

文章编号:1007-2780(2025)09-1333-14

集成小波变换与全局感知的轻量建筑提取网络

邵文^{1,2}, 邵攀^{1,2*}, 宋宝贵³, 熊彪^{1,2}

(1. 三峡大学 水电工程智能视觉监测湖北省重点实验室, 湖北 宜昌 443002;

2. 三峡大学 计算机与信息学院, 湖北 宜昌 443002;

3. 福州大学 物理与信息工程学院, 福建 福州 350108)

摘要:基于深度学习的建筑物提取是遥感领域一个重要研究方向。为有效平衡计算效率和提取精度,提出一种集成小波变换与全局感知的轻量建筑提取网络。首先,将参数共享、星型运算和深度可分离卷积集成,提出一种星型共享深度卷积块,以高效准确地提取建筑物特征。其次,引入小波变换和频域空间注意力,提出一种高效边界增强模块,增强网络对建筑物边界特征的表征能力。最后,借助轻量级非局部注意力机制与层次特征交互策略,提出一种全局上下文感知模块,显著提升了层级特征的融合效果,增强了网络解码阶段整体感知能力。实验结果表明,所提出的网络在 WHU 和 Inria 两个公开建筑物提取数据集上的联合交并比(IoU)指标分别达到 90.08% 和 79.16%,同时模型参数量(Params)为 3.09M,每秒浮点运算数(FLOPs)为 4.93G、推理速度达到 30.24 FPS。与 Swin Transformer、BuildFormer、SDSCUNet、EasyNet、HDNet 以及 CaSaFormerNet 等现有方法相比,该方法在保持低计算复杂度下,实现了更高的提取精度,在计算效率和提取精度之间实现了更好的平衡。

关键词:建筑物提取;轻量级;边界增强;小波变换;全局上下文

中图分类号:TP751;TP391 **文献标识码:**A **doi:**10.37188/CJLCD.2025-0108 **CSTR:**32172.14.CJLCD.2025-0108

收稿日期:2025-05-22;**修订日期:**2025-07-13.

基金项目:国家自然科学基金(No.41901341);湖北省自然科学基金(No.2024AFB867);湖北省自然科学基金青年基金(No.2024AFB217);自然资源部地理国情监测重点实验室开放基金(No.2025NGCM03)

Supported by National Natural Science Foundation of China (No.41901341); Natural Science Foundation of Hubei Province (No. 024AFB867); Youth Fund of Hubei Provincial Natural Science Foundation (No. 2024AFB217); Open Fund for Key Laboratory of Geographic State Monitoring of Ministry of Natural Resources (No.2025NGCM03)

*通信联系人, E-mail: panshao@whu.edu.cn

Lightweight building extraction network integrating wavelet transform and global awareness

SHAO Wen^{1,2}, SHAO Pan^{1,2*}, SONG Baogui³, XIONG Biao^{1,2}

(1. Hubei Key Laboratory of Intelligent Vision Based Monitoring for Hydroelectric Engineering, China Three Gorges University, Yichang 443002, China;

2. College of Computer and Information Technology, China Three Gorges University, Yichang 443002, China;

3. College of Physics and Information Engineering, Fuzhou University, Fuzhou 350108, China)

Abstract: Building extraction based on deep learning is an important research direction in the field of remote sensing. To effectively balance computational efficiency and extraction accuracy, a lightweight building extraction network integrating wavelet transform and global awareness is proposed. First, by integrating parameter sharing, star-shaped operations, and depthwise separable convolution, a star-shared depthwise convolution block is proposed to efficiently and accurately extract building features. Secondly, wavelet transform and frequency-domain spatial attention are introduced to propose an efficient boundary enhancement module that enhances the network's ability to characterize building boundary features. Finally, employing a lightweight non-local attention mechanism and a hierarchical feature interaction strategy, a global context-aware module is proposed. This module significantly improves the fusion effectiveness of hierarchical features and enhances the overall perception capability during the network's decoding stage. Experimental results demonstrate that the proposed network achieves Intersection over Union (IoU) scores of 90.08% and 79.16% on the publicly available WHU and Inria building extraction datasets, respectively. Concurrently, the model maintains a low parameter count (Params) of 3.09 million, FLOPs of 4.93 billion, and an inference speed of 30.24 frames per second (FPS). Compared to existing methods such as Swin Transformer, BuildFormer, SDSCUNet, EasyNet, HDNet, and CaSaFormerNet, the proposed method achieves higher extraction accuracy while maintaining low computational complexity, achieving a superior balance between computational efficiency and extraction accuracy.

Key words: building extraction; lightweight; boundary enhancement; wavelet transform; global context

1 引 言

随着遥感技术的快速发展,基于高分辨率遥感影像的建筑物提取技术在城市规划、环境监测及灾害评估等领域发挥着重要作用^[1],但是也带来精度与效率的双重挑战。在精度方面,由于建筑物尺度差异大、边界模糊以及阴影和类建筑物背景的干扰,现有方法往往难以兼顾完整性和准确性^[2]。而在效率方面,随着影像分辨率的提升,模型在数据加载、计算和存储等方面的负担显著

增加,使模型在计算资源受限场景下难以适用。在此背景下,如何平衡计算效率和提取精度成为建筑物提取的关键问题^[3]。

近年来,深度学习技术凭借其强大的特征学习能力,逐步成为建筑物提取领域的研究主流^[4]。研究初期,学者通过调整经典的卷积神经网络(Convolutional Neural Network, CNN)架构来提升建筑物提取效果^[5]。吕少云等^[6]结合深度可分离卷积和空洞空间金字塔池化(Atrous Spatial Pyramid Pooling, ASPP)改进 UNet++ 以实现

建筑物提取任务。这些方法主要依赖局部卷积,感受野受限,缺乏捕捉远程依赖关系能力,在面对大尺度或密集分布建筑时容易出现结构中断或误分割现象^[7]。

为增强远距离语义建模能力,遥感研究者开始将注意力机制引入建筑物提取框架^[8]。注意力机制通过动态调整特征权重,使模型能够更关注远距离相关性,从而提高分割精度。根据注意力机制的设计角度不同,可以将注意力机制划分为空间注意力、通道注意力和双重注意力^[9]。Zuo等^[10]采用水平和垂直方向池化生成互补注意力图强化多层次特征图之间融合。针对通道注意力机制,Guo等^[11]采用通道-空间双重注意力机制,在增强建筑物信息表征的同时有效抑制背景干扰。这些方法虽然提升了远距离建模能力,但注意力机制的引入增加了模型计算开销,使其没有兼顾精度和效率^[12]。

Transformer通过多头自注意力机制建模全局依赖关系,能够捕捉远距离像素间的关联性。研究者们为发挥两者的优势,将CNN与Transformer结合在一起。Yuan等^[13]在解码阶段堆叠多个Transformer模块,通过在各层级中学习全局特征,增强网络对复杂结构的建模能力。成嘉伟等^[14]融合Transformer全局建模与CNN局部感知提出双分支网络,通过TC模块解决特征异构问题,并嵌入坐标注意力增强建筑定位。Li等^[15]设计了轻量化CaSaFormer网络,其核心的CaSa模块通过自注意力与交叉注意力协同工作机制,在减少参数量的同时保持多尺度特征提取能力。Huang等^[16]提出的EasyNet通过特征融合块(CTF)实现了CNN局部感受野与Transformer长程依赖建模的优势互补。然而,Transformer结构存在计算复杂度高、参数量大的缺陷,限制了其在资源受限场景下的实时推理能力,并且不能有效解决建筑物边缘模糊问题。

尽管上述方法在建筑物提取领域取得了显著进展,但现有方法未能有效协调计算开销、边界完整度和全局建模能力之间的平衡关系。为解决这些问题,本研究提出一种集成小波变换与全局感知的轻量建筑提取网络(Lightweight Building Extraction Network Integrating Wavelet Transform

and Global Awareness,L-WGNet)。本文的主要创新贡献体现在以下3个方面:

