

# 光电工程

## Opto-Electronic Engineering

中文核心期刊 中国科技核心期刊  
Scopus CSCD

### 融合动态特征增强的遥感建筑物分割

肖振久, 田昊, 张杰浩, 曲海成

#### 引用本文:

肖振久, 田昊, 张杰浩, 等. 融合动态特征增强的遥感建筑物分割[J]. 光电工程, 2025, 52(3): 240231.

Xiao Z J, Tian H, Zhang J H, et al. Fusion of dynamic features enhances remote sensing building segmentation[J]. *Opto-Electron Eng*, 2025, 52(3): 240231.

<https://doi.org/10.12086/oee.2025.240231>

收稿日期: 2024-09-30; 修改日期: 2025-01-08; 录用日期: 2025-01-08

### 相关论文

#### 融合ResNeSt和多尺度特征融合的遥感影像道路提取

郝明, 白鹤, 徐婷婷

光电工程 2025, 52(1): 240236 doi: [10.12086/oee.2025.240236](https://doi.org/10.12086/oee.2025.240236)

#### 融合视觉中心机制和并行补丁感知的遥感图像检测算法

梁礼明, 陈康泉, 王成斌, 冯耀, 龙鹏威

光电工程 2024, 51(7): 240099 doi: [10.12086/oee.2024.240099](https://doi.org/10.12086/oee.2024.240099)

#### 特征协同与细粒度感知的遥感图像小目标检测

肖振久, 张杰浩, 林渤翰

光电工程 2024, 51(6): 240066 doi: [10.12086/oee.2024.240066](https://doi.org/10.12086/oee.2024.240066)

#### 基于直接散射光谱的多环境要素激光遥感方法

郭圆新, 梁琨, 徐杨睿, 赵延鹏, 王元庆, 王蓓

光电工程 2024, 51(3): 240003 doi: [10.12086/oee.2024.240003](https://doi.org/10.12086/oee.2024.240003)

更多相关论文见光电期刊集群网站 



<http://cn.oejournal.org/oee>



OE\_Journal



Website



DOI: 10.12086/oee.2025.240231

CSTR: 32245.14.oee.2025.240231

## 融合动态特征增强的遥感建筑物分割

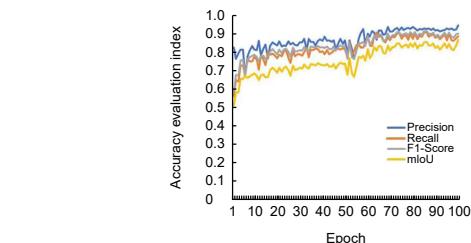
肖振久, 田昊\*, 张杰浩, 曲海成

辽宁工程技术大学软件学院, 辽宁葫芦岛 125105

**摘要:** 针对遥感地物建筑物图像目标尺度差异大、样本空间分布不均衡、地物边界模糊、场景区域跨度大所导致的分割效果不佳问题, 本文提出一种融合动态特征增强高精度遥感建筑物分割算法。首先, 构建 New\_GhostNetV2 网络, 利用自适应上下文感知卷积, 增强算法对样本空间特征的捕捉能力。其次, 采用 Ghost Convolution 结合跳跃连接和特征分支策略设计多层级信息增强模块, 增强特征整合。随后引入级联注意力 CGA (cascaded group attention), 通过组内独立注意力计算, 加强模型对多样化地物形态的适应性。最后, 通过动态深度特征增强器构造特征融合模块, 进一步加强模型捕获能力。在 WHU 数据集上实验结果表明: 改进算法较基线模型 F1-Score 提高 8.57%, mIoU 提高 12.48%, 与其他主流语义分割模型相比, 改进 DeepLabv3+ 具有更好的分割精度。

