良好的城市空间信息库价值。

关键词:高分辨率遥感影像;多尺度;建筑物提取;编码-解码;轻型

中图分类号:P237 **文献标识码:**A **doi:**10.37188/OPE.20233122.3371

Lightweight multi-scale difference network for remote sensing building extraction

LI Guoyan¹, WU Haimiao¹, DONG Chunhua^{2*}, LIU Yi¹

(1. School of Computer and Information Engineering, Tianjin Chengjian University,
Tianjin 300384, China;

2. School of Geology and Mapping, Tianjin Chengjian University, Tianjin 300384, China)

* Corresponding author, E-mail: dch@tcu.edu.cn

Abstract: To address the problem of low accuracy of building extraction in high-resolution remote sensing images due to the diverse shapes and sizes of buildings and large number of parameters in traditional segmentation models, a Lightweight Multi-scale Difference network (LMD-Net) based on encoding-decoding is proposed. First, to avoid the invalid parameters caused by the degraded model performance due to the stacking of single feature processing units, a lightweight differential model is designed to improve the performance by integrating the functional differences of codec structures. Next, a Multi-scale Dilation Perception (MSDP) module is introduced to enhance the ability of the network to capture multi-scale target

收稿日期:2023-06-04;修订日期:2023-07-05.

基金项目:国家自然科学基金资助项目(No. 52178295)

features. Finally, the double fusion mechanism is used to effectively aggregate the feature information of the deep jump connection and deep decoder to enhance the feature recovery ability of the decoder. To verify the validity and applicability of LMD-Net, the open source WHU building dataset was used as the data source to evaluate the accuracy and efficiency of LMD-Net and the common semantic segmentation network as well as the results from recent relevant literature. The results show that LMD-Net has obvious advantages in both efficiency and accuracy, which not only greatly reduces the parameter number and calculation amount of the model but also improves the intersection ratio and accuracy by 3.23% and 2.57%, respectively. Consequently, this model is advantageous in the field of building extraction based on high-resolution remote sensing images to generate an urban spatial information base.

Key words: high resolution remote sensing image; multi-scale; building extraction; coding-decoding; light weight

1 引言

高分辨率遥感影像是地观测目标的重要信息载体,建筑物作为城市空间信息库的重要元素,从高分影像中精确提取的建筑物对土地覆盖分类、城市规划等领域有重要意义^[1-2]。早期提取建筑物主要是人工利用目标建筑物的基础光谱、纹理等特征^[3],然而不同种类的建筑物的几何与纹理信息差异很大,很难获取高精度的建筑物信息。

近年来,基于深度神经网络方法^[4]凭借其强大的特征学习能力在解决各类分割任务中脱颖而出。Long 等^[5]提出的 FCN 网络,该网络通过将引入卷积层从而实现像素级的预测,但其缺乏考虑像素之间的空间关系。Badrinarayanan^[6]提出编解码结构的 SegNet 网络,通过引入池化索引降低内存消耗。Ronnerberger^[7]提出了 U-Net 网络,其主要通过跨层重建的思想将编码器的高级语义信息融入到解码器中,从而提升分割精度。此后,基于编解码的体系结构被逐渐应用于复杂的遥感影像分割任务^[8-9]中。Ma^[10]通过挖掘多尺度信息及学习全局局部特征实现充分提取不同分辨率的建筑物。Shi^[11]提出在 U-Net 中嵌入 FPN 结构提高模型聚合多尺度特征信息的能力。Xu^[12]通过融入胶囊卷积和空间-通道双注意力模块实现增强建筑物高阶特征的表达。Zhang^[13]提出一种语义分割模型 ResUNet,通过在 U-Net 模型中引入高效的残差块结构,有效解决特征信息随网络深度加深而减弱的问题,同时能简化网络的训练,促进信息的传播。Xu^[14]提出

一种融合分形几何特征的 Resnet 网络,通过引入融合分形先验的空洞空间金字塔池化模块和深度可分离卷积注意力融合机制提高分割模型的判别能力。Luo^[15]提出一种 ResUNet+ 网络,通过增强多尺度特征能力提升分割精度。Xu^[16]提出一种改进 ResNet 的模型,通过融合多尺度的遥感建筑物特征从而提升分割精度。Lin^[17]提出一种新型高效网络 ESFNet,旨在以低计算成本和低内存消耗来保持轻微的精度损失。Paszke^[18]提出一种高效神经网络 ENet,其具有高速度低参数的特点,专门为需要低延迟操作的任务所创建。

鉴于此,尝试以 ResUNet 模型为基础,考虑到模型只有学习到充足的全局和局部特征信息及挖掘更多多尺度特征信息,才能充分应对形状各异的建筑物,提升分割精度。本文基于 ResUNet 模型,融入编解码功能差异性、多尺度膨胀感知模块及双融合机制思想设计出一种轻型多尺度差异网络 LMD-Net。从而有效地解决基础模型缺乏考虑局部与全局信息的丰富关联导致建筑物提取精度不高及模型参数量大、不够轻量化等问题,实现优化基础模型的结果。

2 网络及算法

2.1 ResUNet 网络模型

ResUNet^[13] (Deep Residual U-Net) 主要由编-解码网络和跳跃连接结构组成,该网络主要通过应用经典 ResNet^[19] 模型中浅残差单元块搭建模型架构。该模型相比 U-Net 模型有以下几

点不同:第一,ResUNet 通过将 U-Net 原有的裁剪操作去除从而获得更优雅的架构;第二,ResUNet 通过将特征处理单元由普通卷积单元替换成浅残差单元使得网络更能保留目标的精细信息。这种特性的网络不仅能避免普通卷积堆叠导致的信息丢失,更能促进前后向信息的传播,避免退化现象。因此其适用于基于遥感影像的目标提取任务。

2.2 轻型多尺度差异网络(LMD-Net)

为了充分利用残差结构拟合性强的特点和多尺度融合结构增强特征表示的优越性,本文依据编解码结构的功能差异将二者结合提出轻型多尺度差异网络 LMD-Net。该网络是基于编码-解码器结构的端到端建筑物提取网络模型。如图 1 所示,LMD-Net 主要包括编码器、解码器以及跳跃连接部分。首先,依据编解码器结构的功能差异设计出一种轻型差异模型,即编码器使用 BlockA,解码器使用 BlockD;其次,为了使浅层表征图对位置及空间信息更敏感,深层恢复高空