(1)提出星型共享深度卷积块(Star-Shared Depth-wise Convolution Block,SSD-Block)。通过参数共享策略、星型运算和深度可分离卷积结合。该模块在保证特征提取精度同时降低计算复杂度,有效平衡模型效率与性能。

(2)设计高效边界增强模块(Efficient Boundary Enhancement Module,EBEM)。模块结合小波变换与空间注意力机制,利用频域分析和特征自适应加权策略,提升网络对建筑物边界的精细化表达能力。

(3)构建全局上下文感知模块(Global Context-Aware Module,GCAM)。模块基于轻量级自注意力机制与层级特征交互架构,增强长程依赖建模能力,优化多尺度特征融合,提升整体解码质量与语义一致性。

2 相关研究

2.1 轻量级网络

轻量化网络可在维持模型性能的同时显著降低参数量和计算量。其中,MobileNet系列^[17]开创性地提出深度可分离卷积架构,成为轻量化设计的典范。ShuffleNet系列^[18]在此基础上进行创新,引入组卷积与通道混洗机制,在有效降低计算复杂度的同时确保信息交互。与此同时,ESPNet系列^[19]提出基于空间金字塔的分解卷积方法,将大感受野的卷积操作分解为多个并行的低维分支,从而显著减少参数量和计算负担。此外,GhostNet系列^[5]另辟蹊径,基于特征图冗余特性,采用高效的线性变换生成多样化特征。

2.2 注意力机制提升边界特征表达

注意力机制通过动态调整特征的响应,提升网络的表达能力。SE(Squeeze-and-Excitation)模块^[8]利用通道注意力机制,实现特征通道的自适应重标定,增强重要信息的表达,但缺乏空间感知。因此,CBAM(Convolutional Block Attention Module)^[9]在此基础上增加空间注意力,提高特征选择的全面性。然而,基于卷积的注意力机制难以建模长程依赖,Non-local网络^[12]引入全局自注意力以实现全图范围的上下文建模,

但计算复杂度高。而极注意力^[20](Polarized Self-Attention)通过全局池化操作,在保持全局建模能力的同时,将计算复杂度降低,提升效率和适用性。

2.3 建筑物边界优化技术

在遥感影像分割领域,精确的目标边界定位已发展出多种有效方法。其中,罗壮^[21]等提出的边界算子(Sobel)生成边界辅助标签与边界损失函数组合,为解决边界模糊问题提供了重要思路。Zhang等^[22]从特征增强方面入手,设计的多尺度模型配合空间注意力机制增强边界特征的提取能力。Cao等^[7]在后处理阶段采用多策略角点检测与重分类,精细优化边界和角点定位。值得关注的是,Zhou等^[3]将图推理与光流估计的几何校正结合,高效优化边界连续性。

3 方 法

3.1 网络总体结构

图1所示为本文提出的LWGNet整体结构。该网络以U-Net为主干架构,围绕轻量化建模、语义信息建构与边界细节增强3大核心问题,设计SSD-Block、GCAM和EBEM三个功能模块,旨在提升建筑物提取任务中的特征建模能力与边界识别精度,同时显著降低计算开销。为减少输入特征的冗余信息并降低计算复杂度,网络在输入端设置CBR模块,由一个 3×3 卷积(步长为2,输出通道为32)、批归一化和ReLU激活函数构成,作为初始下采样操作。处理后的特征同时输入主干特征提取分支与边界增强分支,实现特征信息的并行建模。

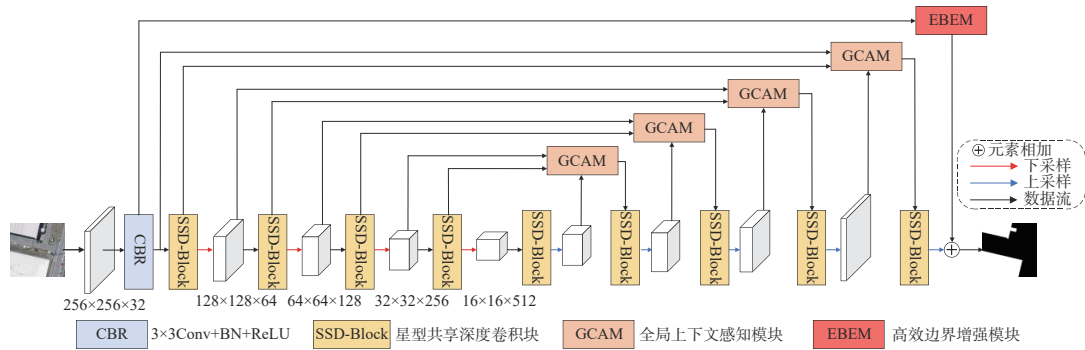


图1 LWGNet总体网络结构

Fig. 1 Overall network structure of LWGNet

主干特征提取分支基于SSD-Block模块构建,其编码器与解码器均由多个SSD-Block模块堆叠构成。该模块融合参数共享机制与深度可分离卷积,有效压缩了模型参数规模并降低了计算成本;引入星型运算,增强了网络的非线性表达能力与提取感知能力。整体设计能够兼顾计算效率与捕获特征信息性能,同时在编码器与解码器之间引入GCAM模块,以增强全局语义建模能力。GCAM核心为轻量级自注意力与层级特征交互机制,通过建立跨尺度的空间-通道关系,提升层级特征之间的信息耦合与语义一致性,在复杂背景中表现出强大的辨别能力。

边界增强路径由EBEM模块构成,用于补偿主干网络在边缘结构建模方面的不足。EBEM通过Haar小波变换将特征图分解为不同频率的

子带。由于高频部分包含图像中的边缘与纹理信息,模块通过频域空间注意力机制对高频分量进行加权建模,增强边缘区域的表达能力。最终将增强后的边界信息与主干解码器融合,以实现更准确的边界还原与细节重建。

总体而言,3个模块在结构和功能上形成互补:SSD-Block构成主干特征提取路径,GCAM强化全局上下文建模与多尺度特征融合,EBEM优化边界细节。这种协同设计使LWGNet在保证效率的同时提升了特征提取能力。

3.2 星型共享深度卷积块

MobileNet给出一种经典的轻量级卷积工具——深度可分离卷积(Depthwise Separable Convolution, DSCConv)。DSCConv将标准卷积分解为深度卷积(Depthwise Convolution, DWConv)

和逐点卷积(Point-wise Convolution, PWConv),显著降低了卷积的参数数量和计算复杂度。然而,DSConv存在以下不足:(1)当特征通道数较大时,逐点卷积仍具有很大的计算开销。(2)深度可分离卷积由于其特性,容易造成通道间交互建模不足。针对上述问题,本文提出如图2所示的SSD-Block(Star-Shared Depthwise Convolution Block, SSD-Block)。首先,SSD-Block引入权重共享策略^[23]到传统的深度卷积和逐点卷积中,并降低逆残差结构中的拓展倍数,从而显著降低计算量。其次,SSD-Block采用高维通道拆分和星型运算操作^[24]增强交互性。在现有增强交互机制中,SE模块通过全局平均池化后,利用两层全连接对通道间信息进行交互建模,随后经Sigmoid生成权重,并与原始特征相乘,实现特征通道间的显式加权交互;CA模块分别对每个通