**关键词:** 遥感图像; 语义分割; 特征增强; 信息整合

中图分类号: TP391.4



文献标志码: A

肖振久, 田昊, 张杰浩, 等. 融合动态特征增强的遥感建筑物分割 [J]. 光电工程, 2025, 52(3): 240231

Xiao Z J, Tian H, Zhang J H, et al. Fusion of dynamic features enhances remote sensing building segmentation[J]. Opto-Electron Eng, 2025, 52(3): 240231

## Fusion of dynamic features enhances remote sensing building segmentation

Xiao Zhenjiu, Tian hao\*, Zhang Jiehao, Qu Haicheng

School of Software, Liaoning University of Engineering and Technology, Huludao, Liaoning 125105, China

**Abstract:** Aiming at the poor segmentation effect caused by the large scale difference of objects, uneven spatial distribution of samples, fuzzy boundary of objects and large span of scene area, this paper proposes a high-precision remote sensing building segmentation algorithm enhanced by integrating dynamic features. Firstly, the New\_GhostNetV2 network is constructed, and the adaptive context-aware convolution is used to improve the algorithm's ability to capture the features of the sample space. Secondly, multi-level information enhancement modules are designed using ghost convolution combined with skip connections and feature branching strategies to enhance the feature integration. Then CGA (cascaded group attention) is introduced to enhance the adaptability of the model to diverse ground object forms through the calculation of independent attention within the group. Finally, the feature fusion module is constructed by the dynamic depth feature enhancer to further enhance the ability of model capture. The experimental results on the WHU data set show that the improved algorithm is 8.57% higher than the baseline model F1-Score and 12.48% higher than mIoU. Compared with other mainstream semantic

收稿日期: 2024-09-30; 修回日期: 2025-01-08; 录用日期: 2025-01-08

基金项目: 辽宁省高等学校基本科研项目 (LJKMZ20220699)

\*通信作者: 田昊, t18846338065@163.com。

版权所有©2025 中国科学院光电技术研究所

segmentation models, the improved DeepLabv3+ has better segmentation accuracy.

**Keywords:** remote sensing image; semantic segmentation; feature enhancement; information integration

## 1 引言

随着计算机视觉的快速发展,深度学习模型在图像分割任务中表现出极大的潜力。如今,作为图像分割任务分支的语义分割被广泛应用于食物检测、路线侦察、田间道路分割、河间监测、建筑物提取等各个方面<sup>[1-5]</sup>。复杂遥感图像的精确分割对城市建筑规划具有更加重要的影响和应用价值<sup>[6]</sup>。语义分割能够将遥感图像中的地物(如建筑物、道路、水体、植被等)逐像素进行分类,从而提供更精细和高分辨率的地物信息<sup>[7-11]</sup>。遥感图像中由于地物形态、大小、纹理等特征的复杂多变,地物间存在相互遮挡、叠加等现象,同时受到大气条件、云层覆盖、光照变化等环境因素的影响,导致图像质量下降,增加了准确分割的难度。现代深度学习技术使得语义分割模型在复杂场景下也能表现出较强的鲁棒性和识别能力,但由于不同地区和应用场景的差异性,仍然存在模型对复杂场景的细节捕捉能力不足、长距离依赖的捕捉能力不足,以及多尺度特征融合困难等问题。因此,语义分割方向的研究有着重要的意义。

近年来,随着深度学习技术的发展,遥感图像语义分割取得了显著进展,众多高效算法被提出。Liu 等<sup>[12]</sup>结合 ResNet 编码器提出了一种多任务学习方法,加强了算法对遥感图像复杂特征的提取能力,解决了传统 UNet 网络由于网络深度较浅,导致了特征提取不充分的问题,且在更深的网络中,存在梯度传播不畅的问题。Wang 等<sup>[13]</sup>提出了改进的双金字塔网络(BiFPN)融合不同尺度的特征信息,结合了空洞空间金字塔池化(ASPP)模块,增强了多尺度信息的处理能力。在利用多层次特征的同时,解决了传统单尺度卷积核对特征提取的限制,虽然 BiFPN 和 Shuffle Attention 模块在计算资源上的增加较小,但仍然提高了内存使用量和 FLOPs。其不足在于处理大规模高分辨率图像时带来了额外的计算负担。Kwak 等<sup>[14]</sup>提出了一种改进的 DeepLabV3-Refiner 模型,通过 LiDAR 测量的 3D 点云进行分析,并利用 DeepLabv3-Refiner 模型增加 3D 点云密度的方法,去除非人类对象的噪声,提供一个更加符合人体形状的高密度分割