三层和解码器第二层使用多尺度膨胀感知模块,从而加剧网络对不同尺寸建筑物目标的提取能力;最后,为了进一步加深解码结构的特征通道的信息恢复能力,在模型的深跳跃连接后嵌入双融合机制,实现充分融合各尺度信息。

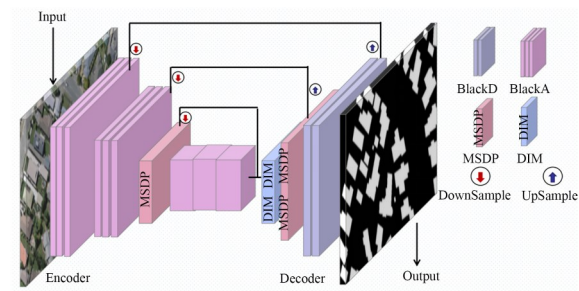


图 1 LMD-Net 的总体框架图

Fig. 1 General framework diagram of LMD-Net

LMD-Net 网络由四层编码器及三层解码器组成。224 尺寸的输入影像经过编码器处理后得到仅有 28 尺寸的高级特征图,最后由三层解码器解析恢复影像空间信息。LMD-Net 网络组件见表 1。其展示各个组件的详细信息。

表 1 LMD-Net 网络组件

Tab. 1 LMD-Net Network Components

Stage	Level	Module type	Filter	Output size
Encode		Input		$3 \times 224 \times 224$
	Level 1	BlockA	$3 \times 3/64$	$64 \times 224 \times 224$
	Level 2	Downsampling	$3 \times 3/128$	$128 \times 112 \times 112$
		BlockA	$3 \times 3/128$	$128 \times 112 \times 112$
	Level 3	Downsampling	$3 \times 3/256$	$256 \times 56 \times 56$
		MSDP	$3 \times 3/256$	$256 \times 56 \times 56$
	Level 4	Downsampling	$3 \times 3/512$	$512 \times 28 \times 28$
		BlockA	$3 \times 3/512$	$512 \times 28 \times 28$
Decode	Level 5	Upsampling	$3 \times 3/256$	$256 \times 56 \times 56$
		DIM	$3 \times 3/256$	$256 \times 56 \times 56$
	Level 6	Upsampling	$3 \times 3/128$	$128 \times 112 \times 112$
		MSDP	$3 \times 3/128$	$128 \times 112 \times 112$
	Level 7	Upsampling	$3 \times 3/64$	$64 \times 224 \times 224$
		BlockD	$3 \times 3/64$	$64 \times 224 \times 224$
		Output	1×1	$1 \times 224 \times 224$

2.3 轻型差异模型

ResUNet 模型主要是通过单一的浅残差单元堆叠而成,这种方式并未考虑编码器解码器的

功能差异,导致编码器不能充分发挥其特征提取的能力和解码器分类效果不佳,其次单一的堆叠网络也使得网络产生很多无用的参数。受目前

轻量化系列结构^[20-21]中的影响,本文引入深度可分离卷积和 1×1 卷积构建一种轻型差异模型,其不仅按编解码结构的功能差异进行设计,而且缩减了很多无效参数,是最终模型轻量化的关键因素。

深度可分离卷积^[22]最大的优势是可大幅度降低卷积参数。如图2所示。可将其分解为深度卷积和逐点卷积两个过程。

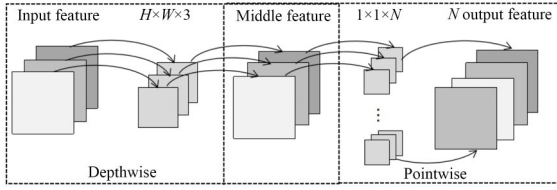


图2 深度可分离卷积过程

Fig. 2 Deeply separable convolution process

假设 $F_h^{\text{in}} \times F_w^{\text{in}} \times C_{\text{in}}$ 尺寸的特征图,经过 C_{out} 个 $F_h^k \times F_w^k$ 卷积操作后,得到 $F_h^{\text{out}} \times F_w^{\text{out}} \times C_{\text{out}}$ 的输出特征图。那么对于标准卷积而言,输出特征图单个像素点需 C_{out} 个卷积核进行点乘操作,而得到 $F_h^{\text{out}} \times F_w^{\text{out}}$ 大小的输出特征图所需要的计算量为:

$$F_h^k \times F_w^k \times F_h^{\text{out}} \times F_w^{\text{out}} \times C_{\text{in}} \times C_{\text{out}}. \quad (1)$$

所需的参数量为:

$$F_h^k \times F_w^k \times C_{\text{in}} \times C_{\text{out}}. \quad (2)$$

而对于深度可分离卷积,首先第一个过程得到输出大小为 $F_h^{\text{out}} \times F_w^{\text{out}} \times C_{\text{in}}$ 的特征图。其次

第二个过程中,通过 C_{out} 个 1×1 卷积核对第一过程特征图加权组合得到输出特征图大小为 $F_h^{\text{out}} \times F_w^{\text{out}}$ 。即深度可分离卷积所消耗的参数量为:

$$F_h^k \times F_w^k \times C_{\text{in}} + C_{\text{in}} \times C_{\text{out}}. \quad (3)$$

所需的计算量为:

$$F_h^k \times F_w^k \times F_h^{\text{out}} \times F_w^{\text{out}} \times C_{\text{in}} + F_h^{\text{out}} \times F_w^{\text{out}} \times C_{\text{in}} \times C_{\text{out}}. \quad (4)$$

相比普通卷积,如式(5)~式(6)所示,深度可分离卷积不仅缩减网络的参数量效果好,而且显著提升模型的训练速度,使网络更趋向于轻量化。

$$(F_h^k \times F_w^k \times F_h^{\text{out}} \times F_w^{\text{out}} \times C_{\text{in}} + F_h^{\text{out}} \times F_w^{\text{out}} \times C_{\text{in}} \times C_{\text{out}}) / (F_h^k \times F_w^k \times F_h^{\text{out}} \times F_w^{\text{out}} \times C_{\text{in}} \times C_{\text{out}}) = \frac{1}{C_{\text{in}} \times C_{\text{out}}} + \frac{1}{C_{\text{in}}}, \quad (5)$$

$$\frac{F_h^k \times F_w^k \times C_{\text{in}} + C_{\text{in}} \times C_{\text{out}}}{F_h^k \times F_w^k \times C_{\text{in}} \times C_{\text{out}}} = \frac{1}{C_{\text{out}}} + \frac{1}{F_h^k \times F_w^k}. \quad (6)$$