道在宽度和高度方向进行平均池化,生成形状为 $C \times H \times 1$ 和 $C \times 1 \times W$ 的特征图,并通过卷积实现空间维度上的通道-位置交互,随后经Sigmoid并与原始特征相乘完成特征通道-位置交互。然而,这两种方法采用显式权重映射机制,增加了模型复杂度。此外,池化操作导致特征细节丢失,影响了局部特征感知能力。相较而言,SSD-Block首先在通道维将特征划分为两部分,然后利用星型运算(逐元素相乘)实现两部分特征的交互^[24]。该运算借鉴多项式核函数的思想,通过对子空间特征逐元素相乘完成信息融合交互,并隐式地将特征映射至高维空间^[24]。该方法在无需显式权重学习的情况下,形成通道间的非线性交互。这种方式不仅避免了SE/CA的额外参数计算,还保留了更完整的特征空间细节信息。

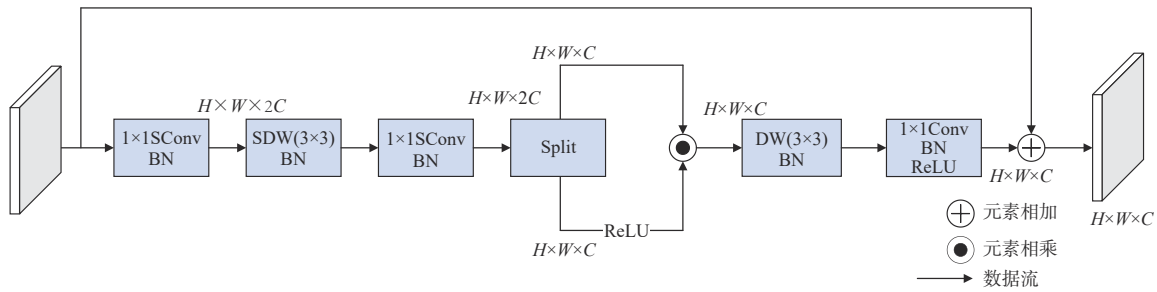


图2 星型共享深度卷积块

Fig. 2 Star-shared depthwise convolution block

具体而言,输入特征首先经由 1×1 分组共享卷积(Shared Convolution, SConv)进行通道扩展,随后通过 3×3 的分组共享深度卷积(Shared Depthwise Convolution, SDW)提取空间局部特征。特征张量再经 1×1 分组共享卷积处理后沿通道维度解耦为基础特征与衍生特征分支,通过元素级乘积操作实现分支特征交互。继而采用 3×3 深度卷积强化局部细节表征,最终通过 1×1 卷积重构通道关系完成多尺度特征聚合,具体过程如式(1)~(4)所示:

$$X_1 = \text{BN}(\text{SConv}_{1 \times 1}(\text{BN}(\text{SDW}_{3 \times 3}(\text{BN} \times (\text{SConv}_{1 \times 1}(X)))))), \quad (1)$$

$$X_2, X_3 = X_1, \text{Split}(), \quad (2)$$

$$X_4 = \text{ReLU}(X_2) \times X_3, \quad (3)$$

$$\text{Output} = \text{Relu}(\text{BN}(\text{Conv}_{1 \times 1}(\text{BN}(\text{DW}_{3 \times 3}(X_4)))) + X, \quad (4)$$

其中: $\text{SDW}_{3 \times 3}$ 表示 3×3 共享深度卷积; $\text{SConv}_{1 \times 1}$ 表示 1×1 共享卷积; $\text{Conv}_{1 \times 1}$ 表示 1×1 卷积; $\text{DW}_{3 \times 3}$ 表示 3×3 深度卷积;BN表示批量归一化;ReLU为激活函数。

深度可分离卷积作为主流轻量结构,由深度卷积与逐点卷积(1×1 卷积)构成。其计算量分别为 $O(9CHW)$ 与 $O(C^2HW)$,对应参数量为 $9C$ 和 C^2 ,其中 C, H, W 分别表示通道数、高度与宽度。为进一步压缩成本,SSD-Block中的共享卷积将通道划分为 G 个分组,并对每组通道共享一组卷积核。该策略下的共享逐点卷积参数量由原始的 C^2 降至 $(C/G)^2$,对应计算量降低为 $O((C/G)^2HW)$,理论压缩比达 $1/G^2$ 。与经典轻量模型MobileNetV2(计算量为 $O(54CHW + 12C^2HW)$,参数量为 $54C + 12C^2$)相比,SSD-Block将计算量压缩至 $O(18CHW + 3C^2HW)$,参数量降至 $18C + 3C^2$,展

现出更优的结构紧凑性与计算效率。

3.3 高效边界增强模块

现有建筑物提取网络在边界优化方面通常采用光流估计或边缘检测算子等^[25]。虽然这些方法能够有效增强边界信息,但其主要考虑影像的空间域,而忽略影像的频率域。在频率域中,可以将空间划分为高频和低频部分,高频部分主要反映图像中强烈变化的区域,如边缘轮廓、纹理以及局部结构细节;低频部分则描述图像的整体结构和背景区域,提供图像的全局形态。因此,频率域

为增强边界信息提供一个有效的工具。基于上述分析,如图3所示,本研究提出高效边界增强模块(Efficient Boundary Enhancement Module, EBEM),借助频率域增强边界信息,特别是高频部分蕴含的边界关键信息,这些高频信息更准确地反映边界细节。EBEM主要贡献包括:(1)引入 Haar 小波变换^[26]进行频域多尺度分解,精确提取频域中的高频边界信息。(2)设计一种轻量化的频域空间注意力机制,用于在频域中强化高频中的边界信息并抑制噪声,实现空间与频域的联合建模。

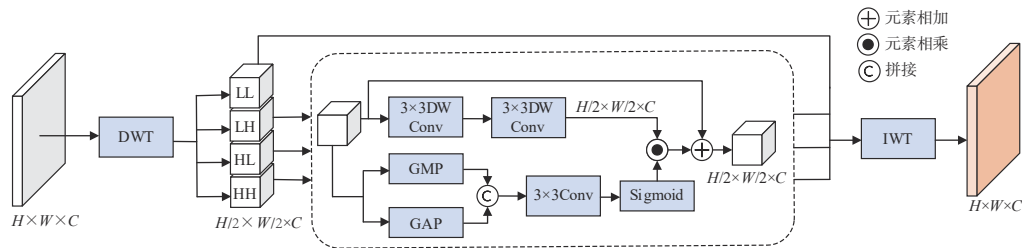


图3 高效边界增强模块

Fig. 3 Efficient boundary enhancement module

首先,对输入图像实施 Haar 小波变换分解生成 4 组子带分量:包含图像近似信息的低频分量(LL)、包含垂直边缘信息的水平方向高频分量(LH)、包含水平边缘信息的垂直方向高频分量(HL)和包含对角边缘信息的对角线方向高频分量(HH)。其次,为降低噪声干扰和强化建筑物边界特征的代表能力,除低频分量(LL)外,分解的 3 组子带分量均经过双支路频域空间注意力机制处理:第一支路由两个级联的 3×3 深度卷积逐步提取局部特征,同时过滤噪声;第二支路采用通道维度的全局最大-全局平均双池化策略,经通道拼接后通过 3×3 卷积与 Sigmoid 激活函数生成空间注意力权重图。接着,两路特征通过矩阵点乘实现特征融合,在增强局部信息的同时有效抑制背景噪声。最后,经注意力增强后的特征与低频分量(LL)通过逆 Haar 小波变换进行融合重构。该操作不仅有效地保留了低频分量的结构完整性,同时将增强后的高频边缘信息注入重建过程,最终输出具有清晰拓扑边界的高质量分割结果,具体操作如下:

$$(LL, LH, HL, HH) = DWT(X), \quad (5)$$

$$C(F) = \sigma(\text{Conv}_{1 \times 1}(\text{Concat}(\text{GMP}(F), \text{GAP}(F))))), \quad (6)$$

$$D(F) = DW_{3 \times 3}(DW_{3 \times 3}(F)), \quad (7)$$

$$A(F) = C(F) \times D(F) + F, \quad (8)$$

$$X_{\text{out}} = IWT(LL, LH_{\text{out}}, HL_{\text{out}}, HH_{\text{out}}), \quad (9)$$