结果,通过 RGB 图像和 LiDAR 点云相结合,进一步提升了分割的精度。但是在背景信息不足或缺失时,模型的表现有所下降。Dong 等<sup>[15]</sup>设计了双分支输入(DB input)模块,分别处理浓雾和轻雾区域。该模块增强了模型在不同雾霾条件下的特征学习能力,同时保留了雾霾图像中的初始信息,确保分割结果的准确性,引入了空间域注意力机制(RCCA),用于增强模型在不同分支中的区域注意能力,但是在一些困难场景上的识别分割仍有困难。Harkat 等<sup>[7]</sup>使用膨胀卷积增强多尺度特征提取能力,通过适应不同编码器的方式进一步优化分割边界,提升了边缘的精细分割,但是模型的鲁棒性有待进一步优化。Basir 等<sup>[16]</sup>选用了 ResNet50 作为 DeepLabv3+模型的编码器,能够更好地捕捉图像中的细节,同时利用 ArcGIS Pro 深度学习工具箱进行数据处理和模型训练,实现更高效的地理信息处理,但模型的泛化能力仍然不高,不同气候条件和稻田种植模式会导致模型在未见过的区域上表现不佳。Nie 等<sup>[17]</sup>引入了 ASPP (atrous spatial pyramid pooling) 来增强对多尺度特征的捕捉能力,同时保持较低的参数量和计算复杂度,但使用较大空洞率时,图像分割结果中出现明显的边缘效应,边缘区域的特征信息可能被稀释或模糊。Cabrera 等<sup>[18]</sup>提出了一种基于 DeepLabv3+迁移学习的图像分割方法,对高分辨率图像中的轮廓进行分割,但模型未能充分捕捉细小的边缘和局部结构,导致分割结果不够精准。Ikeda 等<sup>[19]</sup>提出了一种合作学习方法,共享不同卷积神经网络(CNN)之间的特征图来进行语义分割,以增强每个网络的分割能力,但是多个层之间的特征共享反而导致性能下降,影响模型的表现。Quan 等<sup>[20]</sup>在 DeepLabv3+中结合了 U-Net 的浅层特征融合,提升了样本不平衡条件下的提取效果,但是限制了在大规模场景和更多样化的场景中的泛化能力。Tang 等<sup>[21]</sup>通过将网络修改为能够处理单通图象并执行二分类的像素级分类,增加了其模型的适应性,但是其准确性和特异性并不是很高。Harkat 等<sup>[22]</sup>结合了 DeepLabv3+架构和 Xception 主干网络,主要改进在于通过空洞卷积(ASPP)和深度可分离卷积提升火灾检测的语义分割效果,但是训练时间较长,输入大尺

寸给计算机带来了更多的复杂性和过拟合问题。

针对上述挑战, 本文基于 DeepLabv3+网络, 提出一种融合动态特征增强高精度遥感建筑物分割算法。首先, 主干网络使用 New\_GhostNetV2 网络, 通过优化编码器与解码器中的多模块结构, 优化模型对特征信息的捕捉能力和整体分割性能。其次, 设计多尺度信息整合模块 MS-II module (multi-scale information integration module) 用于信息整合, 通过抑制背景干扰信息, 聚焦重要特征区域信息, 增强前景物体的分割能力。然后, 结合 Dynamic Conv 设计动态深度特征增强器 (dynamic depthwise feature enhancer) 提出 DyMSLFusion (dynamic multi-scale feature fusion) 模块, 动态调整特征权重, 整合多层次的上下文信息, 增强模型对细节和全局信息的感知能力, 实现特征的高效融合与重构。最后, 引入级联分组注意力机制 (CGA), 通过在不同层次和分组之间对全局特征进行精细化提取和注意力加权, 加强算法对重要映射信息的关注度, 抑制冗余特征信息, 从而进一步提升分割的准确性。