由上述分析知,相比普通卷积,深度可分离卷积在压缩模型参数量和计算量方面有明显的优势,而 1×1 卷积所具有的调节通道维度的能力也能进一步缩减模型参数量。基于此理论,本文对比了四种特征处理单元,如图3所示,其中BlockB和BlockD是在BlockA和BlockC的基础上加入深度可分离卷积。ResUNet网络由BlockC单一堆叠而成,本文轻型差异模型的编码部分由BlockA构建,解码部分由BlockD搭建。在实验部分将详细说明四种特征处理单元的性能及轻型差异模型的形成过程。

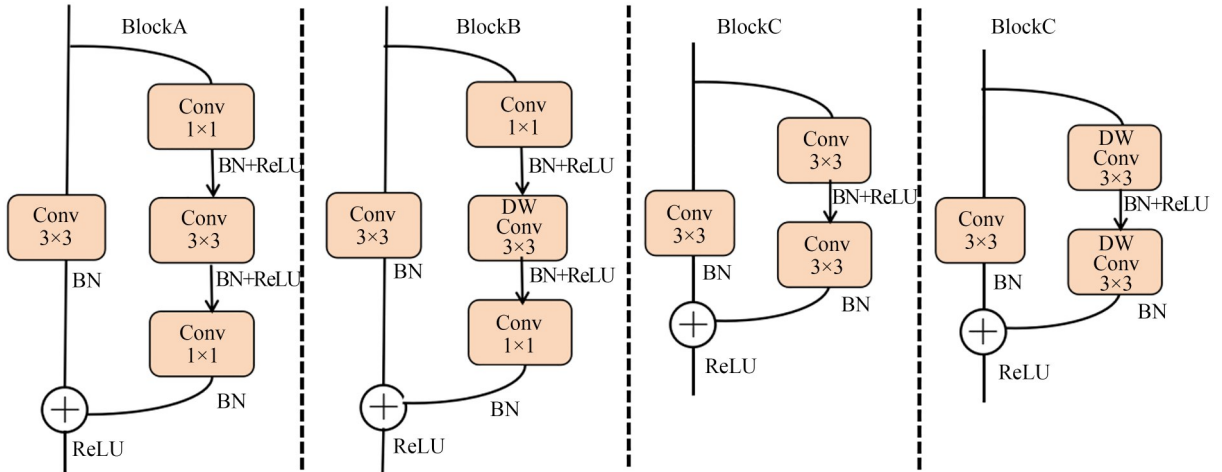


图3 四种特征处理单元

Fig. 3 Four types of feature processing units

2.4 多尺度膨胀感知模块(MSDP)

在遥感影像分割中,增强特征的表示能力是举足轻重的。伴随网络层数的不断加深,深卷积对特征的敏感程度逐渐从局部特征向全局特征的趋势发展。为了能够增强浅卷积层对定位及空间信息的敏感度及深卷积恢复空间信息的表征能力,本文基于膨胀卷积^[23]的思想,提出多尺度膨胀感知模块(Multi-Scale Dilation Perception, MSDP)。

由于网络最浅层涵盖较多粗糙特征,最浅层的并行多尺度策略会加快粗糙信息的传递而减弱网络的性能,因此本文在编码器特征信息较丰富的第三层引入MSDP模块,增强编码器对定位及空间信息的提取。此外,MSDP模块虽能加剧网络融合多尺度特征信息,但沿用连续的特征恢复层更能促进网络恢复空间信息。因此本文在解码器第二层融入MSDP模块,经实验验证连续性有效。

MSDP模块的结构图如图4所示。该模块采用级联采样法,并行三级不同的空洞卷积对特征图采样,使MSDP模块能学习到来自不同感受野的图像特征,所聚合的多个尺度的语义信息能有效地增强深浅卷积的表征能力。其中,三个分支所设置的膨胀系数分别为(1,3,5),所能获得的感受野大小为(3,7,11)。并行多个大感受野分支使模型能有效捕获多尺度的上下文信息,提升对不同尺度的建筑物目标的识别能力。

膨胀卷积的主要优势是扩大感受野不消耗额外的参数量,其可视化效果如图5所示,相比标准卷积,膨胀率为2的卷积核的感知范围明显比

标准卷积的更大,能表达的语义信息也越多。

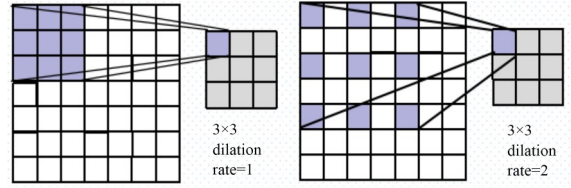


图5 膨胀卷积可视化效果

Fig. 5 Inflated convolution visualization effect

实际上,采用膨胀率的卷积核大小会发生改变,其扩张后的卷积核的尺寸计算公式为:

$$K = K_i + (K_i - 1)(R - 1), \quad (7)$$

式中,第 i 层的卷积核大小 K_i , R 为膨胀率。由公式(7)进而可得第 i 层特征图的感受野。原始影像的感受野记为1,那么第 i 层输出特征图的感受野大小 RF_i 的递推公式为:

$$\begin{cases} RF_0 = 1, \\ RF_1 = K_1, \\ RF_i = RF_{i-1} + (K_i - 1) \prod_{j=1}^{i-1} S_j, i \geq 2 \end{cases}, \quad (8)$$

其中: S_i 为第 i 层的步长, RF_{i-1} 为 $i-1$ 层的感受野。

2.5 双融合机制

原模型的解码器仅通过密集连接融合同层编码器的特征图实现语义信息的融合,融合后的特征图仅通过单一路径提取语义特征,未能充分接收强大的特征提取编码器网络的所有有效信息。因此,本文将在解码器第一深层引入双融合机制(Dual Integration Mechanism, DIM),增强解码器的特征恢复能力。