其中: F 表示 $\{LH, HL, HH\}$, DWT表示 Haar 小波变换, IWT表示逆 Haar 小波变换, Concat表示拼接操作, σ 表示 Sigmoid 函数, GMP表示全局最大池化, GAP表示全局平均池化。

针对上述提出模块,本研究进一步对其计算复杂度进行详细分析。该模块的计算复杂度可分解为以下组成部分:首先,基于 Haar 小波变换的频域分解与重构操作具有 $O(2CHW)$ 的计算复杂度^[27],其中 C 、 H 、 W 分别表示特征图的通道数、高度和宽度。其次,双支路频域空间注意力机制包含两个并行计算路径:在深度卷积支路中,通过双级 3×3 卷积核进行特征提取,其计算复杂度为 $O(4.5CHW)$;第二支路通道双池化后采用跨通道特征融合与 3×3 卷积运算的级联,其计算复杂度为 $O(4.5HW)$ 。EBEM 模块整体计算复杂度为 $O(15.5CHW + 13.5HW)$ 。

3.4 全局上下文感知模块

UNet通过跳跃连接融合不同层级的特征,有效缓解编码过程中的空间信息损失^[28]。然而,现有基于 UNet 的建筑物提取方法通常依赖于简单的

特征拼接策略,缺乏对全局上下文的建模能力,从而限制了特征融合的表达效果。为解决上述问题,本文在 Non-local 与极自注意力机制的启发下,提出一种全局上下文感知模块(Global Context-Aware Module, GCAM)。该模块采用高效的空间-通道混合自注意力机制,能够充分建模跨层级全局依赖关系,从而实现更有效的特征融合。

与传统注意力机制相比,GCAM具有以下优势:(1)现有主流注意力(如 Swin Transformer 注意力、极自注意力、Non-local)通常基于单一特征图进行注意力建模,难以有效建模跨尺度的上下文依赖关系。与此不同,本文提出的 GCAM 专门设计用于处理多输入特征,核心创新在于“分级解耦”策略:将通道自注意力和空间自注意力解耦,并分别应用于不同尺度的输入特征情况。这一机制使得 GCAM 能够更充分地挖掘和融合各尺度特征的判别性信息。(2)自注意力机制通常由于卷积映射 K 、 Q 、 V 值以及全局建模的特性,计算开销较大。为降低这一成本,GCAM 通过减少卷积操作,并结合通道压缩与全局池化策略,有效减轻注意力计算负担。GCAM 通道/空间分支计算量分别为 $O(2CHW+0.75C^2)$ 和 $O((C+4)HW)$,

合并后总计算量为 $O((3C+4)HW+0.75C^2)$,总参数量为 $0.75C^2+C+4$ 。相比之下, Swin Transformer 虽在 $M \times M$ 窗口机制下有效限制了注意力交互范围,但仍需完成多次通道映射与局部窗口内的自注意力计算,其计算量约为 $O(4C^2HW+2CHWM^2)$,参数量为 $4C^2$;而 Non-local 模块虽通过将 $Q/K/V$ 映射通道压缩至 $C/2$ 以降低映射开销,但其全局范围的像素交互仍保持着高额的计算量,其计算量为 $O(2C^2HW+2C(HW)^2)$,参数量为 $2C^2$ 。对比可见,GCAM 在保持建模能力的同时,有效减少了计算量和参数量,展现出更高的计算效率与部署适应性。

为直观展示 GCAM 如何实现全局上下文建模,图 4(a) 详细描述了模块的处理流程:模块输入端接收低分辨率特征图 y 及两个高分辨率特征图 x_1, x_2 ,通过分层特征交互实现跨分辨率信息增强。

首先,对低分辨率特征图 y 进行双线性上采样操作,并将其分解为 y_1 和 y_2 两个分支,分别与高分辨率特征图 x_1, x_2 进行 CCSA (Context Channel Self-Attention) 处理,如图 4(b) 所示。以 y_1 与 y_2 的交互过程为例:两者首先相加生成融合特征图,分别经过全局最大池化和平均池化操作获取

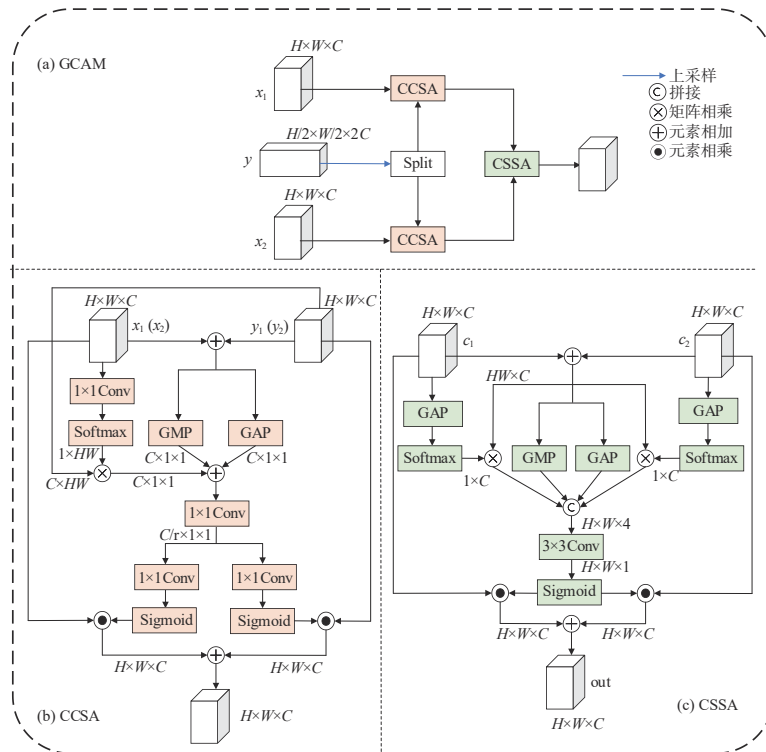


图4 全局上下文感知模块。(a) GCAM 模块;(b) CCSA 模块;(c) CSSA 模块。

Fig. 4 Global context-aware module. (a) GCAM module; (b) CCSA module; (c) CSSA module.

特征通道信息;同时,对 x_1 施加 1×1 卷积生成空间注意力掩膜,经Softmax后与 y_1 进行矩阵乘法计算,获得具有上下文信息的重组特征。该重组特征与池化特征进行加权融合后,通过 1×1 卷积结合Sigmoid激活函数生成通道注意力权重,再分别与 x_1 和 y_1 相乘,得到双路径特征图,两者相加构成通道上下文增强的特征图 c_1 输出。另一组输入对 (x_2,y_2) 以对称结构处理,输出特征图 c_2 。随后再将 c_1 和 c_2 输入到CSSA(Context Space Self-attention)中,如图4(c)所示。

在CSSA中,将 c_1 与 c_2 相加生成融合特征图,融合特征图沿通道维度分别执行全局平均池化与全局最大池化提取空间特征信息。同时,对原始 c_1 与 c_2 分别进行自适应平均池化,并通过Softmax生成自注意力权重。该权重矩阵与融合特征图进行矩阵相乘,构建空间感知的加权上下文特征。通过通道拼接将加权上下文特征与空间特征进行整合,经 3×3 卷积实现通道维度压缩和非线性特征融合并采用Sigmoid生成权重,该权重分别与 c_1 和 c_2 相乘得到双路径特征,并最终相加形成空间增强的特征图 F_{out} 。这种双阶段注意力架构能有效实现跨分辨率特征的全局上下文建模,在保持计算效率的同时提升特征的判别表达能力。