## 2 融合动态特征增的强遥感建筑物分割

为解决建筑物分割的定位失准及分割效果不佳问题, 提高算法的分割效率, 本文以 DeepLabv3+算法为基线, 提出一种融合动态特征增强的遥感建筑物分割算法。

基线模型 DeepLabv3+由主干网络、空洞空间金字塔池化、解码器和输出层构成。主干网络 Backbone 利用一个核大小为  $3 \times 3$ , 步幅为 2 的标准卷积对  $512 \times 512$  的输入特征图进行初步的特征学习, 减少特征图的空间分辨率, 提高后续层的计算效率, 同时保留重要的空间信息用于进一步的特征学习和语义分割任务。随后, 经最大池化操作, 进一步降低特征图的空间分辨率, 同时保留显著特征, 对特征图进行融合。ASPP 模块使用多个并行的空洞卷积, 通过动态调整空洞率以捕捉更广泛的全局上下文信息, 并沿通道维度利用  $1 \times 1$  卷积进行特征映射的融合。在解码器中: 首先对 ASPP 输出特征图进行上采样操作, 从主干的浅层网络提取特征图, 将其通过一系列的卷积进一步增强特征后再次进行上采样; 随后通过一个  $1 \times 1$  卷积层将解码器的特征图映射到目标类别的维度并且生成最终预测。

本文算法如图 1 所示, 较基线模型 DeepLabv3+做出四项改进: 1) 采用改进 New\_GhostNetV2<sup>[23]</sup> 作为主干网络, 利用 Ghost module 及多种卷积减少冗余计算。2) 设计 MS-II 模块, 通过分组卷积与多分支结构, 降低模型计算量并增强多尺度特征间的融合效率。3) 引入 CGA<sup>[24]</sup> 级联分组注意力机制。在编码器中, 级联分组注意力机制使模型更好的捕捉特征图中的全局上下文信息, 提升模型对于细节特征的捕捉能力; 在解码器中, 级联分组注意力机制增强了上采样