受2.3小节中四种特征处理单元的对比实验启发,BlockA的提取能力相当高,因此将原仅通过BlockD单一路径恢复特征信息变成融合BlockA与BlockD双路径提取丰富特征信息,如图6所示, X_1 为同层编码器同尺寸的特征图, X_2 为解码器上一层特征图,首先通过跳跃连接实现第一次融合,得到 $56 \times 56 \times 512$ 大小的特征图,其次转入双路径的特征提取路径中,得到蕴含丰富信息的两个 $56 \times 56 \times 256$ 大小的特征图,最后通过第二次融合操作,实现增强特征恢复网络的效果,随后,上采样操作将通道数恢复到下一次所需的数值。

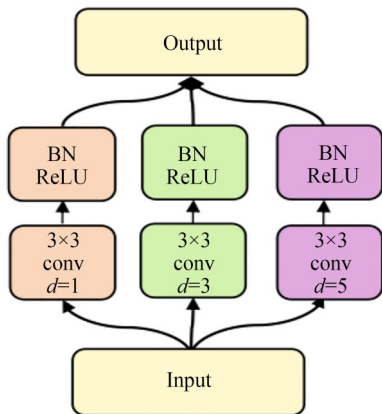


图4 MSDP模块结构图

Fig. 4 MSDP module structure diagram

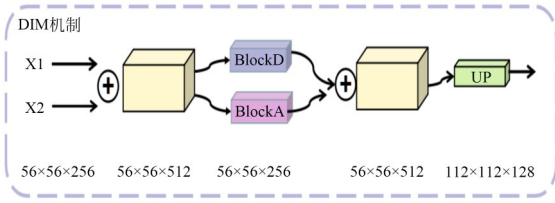


图 6 双融合机制

Fig. 6 Dual integration mechanism

3 实验与分析

3.1 实验数据及实验平台

本实验采用 WHU building dataset^[24] 开源高分遥感影像建筑物数据集,其包含 8 189 幅 RGB 遥感影像,地面空间分辨率为 0.3 m,图片大小为 512×512 。该数据集由 4 736 幅训练集、2 416 幅测试集和 1 036 幅验证集组成。本实验将数据集按照 8:1:1 比例划分,并将其裁剪为 224×224 。考虑到训练数据样本比较少,对训练集进行随机旋转、随机翻转、按照归一化概率选择执行色调或亮度操作扩充数据样本集。

本实验基于 Pytorch 框架,硬件环境为 Intel (R) Core(TM) i7-7700 CPU @3.60 GHZ,内存 16G, NVIDIA GeForce GTX1080(8 G 显存),操作系统为 Windows10。优化器 Adam 代替 SGD 训练网络,初始学习率为 0.001,批次设为 8、迭代次数为 100。如图 7 所示,在训练及验证 LMD-Net 模型中,随着迭代次数增加,损失值逐渐降低。准确率和交并比呈平稳上升趋势。

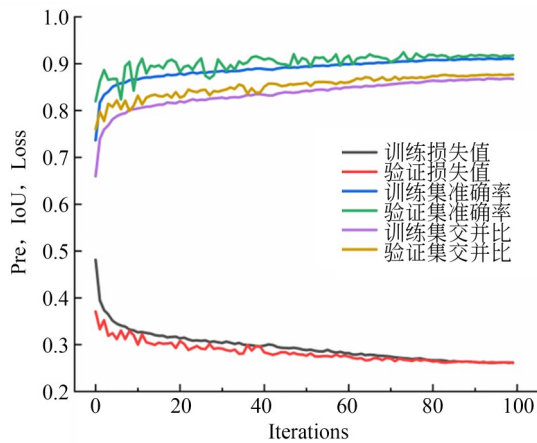


图 7 模型训练、验证阶段指标变化趋势图

Fig. 7 Model training and validation phase metrics change trend graph

3.2 分割性能评价指标

为了对模型的性能做定量评价,本文选用交并比 (Intersection over Union, IoU)、Dice 系数、准确率 (Precision, Pre) 和召回率 (Recall, R) 作为建筑物像素级语义分割的评价指标。具体计算公式如式(9)~式(12):

$$Precision = \frac{TP}{TP + FP}, \quad (9)$$

$$Recall = \frac{TP}{TP + FN}, \quad (10)$$

$$IoU = \frac{TP}{TP + FN + FP}, \quad (11)$$

$$Dice = \frac{2 \times TP}{FP + 2 \times TP + FN}, \quad (12)$$

其中: TP 为建筑物像素数被正确预测的数量, FP 为将非建筑物像素被错误预测为建筑物像素数量, TN 为非建筑物像素数被正确识别的数量, FN 为将建筑物像素预测为建筑物像素数量。

3.3 消融实验结果和分析

3.3.1 验证轻型差异模型的有效性

为了验证轻型差异模型的设计对网络整体性能及效率的影响,本文设计了四组对比实验,该实验是将编码器解码器的特征处理单元依次更换为 BlockA, BlockB, BlockC 和 BlockD 块,通过分析各项指标的数据及结合编码器和解码器的特性进而设计出本文的基础轻型差异模型。实验中使用效率评价指标 Params, Flops 衡量模型的时间复杂度,使用 IoU 衡量模型的性能。

如图 8 展示了在 WHU 数据集上六种特征处

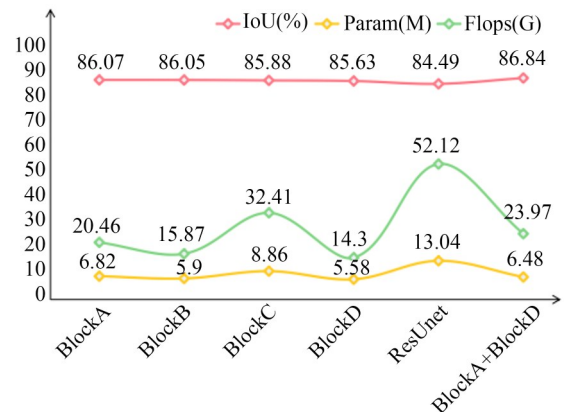


图 8 在 WHU 数据集上六种特征处理单元的分割效率及性能

Fig. 8 Segmentation efficiency and performance of six feature processing units on the WHU dataset