4 实 验

4.1 实验数据集与实验设计

本研究选择WHU数据集^[29]和Inria数据集^[30]。实验数据集具体构成如下:WHU数据集源自新西兰地区的航空影像,原始数据经标准化预处理后统一分割为512像素 $\times$ 512像素,形成包含4 736张训练图像、1 036张验证图像及2 416张测试图

像的标准数据集;Inria数据集涵盖美国与奥地利5个城市的城市建成区,原始影像分辨率达5 000像素 $\times$ 5 000像素。本研究从每个城市中选取5张原始图像构建测试集,数据经填充策略后,以512像素 $\times$ 512像素为单元进行非重叠裁剪,通过滤除无建筑物样本最终形成包含9 294张训练图像与2 325张测试图像的高质量数据集。

实验平台基于PyTorch深度学习框架构建,硬件配置采用NVIDIA GeForce RTX 4060 GPU(8 GB显存)。网络训练采用Adam优化器并实施多阶段学习率衰减策略,初始学习率设定为0.001。损失函数为边界加权二进制交叉熵损失函数(权重0.7)与骰子损失函数(权重0.3)联合优化。训练阶段设置批大小为4,验证阶段设置为2以适配显存限制。针对样本不足问题,构建组合式数据增强策略^[31]:集成随机旋转、水平/垂直翻转以及尺度缩放等空间变换方法有效扩充训练样本多样性。模型在完整训练集上执行200个训练周期以确定最优参数。

为全面评估LWGNet的算法性能,在模型评估方面,采用4个量化指标进行性能分析^[7]:查准率(Precision)表示预测为建筑物的样本中真正的建筑物占比率;查全率(Recall)表示所有建筑物样本中,正确预测出建筑物的样本占比率;F1-score(F1)表示查准率和查全率的调和平均数;联合交并比(IoU)表示预测结果与真实情况的重叠程度。此外,还使用每秒浮点运算数(FLOPs)、模型参数量(Params)和每秒帧数(Frames Per Second,FPS)来衡量不同模型的计算效率和内存消耗。同时,未来代码将在<https://github.com/shaowen1234/lwgnet>中开源,网络实验具体部署如表1所示。

表 1 网络模型部署

Tab. 1 Network model deployment

项目	具体细节
数据构成	WHU:4 736张训练图像,1 036张验证图像,2 416张测试图像;Inria:9 294张训练图像,2 325张测试图像
数据预处理	WHU:标准化处理,分割为512像素 $\times$ 512像素;Inria:非重叠裁剪为512像素 $\times$ 512像素,滤除无建筑物样本
实验平台	PyTorch框架, NVIDIA GeForce RTX 4060 GPU(8 GB显存)
优化器	Adam优化器,初始学习率0.001,多阶段衰减
损失函数	边界加权二进制交叉熵损失(权重0.7)+骰子损失(权重0.3)
训练设置	批大小:4(训练),2(验证),200训练周期
数据增强	随机旋转、水平/垂直翻转、尺度缩放等空间变换
评估指标	Precision、Recall、F1-score、IoU、FLOPs、Params、FPS

4.2 实验结果与分析

本研究选取6个代表性模型构建多维度评估基准,包括Swin Transformer^[32]、BuildFormer^[33]、SDSCUNet^[34]、HDNet^[2]、EasyNet^[16]和CaSaFormer^[15]。其中,Swin Transformer为经典语义分割模型,具有较好的特征提取能力;BuildFormer、

SDSCUNet、HDNet、EasyNet和CaSaFormer为遥感影像建筑物提取领域近年提出的模型。

图5、图6分别展示了各种模型在WHU和Inria数据集上的可视化结果(每个数据集选取4组结果,每一行代表一组),其中图像白色和黑色范围分别表示建筑物区域和非建筑物区域;红色为误