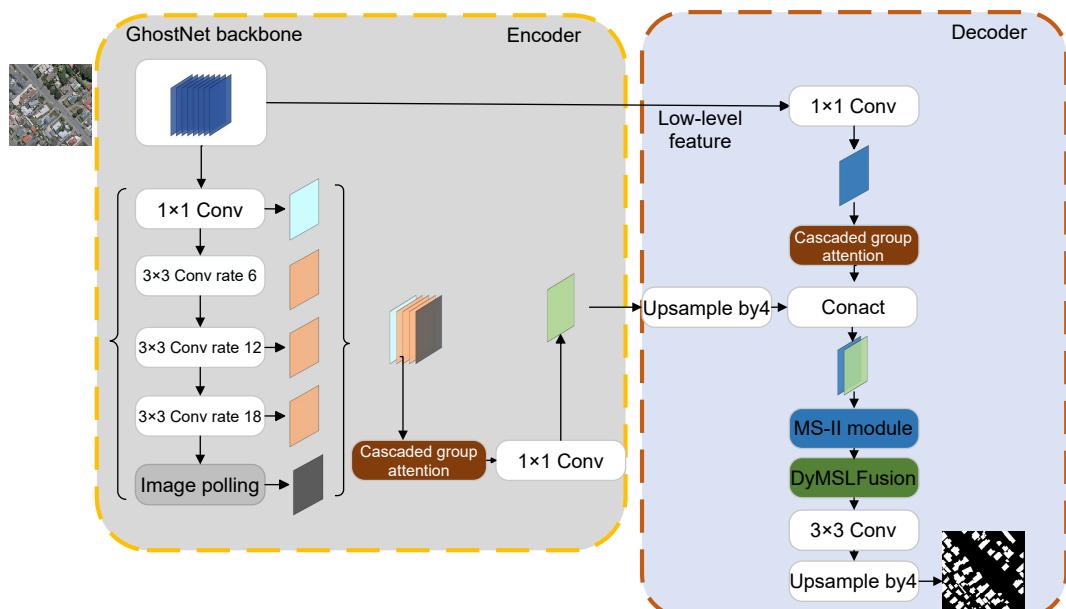


图 1 模型总体结构  
Fig. 1 Overall model structure

特征的恢复能力。4) 采用动态卷积技术, 提升对不同输入数据的适应性, 使得特征融合更加高效灵活, 此外搭建动态深度特征增强器, 通过动态特征提取和增强机制, 增强算法对不同尺度和复杂特征的捕捉能力。

## 2.1 New\_GhostNetV2

基线模块 GhostNetV2 通过 Ghost 模块来减少神经网络中的冗余计算, 通过生成“Ghost”特征来减少计算资源开销, 降低模型的复杂度, 同时利用深度可分离卷积与注意力机制优化算法分割性能, 解决模型计算量大和推理速度慢的问题。但其在捕捉全局特征和细节信息时效果不佳, 虽然降低了计算复杂度, 但在处理复杂场景时表达能力仍然有限, 对此本文对其进行了优化, 提出 New\_GhostNetV2, 结构如图 2 所示。

New\_GhostNetV2 中 Ghost module 模块输出的特征会通过 SC\_FC 模块, 对通道进行自适应的权重调整, 增强重要特征的表达, 从而增强网络的表示能力, 在保证计算效率的同时, 增强网络对于全局和局部特征的整合能力, 使得特征表达更加全面。随后经过 Deformable 卷积后输出, 使其能够更好地调整卷积操作的感受野, 增强对复杂和不规则地形的特征提取能力。在输入特征图的同时经过 Pointwise 卷积进行逐

点卷积并输出, 从而有效聚合通道信息, 减少不必要的信息冗余。将两分支进行相加后进行最终输出。

$$a'_{hw} = \sum_{h'=1}^{H'} F_{h,h'w}^{H'} \odot Z_{h'w}, h = 1, 2, \dots, H', w = h = 1, 2, \dots, W', \quad (1)$$

$$a_{hw} = \sum_{w'=1}^{W'} F_{w,hw'}^{W'} \odot a'_{hw'}, h = 1, 2, \dots, H', w = h = 1, 2, \dots, W', \quad (2)$$

式中:  $F^H$  和  $F^W$  是变换权重。将原始特征  $Z$  作为输入, 式(1)和式(2)顺序地应用于特征, 分别捕获沿着两个方向的长程相关性。

## 2.2 信息整合模块

针对分割建筑物时可能无法有效结合全局与局部特征信息, 进而导致融合不充分, 以及对细节部分的捕捉能力减弱等问题, 本文引入信息整合模块 (MS-II module), 结构如图 3 所示。首先, 输入特征图经过一个卷积操作生成初步的特征图; 继而通过 Split 操作对输入特征图进行切片操作; 每个子特征图进入不同卷积路径, 分别提取不同尺度信息以及不同感受野下的特征信息; 每条路径处理后的特征图在加法节点(④)处进行逐点相加, 融合各个分支特征; 随后通过跳跃连接以及 Concat 进行通道维度整合, 将所有提