理单元的分割效率及性能。对比 BlockA 与 BlockC, BlockA 与 BlockB 和 BlockC 与 BlockD 可知,深度可分离卷积及 1×1 卷积在降低模型的参数量及计算量方面表现突出;由于编码器特征提取的质量对建筑物分类结果十分关键,而分类解码网络需要轻便快捷。分析图知,BlockA 的交并比 Iou 最高,BlockD 具有最低的参数量和计算量,因此所设计的基础轻型差异模型的编码器特征提取单元选为 BlockA,解码器的特征恢复单元选用 BlockD。对比 BlockA+BlockD 与 ResUnet 模型可得,轻型差异模型交并比提升了 2.35%,降低约 1/2 的参数量,提升大约一半的速率。实现初步优化性能及效率。

3.3.2 验证 MSDP 模块的有效性

为了验证在 3.3.1 小节中所设计的轻型差异模型中添加多尺度膨胀感知 MSDP 模块的有效性,如表 2 详细列出在 WHU 数据集上 MSDP 模块的分割精度。其中,Model 表示不同模块所组

合的模型,Position 体现 MSDP 模块的放置位置,即值为 0,1,2 分别表示在轻型差异模型的第三层、第六层和第七层融入 MSDP 模块。

通过表 2 可以发现,在轻型差异模型中加入 MSDP 组合后精度得到明显提升,观察组 1,2 和组 1,3,单向使用 MSDP 时,两组的 Iou 和 Dice 分别提高了 0.19%,1.23% 和 0.35%,0.40%。对比组 1,4 可知,双向融入 MSDP 效果提升更明显,其 IoU, Recall 和 Dice 提高了 0.42%,0.77% 和 0.3%。这说明组合 MSDP 模块在 LMD-Net 中发挥了积极作用,

能增强模型提取多尺度建筑物的能力。对比组 4,5 可知,沿用连续的特征恢复层能促进 MSDP 模块发挥作用,解码器仅第三层使用 MSDP 模块比解码器第二层使用 MSDP 模块和解码器第三层使用 BlockD 性能下降明显, IoU 降低 0.37%, Recall 下降 0.76%。

表 2 在 WHU 数据集上 MSDP 的分割精度

Tab. 2 Segmentation accuracy of MSDP on WHU dataset

(%)

Group	Model	Position	IoU	Pre	R	Dice
		MSDP				
1	BlockA+BlockD	无	86.84	91.34	90.31	90.80
2	Group1+MSDP	0	87.03	91.35	90.66	90.98
3	Group1+MSDP	1	87.07	91.43	90.71	91.05
4	Group1+MSDP	0,1	87.26	91.13	91.08	91.10
5	Group1+MSDP	0,2	86.89	91.20	90.32	90.84

图 9 显示了在轻型差异模型中添加 MSDP 模块的可视化效果, G1-G4 分别对应表 2 的四组模型,第一列为中型建筑物预测图。第二至五列为小型建筑物预测图,第五列为大型建筑物预测图。观察可知,针对不同尺度的建筑物数据集,在不同位置中所添加的 MSDP 模块,能不同程度地增强模型的提取性能。对比 G1-G3 和 G4, G4 能有效地改善各形状大小的建筑物出现的错检现象,但针对一些极小建筑物提取效果还是有待提升。

3.3.3 验证双融合机制的有效性

为了验证在 3.3.2 小节所改进的模型中添加双融合机制 DIM 的有效性,在 WHU 数据集上对该机制进行实验。如图 10 对比可得,添加 DIM

机制后,网络的性能得到进一步的提升,其中 IoU 及 Pre 提升了 0.46%,0.65%。表明 BIM 能有效增强解码器的像素恢复能力。

3.4 模型实验结果和分析

通过对三组消融实验的分析可知,所构建的轻型差异模型、MSDP 模块以及 DIM 机制对模型整体的性能及效率均有促进作用。本实验主要针对三种典型的语义分割网络 SegNet^[6], U-Net^[7], ResUNet^[13] 和最近几年的两篇参考文献 Enet^[18], ESFNet^[17] 进行效率及性能两个方面的对比实验。