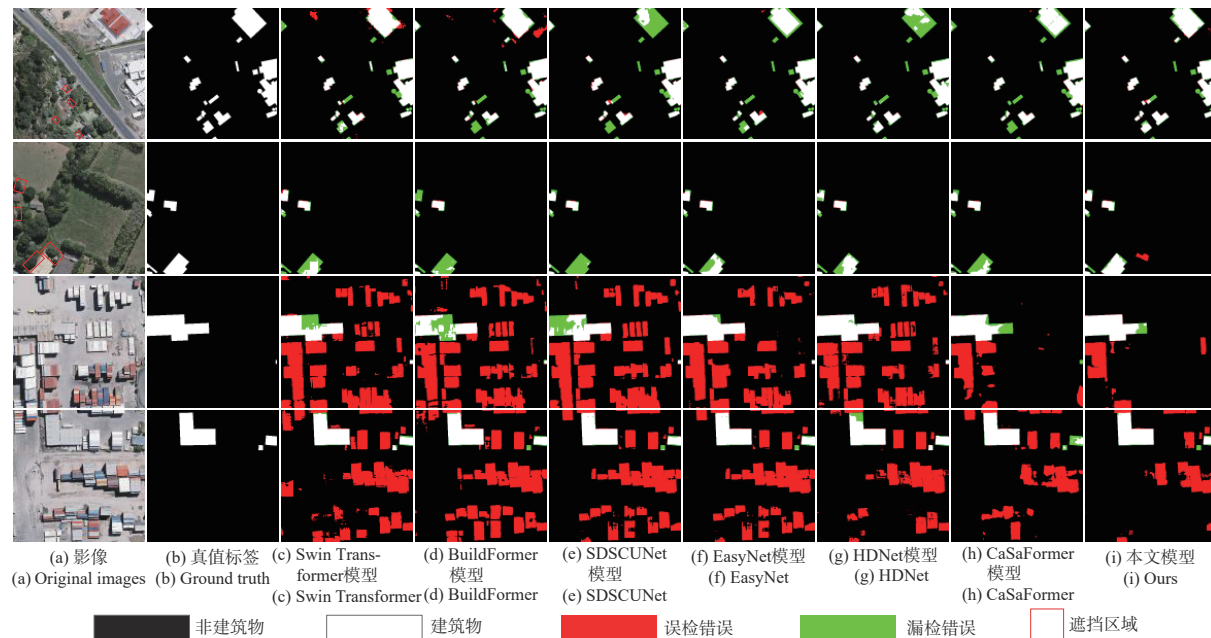


图5 不同模型在WHU数据集上的建筑物提取图

Fig. 5 Building extraction maps for different models on WHU dataset

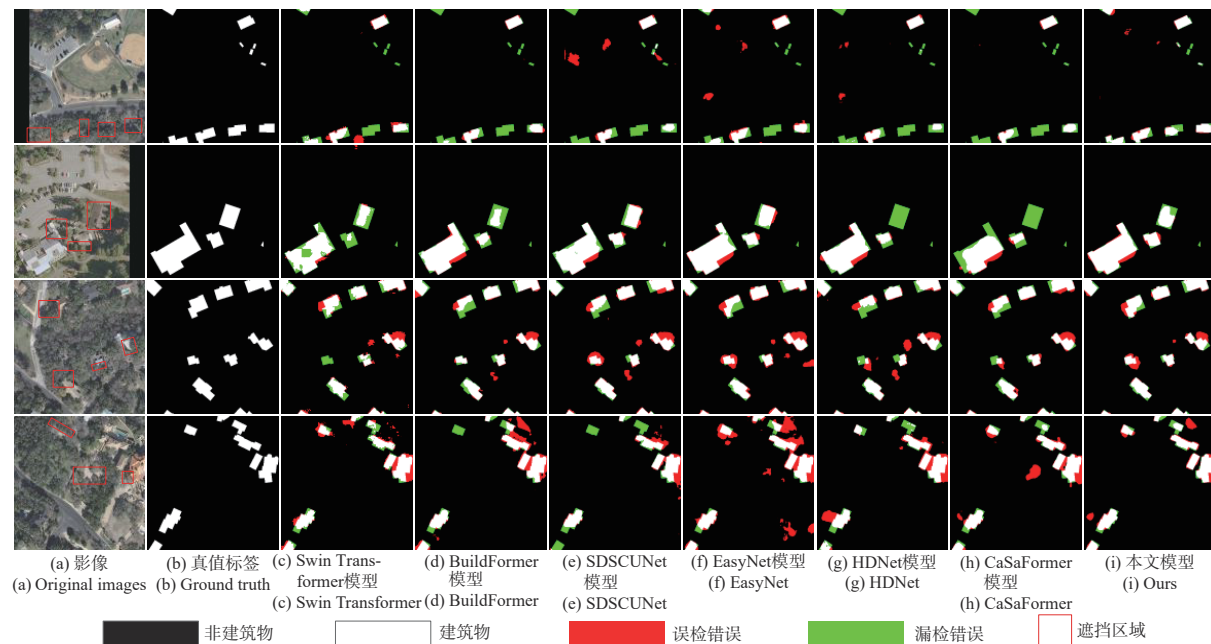


图6 不同模型在Inria数据集上的建筑物提取图

Fig. 6 Building extraction maps for different models on Inria dataset

检错误(即非建筑物区域被误判为建筑物区域),绿色为漏检错误(即建筑物区域被误判为非建筑物区域)。

LWGNet在两个数据集上的建筑物提取任务中展现出显著的性能提升。该模型在应对大规模建筑目标与复杂高噪声背景场景时取得突破性进展,具体表现为以下两个维度:(1)漏检控制层面。通过图5的第1、2组实验的对比分析发现,模型BuildFormer和SDSCUNet在提取小型建筑时存在缺失问题。LWGNet引入全局上下文感知模块,有效提升了建筑物结构提取的完整性。在Inria数据集中,通过图6的第1、2组可知,对于被大量遮挡的建筑物,各模型虽然都容易出现遗漏情况,但是LWGNet通过星型共享深度卷积块,相对于其他网络有着更好的识别能力,遗漏的情况最低。(2)误检抑制层面。对比

实验数据(图5第3、4组)可知,Swin Transformer、HDNet等模型受限于特征表征能力,易将集装箱堆场、高密度停放车辆等干扰目标误判为建筑实体。相较之下,LWGNet通过星型共享深度卷积块优化以及全局上下文感知模块特征提取,实现了对非建筑区域的精准抑制,展现出更优的建筑物识别特异性。在面对建筑物部分边界被遮挡的情况(图6第3、4组),大部分模型都出现了边界虚检的情况。相比于其他模型,本研究提出模型LWGNet中高效边界增强模块有效地缓解了该情况,在边界检查中付出最小的误检代价。

为更客观地比较不同建筑物提取模型的表现,表2和表3分别列出各模型在WHU数据集和Inria数据集的定量评价指标,包括精度(Precision)、召回率(Recall)、F1值和交并比(IoU)等。

表 2 不同网络模型实验结果在 WHU 数据集上的精度指标

Tab. 2 Accuracy indicators of experiment results from different network models on WHU dataset

Methods	Precision/%	Recall/%	F1/%	IoU/%	Params/M	FLOPs/G	FPS
Swin Transformer	92.83	93.79	93.31	87.45	27.18	30.88	22.71
BuildFormer	93.80	94.32	94.06	88.79	40.52	116.36	15.28
SDSCUNet	93.92	94.45	94.19	89.01	21.32	29.81	24.97
EasyNet	93.97	94.67	94.32	89.25	3.90	17.30	28.67
HDNet	94.61	93.96	94.28	89.18	13.89	201.37	9.89
CaSaFormerNet	93.86	92.15	93.00	86.91	4.44	1.99	29.93
LWGNet	94.20	95.37	94.78	90.08	3.09	4.93	30.24

表 3 不同网络模型实验结果在 Inria 数据集上的精度指标

Tab. 3 Accuracy indicators of experiment results from different network models on Inria dataset

Methods	Precision/%	Recall/%	F1/%	IoU/%	Params/M	FLOPs/G	FPS
Swin Transformer	83.68	88.69	86.11	75.61	27.18	30.88	22.71
BuildFormer	85.79	88.38	87.07	77.09	40.52	116.36	15.28
SDSCUNet	86.70	88.09	87.39	77.60	21.32	29.81	24.97
EasyNet	87.24	87.54	87.39	77.61	3.90	17.30	28.67
HDNet	87.61	86.64	87.03	77.04	13.89	201.37	9.89
CaSaFormerNet	87.51	85.48	86.48	76.19	4.44	1.99	29.93
LWGNet	87.50	89.26	88.37	79.16	3.09	4.93	30.24

从结果可知,LWGNet在WHU数据集上的IoU和F1等关键指标均表现最优,其中IoU达到90.08%,较次优模型EasyNet提升0.83%。LWGNet在Inria数据集上的建筑分割任务中同样展现出卓越的性能。在4项关键评估指标上,该模型取得了优异表现:精度(Precision)为87.50%、

召回率(Recall)为89.26%、F1值为88.37%、交并比(IoU)为79.16%。特别值得注意的是,其IoU指标较次优模型提升1.55%,这一优势幅度甚至超过在WHU数据集上的表现。

尽管LWGNet在精度指标上表现优异,模型在实际应用中的部署适应性仍是衡量其实用价值

的重要指标。为此,本文整理并概括了当前主流的移动端与嵌入式部署流程,包括两个阶段:首先,将训练完成的模型通过 TorchScript 序列化为 .pt 格式。随后,移动端通过 Android Studio 将模型集成到应用中,并借助 PyTorch Mobile API 完成模型加载与推理;嵌入式设备则采用 libtorch 的 C++ 接口实现模型的加载部署与执行。

尽管当前多数移动端设备具备 4~12 GB 内存及数十至上千 GFLOPs 的计算能力,但受限于多任务运行、功耗和散热等因素,实际单个任务可用资源远低于理论峰值。因此,模型在计算复杂度和参数规模上的轻量化设计成为实现高效端侧推理的关键保障。

在上述分析基础上,LWGNet 展现出显著的计算效率优势:参数量仅为 3.09M,分别相当于 Swin

Transformer (27.18M) 和 BuildFormer (40.52M) 的 11.4% 和 7.6%;FLOPs 为 4.93G,仅为 HDNet 的 2.4%;推理速度达到 30.24 FPS,在所有对比方法中表现最优。

综上,LWGNet 在保持高精度的同时,兼顾结构简洁与计算开销控制,为实际应用中的端侧部署提供了理论支撑。

4.3 消融实验

本研究以基于 MobileNet_V2 的 Unet 架构作为基础模型,如表 4 第 1 行所示,通过逐步引入星型共享深度卷积块(SSD-Block)、全局上下文感知模块(GCAM)以及高效边界增强模块(EBEM),对各模块在精度和计算效率方面的贡献进行定量分析,验证其在建筑物提取任务中的有效性与协同性。

表 4 在 WHU 数据集上的消融实验

Tab. 4 Quantitative evaluation results of ablation experiments on the WHU dataset

Methods	Precision/%	Recall/%	F1/%	IoU/%	Params/M	FLOPs/G	FPS
baseline	93.08	93.89	93.48	87.76	6.60	13.50	24.18
baseline+EM	93.95	94.07	94.01	88.07	6.60	13.53	23.56
baseline+GM	92.44	95.35	93.87	88.45	6.43	12.93	25.60
baseline+SSD	93.11	95.37	94.23	89.09	3.26	5.47	29.84
baseline+EM+GM	92.76	95.50	94.11	88.88	6.43	12.96	25.16
baseline+EM+SSD	94.04	94.66	94.35	89.30	3.26	5.50	29.22
baseline+GM+SSD	93.68	95.27	94.47	89.51	3.09	4.90	31.46
baseline+EM+SSD+CCSA	94.81	94.26	94.53	89.63	3.09	4.79	29.56
baseline+EM+SSD+CSSA	94.01	94.98	94.50	89.57	3.09	4.62	29.77
baseline+GM+EM+SSD($G=2$)	94.20	95.37	94.78	90.08	3.09	4.93	30.24
baseline+GM+EM+SSD($G=1$)	94.49	95.14	94.81	90.13	5.07	6.75	29.27