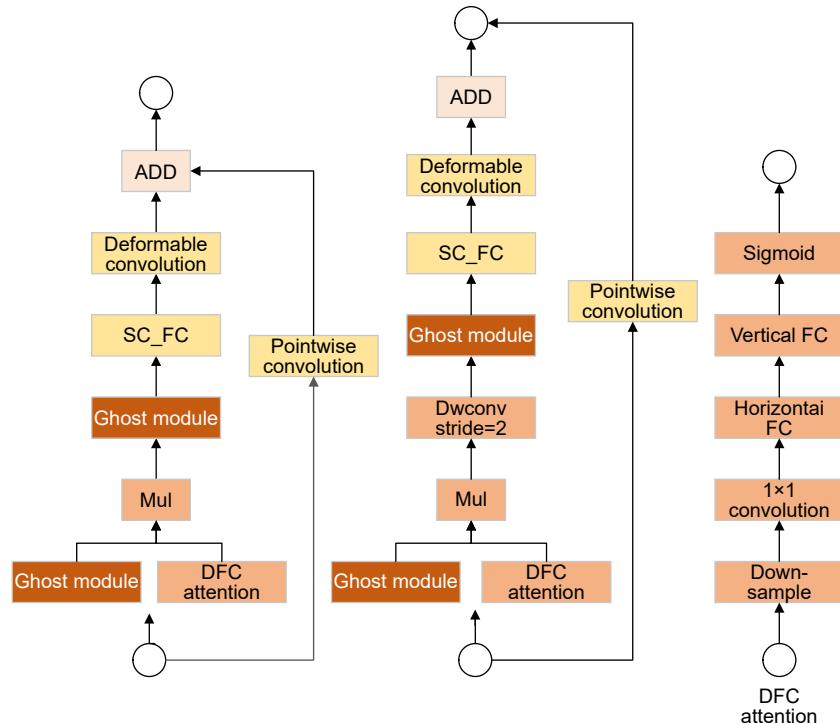


图 2 主干网络 (New\_GhostNetV2) 结构  
Fig. 2 Structure of backbone network (New\_GhostNetV2)

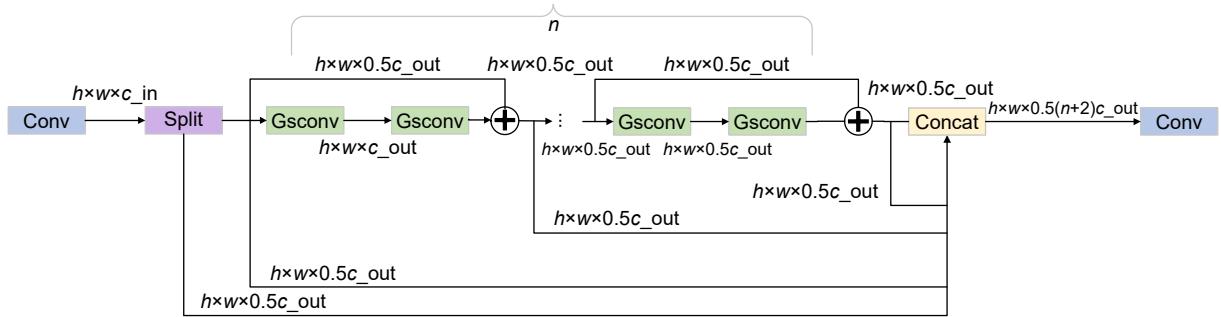


图 3 信息整合 (MS-II) 模块  
Fig. 3 Information integration (MS-II) module

取的多尺度信息整合在一起，在解决梯度消失问题的同时保留浅层信息<sup>[25]</sup>。通过多层 GSConv 提供轻量化的卷积运算，有效减少参数量和计算量，同时保持对特征的敏感性。这种跨层次的特征整合可以在捕捉细节特征的同时保持全局上下文信息，提升网络对建筑物边界和细节的分割效果。

### 2.3 特征融合模块

对于信息整合模块后的特征图，本文对其进行交

互和建模操作，以捕捉各个阶段特征映射之间的依赖关系。借鉴 MSLF (multi-stage local fusion)<sup>[26]</sup>思想并对其进行改进，整体结构如图 4 所示。首先，将 MS-II module 输出分成两个支路输入到 MSLF 中，前两个输出  $F^r$  和  $F^v$  作为响应特征，最后一个输出  $F^v$  作为验证特征，其中  $v$  是输入特征的一个表示， $r$  是用于表征不同阶段输出特征的变量， $r1$ 、 $r2$  是两个不同阶段输出的特征图。为了捕获特征映射的远程依赖