3.4.1 各模型复杂度对比实验

在效率对比实验中,主要使用的性能评估指标有 Params 参数量、MADDs 乘法累加运算总量

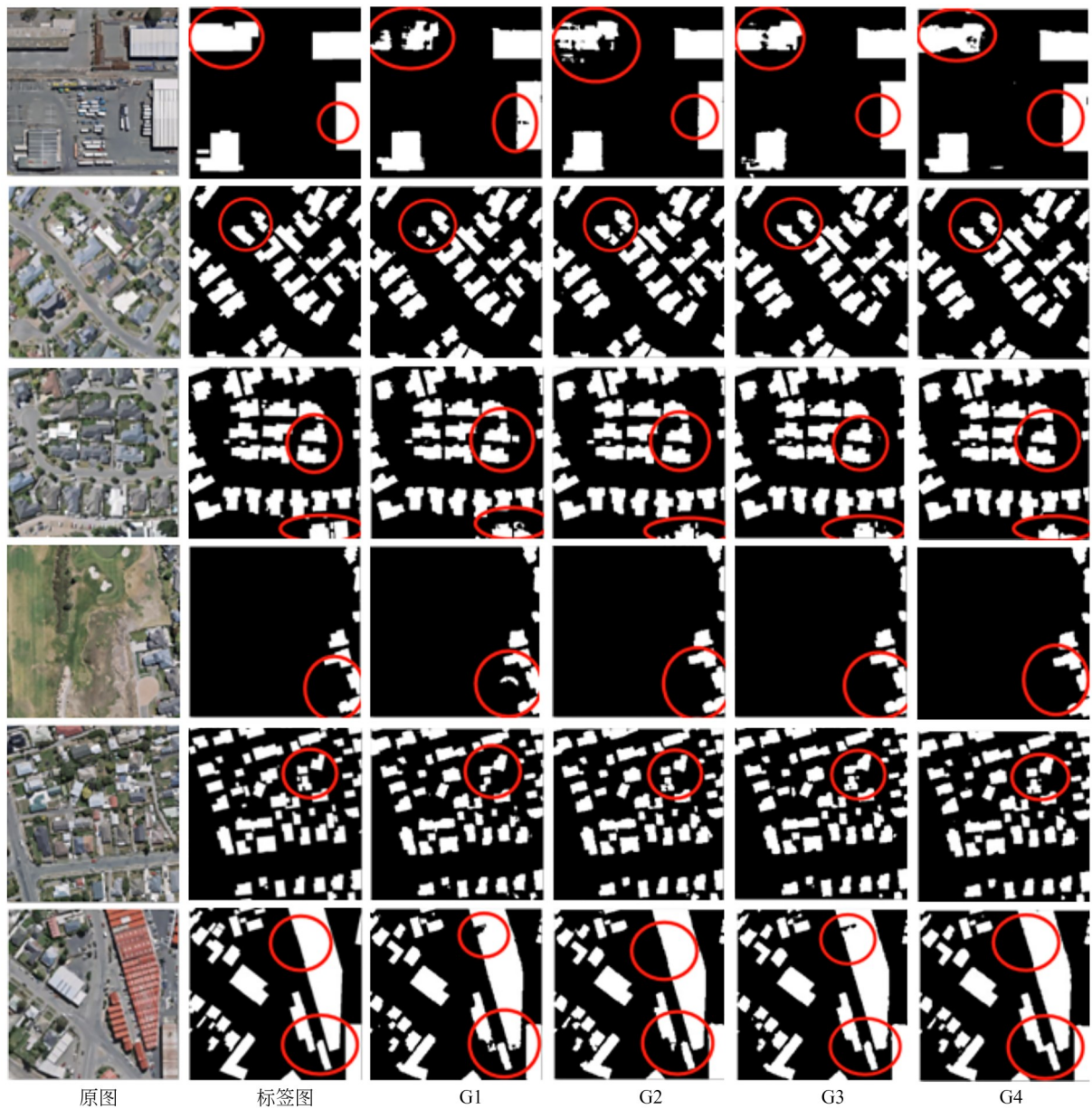


图9 MSDP模块可视化分割结果
Fig. 9 Visual segmentation results of MSDP module

和 Flops 计算量来考核模型的时间及空间复杂度,所有对比实验的实验环境均一致。分析表 3 可知,U-Net 网络的参数量最多,SegNet 网络次之,而 ResUNet 因其引入残差结构等因素,虽参数量相比于 U-Net 网络减少一半多,但计算量却大大增加。而观察组 3 与组 4 的轻量化 ESFNet 及 ENet 网络,两组网络的参数量和计算量都极低,但依据后面实验的验证,两组轻量化网络的性能指标却损失严重,对比组 5 和 6 的 LMD-Net 模型与基础网络 ResUNet 可得,

表 3 各模型的复杂度对比实验
Tab. 3 Experiments comparing complexity of each model

Group	Model	Params/M	MADDs/G	Flops
1	SegNet	29.44	61.36	30.7 G
2	U-Net	32.08	77.53	38.81 G
3	ENet	0.36	1.05	434.83 M
4	ESFNet	0.68	1.69	850.33 M
5	ResUNet	13.04	109.07	52.12 G
6	LMD-Net	10.42	81.68	39.23 G

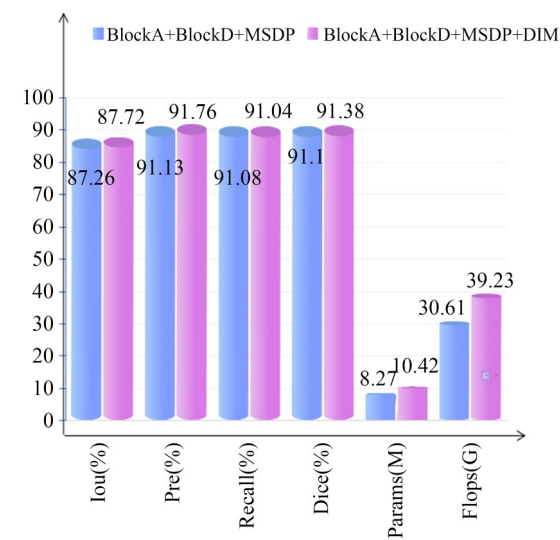


图 10 在 WHU 数据集上 SPFR 机制的对比实验
Fig. 10 Contribution of SPFR mechanism on the WHU dataset

LMD-Net 在各个效率指标中都优于 ResUNet, 不仅参数量降低 2.62 M, 计算量也降低 12.89 G; 表明 LMD-Net 网络实现在效率方面优化 ResUNet 网络。

3.4.2 各模型性能评估实验

为了定量评估 LMD-Net 网络结构在建筑物目标提取时的有效性, 实验中采用 IoU, Precision, Recall 和 Dice 指标对各模型进行精度评估。由表 4 可知, 本文所提 LMD-Net 网络在基于遥感影像建筑物提取领域中表现更优, 相较于对比网络 ResUNet, LMD-Net 网络在交并比、准确率、

召回率、Dice 系数上分别提高了 3.23%, 2.57%, 1.56%, 2.07%; 而轻量化网络 ESFnet 和 ENet 的性能却很低, 在 Iou 上, LMD-Net 相比 ESFnet 模型提高了 11.27%, 因此并不是参数量越低性能就越好。综上所述, LMD-Net 网络实现了在性能方面优化 ResUNet 网络, 更适用于基于遥感影像的建筑物提取。

表 4 不同网络的分割精度对比实验/%
Tab. 4 Comparison experiments of segmentation accuracy of different networks (%)

Model	Iou	Pre	R	Dice
U-Net	83.61	90.09	87.57	88.78
SegNet	79.45	86.76	85.24	85.96
ENet	81.34	87.91	86.92	87.37
ESFnet	76.45	83.95	83.54	83.72
ResUNet	84.49	89.19	89.48	89.31
LMD-Net	87.72	91.76	91.04	91.38

为了能直观体现 LMD-Net 模型的优越性, 如图 11 和图 12 所示, 分别展示了六种模型在不同尺度建筑物上的可视化分割性能。

通过对比图 11 和图 12 的红框区域可知, LMD-Net 网络能有效降低分割噪声, 其误检率更低。由于 SegNet 网络结构简单, 其分割预测图破碎最严重。U-Net 与 SegNet 相比, 整体效果有所提升, 但对于小型建筑物仍存在背景误分。而针对轻型的 ENet 与 ESFnet 网络, ESFnet 网络