边界特征增强的实验结果(表 4 第 2 行)显示,引入 EBEM 模块后,IoU 指标从 87.76% 提升至 88.70%,证明了该模块通过小波变换与频域空间注意力的协同作用在边界特征增强方面的有效性。值得注意的是,EBEM 模块仅增加了 2.8k 的参数量,模型仍保持良好的计算效率(同基准模型的 FPS 相比仅降低 0.62,在 LWGNet 架构中 FPS 降低 1.22),实现了精度提升与计算效率的良好平衡。

在上下文建模机制优化方面(表 4 第 3 行),引入 GCAM 模块替代 UNet 中原始的编解码跳跃连接。实验数据显示,引入 GCAM 后,模型参数量从 6.60M 降至 6.43M,FLOPs 由 13.50G 降至 12.93G。在性能方面,模型 F1 和 IoU 分别提升

至 93.87% 和 88.45%,同时 FPS 提升至 25.60,说明 GCAM 不仅提升了特征表达力,而且还保持了高效的推理速度。

在特征提取组件优化方面(表 4 第 4 行),本研究针对基础模型中采用的 SSD-Block 进行了改进。结果表明,改进后的模型参数量从 6.60M 降至 3.26M(降幅 50.6%),计算量由 13.50G 减少至 5.47G(降幅 59.5%)。在性能表现方面,F1 和 IoU 指标分别提升 0.75% 和 1.33%,证明该结构比 MobileNet_V2 拥有更优的提取能力。同时推理速度从 24.18 FPS 提升至 29.84 FPS,成功实现了特征提取能力增强与模型轻量化的双重优化目标。

在两两模块混合的消融实验(表 4 第 5、第 6、第 7 行)中,EBEM 与 GCAM 模块的组合首先带

来精度和上下文建模的提升, IoU 从 87.76% 提升至 88.88%, F1 为 94.11%。GCAM 通过增强长程依赖建模, 进一步优化了上下文信息的表达, 而 EBEM 则在边界特征增强方面起到积极作用, 展示了精度与计算效率的平衡。EBEM 与 SSD-Block 模块的组合使 IoU 提升至 89.30%, F1 为 94.35%, SSD-Block 通过减小计算复杂度和提升推理速度, 有效优化了特征提取能力和计算效率, 边界增强模块的引入进一步提升了精度。最后, SSD-Block 与 GCAM 组合的协同作用最为显著, IoU 达到 89.51%, F1 为 94.47%。SSD-Block 显著减少了计算开销, GCAM 则增强了上下文建模能力, 推理速度提升至 29.84 FPS, 展现了两者在提高精度与计算效率方面的高度协同。

从表 4 第 6、8、9、10 行 4 组实验的 IoU 指标可见, baseline+EM+SSD 的 IoU 为 89.30%, 在其基础上分别引入 CCSA 与 CSSA 后提升至 89.63% 和 89.57%, 提升幅度为 0.33% 和 0.27%; 进一步引入 GM (即同时包含 CCSA 和 CSSA) 后, IoU 达到 90.08%, 较基线提升 0.78。由此可知, CCSA 和 CSSA 分别贡献约 42% 与 35% 的提升, 其余约 23% 来自二者的协同作用。该结果表明, CCSA 在 IoU 提升中占主导, CSSA 次之, 二者联合可实现超过线性叠加的性能增益, 体现出良好的互补性与协同效应。

在最后两组实验中, 本研究对 SSD-Block 模块是否分组共享进行定量分析, 以评估不同分组策略对模型性能的影响。具体而言, $G=1$ (不分组共享) 配置下, 模型的 IoU 为 90.13%, F1 达到 94.81%。然而, 相较于分组共享配置, 该方案的计算复杂度较高, FLOPs 为 6.75G, 推理速度为 29.27 FPS。这表明, 尽管不分组共享策略在精

度上表现较好, 但计算开销较大。另一方面, $G=2$ (分为两组共享数据) 配置通过分组共享策略显著降低计算成本, FLOPs 降至 4.93G, 推理速度提升至 30.24 FPS, 同时精度保持在相对较高水平, IoU 为 90.08%, F1 为 94.78%。这一结果表明, SSD-Block 分组共享策略有效减少了计算复杂度, 同时保持了较高的精度, 证明了该策略在提升计算效率方面的优势。

5 结 论

本文针对遥感影像中建筑物提取任务, 提出一种轻量化网络结构 LWGNet, 旨在实现特征建模精度与计算效率的有效平衡。所设计的 3 项核心模块各具功能: SSD-Block 融合参数共享、星型运算与深度可分离卷积结构, 兼顾非线性建模能力与模型紧凑性; GCAM 基于轻量级自注意力机制构建全局上下文关联, 提升语义一致性与解码性能; EBEM 结合小波变换与空间注意力机制, 有效增强边界区域的表达能力, 改善建筑物轮廓模糊问题。

在 WHU 和 Inria 两个公开遥感数据集上的实验验证了 LWGNet 的优越性能。在 WHU 数据集上, 模型的 IoU 达到 90.08%; 在 Inria 数据集上, IoU 为 79.16%。此外, LWGNet 在计算效率方面也展现出显著优势: 参数量仅为 3.09M; FLOPs 为 4.93G; 推理速度为 30.24 FPS, 在所有对比方法中表现最优。综上, LWGNet 通过结构上的紧凑设计与模块间的功能互补, 在保持高精度提取能力的同时显著降低了计算负担, 具备良好的泛化性与部署价值。

致谢: 本文受到三峡大学先进计算中心资助支持。

参 考 文 献:

- [1] 潘梦洋, 杨航, 范祥晖. 基于 CNN-Transformer 结构的遥感影像变化检测[J]. 液晶与显示, 2024, 39(10): 1361-1379. PAN M Y YANG H, FAN X H. Remote sensing image change detection based on CNN-Transformer structure [J]. *Chinese journal of liquid crystals and displays*, 2024, 39(10): 1361-1379. (in Chinese)
- [2] LI Y, HONG D, LI C, *et al.* HD-Net: High-resolution decoupled network for building footprint extraction via deeply supervised body and boundary decomposition [J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2024, 209: 51-65.
- [3] ZHOU Y, CHEN Z, WANG B, *et al.* BOMSC-Net: Boundary optimization and multi-scale context awareness based building extraction from high-resolution remote sensing imagery [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, 60: 1-17.