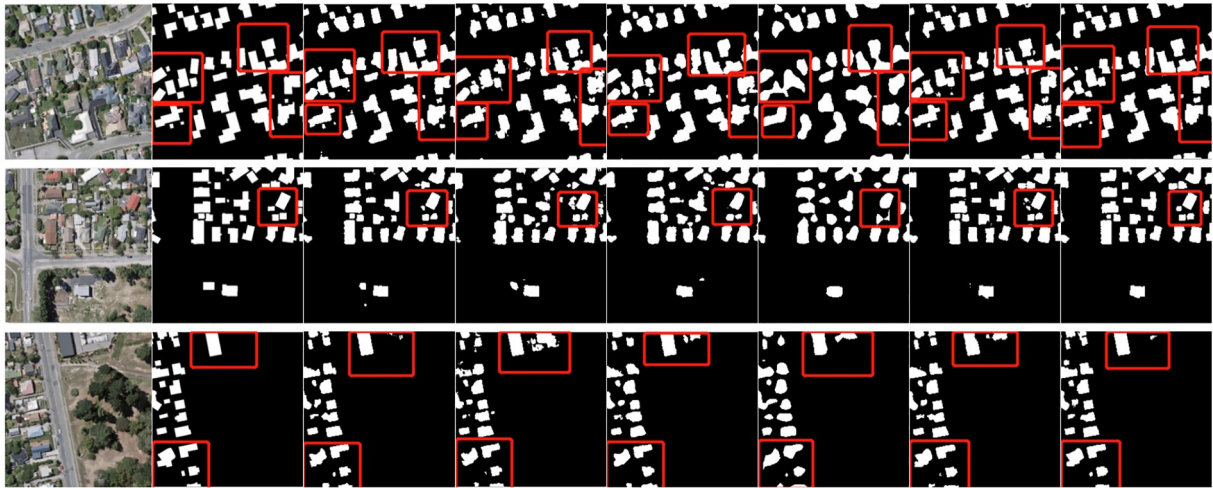


图 11 小型建筑物上预测图
Fig. 11 Forecast map on small buildings

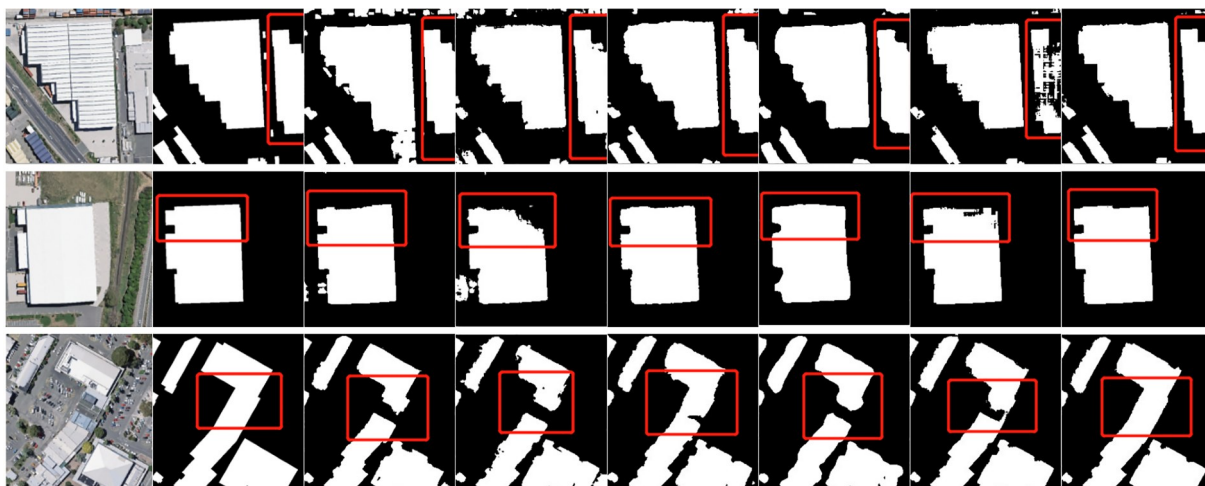


图 12 大型建筑物上预测图

Fig. 12 Prediction map on large buildings

没有完整的建筑物边界,严重缺失边缘细节信息,ENet虽能保持良好边缘信息,但仍会将毗邻的建筑物粘连在一起。而基础模型 ResUNet 虽能有所提升分割性能,但小型建筑物的多检、错检及大型建筑物的漏检现象明显存在,而本文所提的 LMD-Net 网络,能进一步增强建筑物边缘检测能力和降低建筑物分割的错误率,但依然有提升的空间。

4 结 论

针对传统建筑物提取方法存在参数量大、计算量大以及遥感影像建筑物形状多样导致分割存在粘连现象,提出一种轻型多尺度差异网络 LMD-Net 网络模型,其通过编解码结构的功能

差异性引入轻型差异网络,实现模型轻便化的同时增强模型语义表达能力。在此基础上,在编码器第三层和解码器第二层融入多尺度感知模块,增强模型聚合多尺度信息的能力,最后,引入双融合机制进一步增强特征恢复网络的性能。

通过在航空建筑物影像 WHU 数据集上实验验证,所提模型在精度指标交并比值达到 87.26%,准确率达到 91.76%,召回率为 91.04%,Dice 系数是 91.38%,且参数量及计算量均大幅度得到提升,表明,LMD-Net 模型相较其余模型在性能及效率两方面具有更高的可行性。

在后续研究中,在保证模型轻量化的前提下,进一步提升网络对前景像素的正确捕捉能力,从而能提取到更连续的建筑物区域和清晰的建筑物边缘部分。

参考文献:

- [1] WANG Y, ZENG X Q, LIAO X H, *et al.* B-FGC-net: a building extraction network from high resolution remote sensing imagery[J]. *Remote Sensing*, 2022, 14(2): 269.
- [2] HOSSEINPOUR H, SAMADZADEGAN F, JAVAN FD. CMGFNet: a deep cross-modal gated fusion network for building extraction from very high-resolution remote sensing images[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2022, 184: 96-115.
- [3] JIN X Y, DAVIS C H. Automated building extrac-
- tion from high-resolution satellite imagery in urban areas using structural, contextual, and spectral information[J]. *EURASIP Journal on Advances in Signal Processing*, 2005, 2005(14): 1-11.
- [4] 岳照溪,潘琛,郭功举. 基于深度学习的高分辨率光学卫星遥感影像建筑物变化检测方法[J]. *测绘科学与工程*, 2021(2): 30-38.
- YUE Z X, PAN C, GUO G J. A change detection method for high resolution optical satellite remote sensing image based on deep learning[J]. *Geomatics Science and Engineering*, 2021(2): 30-38. (in Chinese)
- [5] SHELHAMER E, LONG J, DARRELL T. Fully

- convolutional networks for semantic segmentation [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(4): 640-651.
- [6] BADRINARAYANAN V, KENDALL A, CIPOLLA R. SegNet: a deep convolutional encoder-decoder architecture for image segmentation [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(12): 2481-2495.
- [7] RONNEBERGER O, FISCHER P, BROX T. *U-Net: Convolutional Networks for Biomedical Image Segmentation* [M]. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2015: 234-241.
- [8] 王舒洋,慕晓冬,杨东方,等. 融合高阶信息的遥感影像建筑物自动提取[J]. 光学精密工程, 2019, 27(11): 2474-2483.
WANG S Y, MU X D, YANG D F, *et al.* High-order statistics integration method for automatic building extraction of remote sensing images [J]. *Opt. Precision Eng.*, 2019, 27(11): 2474-2483. (in Chinese)
- [9] 陈凯强,高鑫,闫梦龙,等. 基于编解码网络的航空影像像素级建筑物提取[J]. 遥感学报, 2020, 24(9): 1134-1142.
CHEN K Q, GAO X, YAN M L, *et al.* Building extraction in pixel level from aerial imagery with a deep encoder-decoder network [J]. *Journal of Remote Sensing*, 2020, 24(9): 1134-1142. (in Chinese)
- [10] MA J J, WU L L, TANG X, *et al.* Building extraction of aerial images by a global and multi-scale encoder-decoder network [J]. *Remote Sensing*, 2020, 12(15): 2350.
- [11] SHI M Y, GAO J. Research on high altitude remote sensing building segmentation based on improved u-net algorithm [J]. *Instrumentation*, 2021, 8(04): 47-54.
- [12] 许正森,管海燕,彭代锋,等. 高分辨率遥感影像建筑物提取的注意力胶囊网络算法[J]. 遥感学报, 2022, 26(8): 1636-1649.
XU Z S, GUAN H Y, PENG D F, *et al.* A dual-attention capsule network for building extraction from high-resolution remote sensing imagery [J]. *Journal of Remote Sensing*, 2022, 26(8): 1636-1649. (in Chinese)
- [13] 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.
- [14] 徐胜军,张若暄,孟月波,等. 融合分形几何特征 Resnet 遥感图像建筑物分割[J]. 光学精密工程, 2022, 30(16): 2006-2020.
XU S J, ZHANG R X, MENG Y B, *et al.* Fusion of fractal geometric features Resnet remote sensing image building segmentation [J]. *Opt. Precision Eng.*, 2022, 30(16): 2006-2020. (in Chinese)
- [15] 罗松强,李浩,陈仁喜. 多尺度特征增强的 ResU-Net+ 遥感影像建筑物提取[J]. 激光与光电子学进展, 2022, 59(8): 0828007.
LUO S Q, LI H, CHEN R X. Building extraction of Remote sensing images using ResUNet+ with enhanced multiscale features [J]. *Laser & Optoelectronics Progress*, 2022, 59(8): 0828007. (in Chinese)
- [16] 徐胜军,欧阳朴衍,郭学源,等. 多尺度特征融合空洞卷积 ResNet 遥感图像建筑物分割[J]. 光学精密工程, 2020, 28(7): 1588-1599.
XU S J, OUYANG P Y, GUO X Y, *et al.* Building segmentation in remote sensing image based on multiscale-feature fusion dilated convolution resnet [J]. *Opt. Precision Eng.*, 2020, 28(7): 1588-1599. (in Chinese)
- [17] LIN J B, JING W P, SONG H B, *et al.* ESF-Net: efficient network for building extraction from high-resolution aerial images [J]. *IEEE Access*, 2019, 7: 54285-54294.
- [18] PASZKE A, CHAURASIA A, KIM S, *et al.* *ENet: a Deep Neural Network Architecture for Real-Time Semantic Segmentation* [EB/OL]. 2016: *arXiv*: 1606.02147. <https://arxiv.org/abs/1606.02147.pdf>
- [19] 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 (CVPR)*. 27-30, 2016, Las Vegas, NV, USA. IEEE, 2016: 770-778.
- [20] HOWARD A G, ZHU M, CHEN B, *et al.* *MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications* [EB/OL]. 2017: *arXiv*: 1704.04861. <https://arxiv.org/abs/1704.04861.pdf>
- [21] RAZA A, HUO H, FANG T. EUNet-CD: efficient UNet for change detection of very high-resolution remote sensing images [J]. *IEEE Geoscience*

- and Remote Sensing Letters*, 2022, 19: 1-5.
- [22] SHAHEED K, MAO A, QURESHI I, *et al.* Finger-vein presentation attack detection using depth-wise separable convolution neural network[J]. *Expert Systems with Applications*, 2022, 198: 116786.
- [23] WANG S H, WANG X Y, GUO X. Advanced face mask detection model using hybrid dilation convolution based method [J]. *Journal of Software Engineering and Applications*, 2023, 16 (1) : 1-19.
- [24] 季顺平, 魏世清. 遥感影像建筑物提取的卷积神经网络与开源数据集方法[J]. *测绘学报*, 2019, 48(4): 448-459.
- JI S P, WEI S Q. Building extraction via convolutional neural networks from an open remote sensing building dataset[J]. *Acta Geodaetica et Cartographica Sinica*, 2019, 48(4): 448-459. (in Chinese)

作者简介:



李国燕(1984—),女,河北石家庄人,博士,硕士生导师,副教授,2006年于河北师范大学获得学士学位,2009年于河北工业大学获得硕士学位,2013年于河北工业大学获得博士学位,主要从事机器视觉、下一代网络技术的研究。E-mail:ligy@tcu.edu.cn

通讯作者:



董春华(1982—),女,山东聊城人,硕士,讲师,2009年于天津城建大学获得硕士学位,主要研究方向为GIS、遥感信息分析。E-mail:dch@tcu.edu.cn