- [4] 周槃入,王哲涛,杨思危,等. 化学计量学和深度学习方法在拉曼光谱处理方面的应用研究进展[J]. 分析化学, 2023, 51(8): 1232-1242.
ZHOU C R, WANG Z T, YANG S W, *et al.* Application progress of chemometrics and deep learning methods in raman spectroscopy signal processing [J]. *Chinese Journal of Analytical Chemistry*, 2023, 51(8): 1232-1242. (in Chinese)
- [5] LIU Z, HAO Z, HAN K, *et al.* Ghostnetv3: Exploring the training strategies for compact models [J/OL]. *arXiv*, 2024: 240411202.
- [6] 吕少云,李佳田,阿晓荟,等. Res_ASPP_UNet++: 结合分离卷积与空洞金字塔的遥感影像建筑物提取网络[J]. 遥感学报, 2023, 27(2): 502-519.
LV S Y, LI J T, A X H, *et al.* Res_ASPP_UNet++: Building an extraction network from remote sensing imagery combining depthwise separable convolution with atrous spatial pyramid pooling [J]. *National Remote Sensing Bulletin*, 2023, 27(2): 502-519. (in Chinese)
- [7] CAO Y, ZHANG S, SUI B, *et al.* IBCO-Net: Integrity-boundary-corner optimization in a general multistage network for building fine segmentation from remote sensing images [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2023, 61: 1-19.
- [8] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018. Salt Lake City: IEEE, 2018: 2011-2023.
- [9] WOO S, PARK J, LEE J Y, *et al.* Cbam: Convolutional block attention module [C]//*Proceedings of the European Conference on Computer Vision (ECCV)*, 2018. German: Springer, Cham, 2018:3-19.
- [10] ZUO X, SHAO Z, WANG J, *et al.* A cross-stage features fusion network for building extraction from remote sensing images [J]. *Geo-Spatial Information Science*, 2024: 1-15.
- [11] GUO H, SU X, WU C, *et al.* Decoupling semantic and edge representations for building footprint extraction from remote sensing images [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2023, 61: 1-16.
- [12] WANG X, GIRSHICK R, GUPTA A, *et al.* Non-local neural networks [C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018. Salt Lake City: IEEE, 2018: 7794-7803.
- [13] YUAN W, RAN W, SHI X, *et al.* Multiconstraint transformer-based automatic building extraction from high-resolution remote sensing images [J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2023, 16: 9164-9174.
- [14] 成嘉伟,郭荣佐,吴建成,等. 双分支特征融合的遥感建筑物检测模型[J]. 计算机工程与应用, 2024, 60(22): 145-153.
CHENG J W, GUO R Z, WU J C, *et al.* Dual-branch feature fusion remote sensing building detection model [J]. *Computer Engineering and Applications*, 2024, 60(22): 145-153. (in Chinese)
- [15] LI J, HU Y, HUANG X. CaSaFormer: a cross-and self-attention based lightweight network for large-scale building semantic segmentation [J]. *International Journal of Applied Earth Observation and Geoinformation*, 2024, 130: 103942.
- [16] HUANG H, LIU J, WANG R. Easy-Net: A lightweight building extraction network based on building features [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2023, 62: 1-15.
- [17] QIN D, LEICHNER C, DELAKIS M, *et al.* MobileNetV4: universal models for the mobile ecosystem [C]//*Proceedings of the European Conference on Computer Vision*, 2024. Milan, Italy: Springer, Cham, 2024:78-96.
- [18] MA N, ZHANG X, ZHENG H-T, *et al.* Shufflenet v2: Practical guidelines for efficient cnn architecture design [C]//*Proceedings of the European Conference on Computer Vision (ECCV)*, 2018. German: Springer, Cham, 2018: 122-138.
- [19] WATANABE S, HORI T, KARITA S, *et al.* ESPnet: End-to-end speech processing toolkit [J/OL]. *arXiv*, 2018: 180400015.
- [20] LIU H, LIU F, FAN X, *et al.* Polarized self-attention: Towards high-quality pixel-wise regression [J/OL]. *arXiv*, 2021: 210700782.
- [21] 罗壮,李明,张德朝. 结合边界约束网络和分水岭分割算法的建筑物提取[J]. 遥感学报, 2022, 26(7): 1459-1468.
Luo Z, Li M and Zhang D Z. Building detection based on a boundary-regulated network and watershed segmentation [J]. *National Remote Sensing Bulletin*, 2022, 26(7): 1459-1468. (in Chinese)

- [22] ZHANG H, ZHENG X, ZHENG N, *et al.* A multiscale and multipath network with boundary enhancement for building footprint extraction from remotely sensed imagery [J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2022, 15: 8856-8869.
- [23] WANG X, STELLA X Y. Tied block convolution: Leaner and better CNNs with shared thinner filters [C]//*Proceedings of the AAAI Conference on Artificial Intelligence*, 2021. Palo Alto, California USA: AAAI Press, 2021:10227-10235.
- [24] MA X, DAI X, BAI Y, *et al.* Rewrite the stars [C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024. Seattle, WA, USA:IEEE, 2024:5694-5703.
- [25] 于水, 宦克为, 王磊, 等. 基于卷积神经网络的近红外光谱多组分定量分析模型研究[J]. *分析化学*, 2024, 52(5): 695-705.
- YU S, HUAN K W, WANG L, *et al.* Multicomponent quantitative analysis model of near infrared spectroscopy based on convolution neural network [J]. *Chinese Journal of Analytical Chemistry*, 2024, 52(5): 695-705. (in Chinese)
- [26] 孔德明, 何绍炜, 李心怡, 等. 基于离散小波变换算法和 Inception 卷积模块一维卷积神经网络的石油类污染物红外光谱定性分析[J]. *分析化学*, 2024, 52(9): 1287-1297.
- KONG D M, HE S W, LI X Y, *et al.* Classification method for petroleum pollutants based on inception-one-dimensional convolutional neural network and infrared spectroscopy [J]. *Chinese Journal of Analytical Chemistry*, 2024, 52(9): 1287-1297. (in Chinese)
- [27] FINDER S E, AMOYAL R, TREISTER E, *et al.* Wavelet convolutions for large receptive fields [C]//*Proceedings of the European Conference on Computer Vision*, 2024. Milan, Italy: Springer, Cham, 2024:363-380.
- [28] 刘振邦, 张硕, 包宇, 等. 深度学习在化学信息学中的应用研究进展[J]. *应用化学*, 2023, 40(3): 360-373.
- LIU Z B, ZHANG S, BAO Y, *et al.* Progress of application research on cheminformatics in deep learning [J]. *Chinese Journal of Applied Chemistry*, 2023, 40(3): 360-373. (in Chinese)
- [29] JI S, WEI S, LU M. Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery dataset [J]. *IEEE Transactions on geoscience and remote sensing*, 2018, 57(1): 574-586.
- [30] MAGGIORI E, TARABALKA Y, CHARPIAT G, *et al.* Can semantic labeling methods generalize to any city the inria aerial image labeling benchmark [C]//*Proceedings of the 2017 IEEE International Geoscience and Remote Sensing Symposium*, 2017. Fort Worth, TX, USA: IEEE, 2017:3226-3229.
- [31] 郑运, 杨思雨, 王涛, 等. 基于数据增强策略和卷积神经网络的近红外光谱分析研究[J]. *分析化学*, 2024, 52(9): 1266-1276.
- ZHENG Y, YANG S Y, WANG T, *et al.* Near infrared spectral analysis based on data augmentation strategy and convolutional neural network [J]. *Chinese Journal of Analytical Chemistry*, 2024, 52(9): 1266-1276. (in Chinese)
- [32] FAN C M, LIU T J, LIU K H. SUNet: Swin Transformer UNet for image denoising [C]//*Proceedings of the 2022 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2022. Austin, TX, USA: IEEE, 2022: 2333-2337.
- [33] WANG L, FANG S, MENG X, *et al.* Building extraction with vision Transformer [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, 60: 1-11.
- [34] ZHANG R, ZHANG Q, ZHANG G. SDSC-UNet: Dual skip connection ViT-based U-shaped model for building extraction [J]. *IEEE Geoscience and Remote Sensing Letters*, 2023, 20: 1-5

作者简介:



邵 文,男,硕士研究生,2022年于东华理工大学长江学院获得学士学位,主要从事深度学习、遥感影像建筑物提取的研究。E-mail: shaow@ctgu.edu.cn



邵 攀,男,博士,副教授,2016年于武汉大学获得博士学位,主要从事深度学习、遥感影像处理等方面的研究。E-mail: panshao@whu.edu.cn