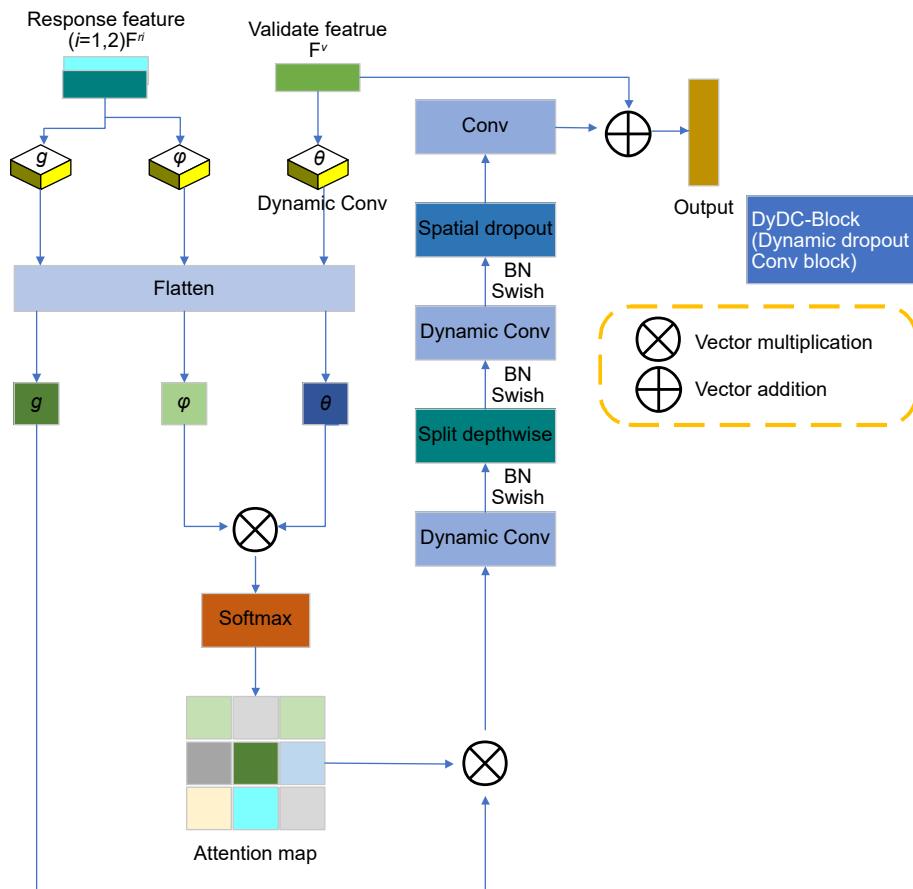


图 4 特征融合模块 (DyMSLFusion)  
Fig. 4 Feature fusion module (DyMSLFusion)

关系, 需要计算注意力模块三个输出之间的空间位置关系, 其中  $\mathbf{W}_\theta$  是需要学习的权重矩阵。

$$\theta = \mathbf{W}_\theta F^v. \quad (3)$$

鉴于三个注意力模块输出在空间维度上的差异, 有必要分别学习独立的权重矩阵:

$$\varphi = \mathbf{W}_\varphi F^{v(r)}, \quad (4)$$

$$g = \mathbf{W}_g F^r. \quad (5)$$

将三个特征图投影到嵌入空间后, 原始特征图被缩小到原始通道大小的一半, 特征图的高度和宽度被平展成  $H \times W$ 。然后对平面化的验证特征图和响应特征图进行乘积运算。计算该矩阵的相似度 ( $f$ ), 记为

$$f(F^v, F^r) = \theta(F^v)\varphi(F^r). \quad (6)$$

为了进一步提升特征融合的效果, 通过 softmax 得到一个注意映射, 减少不重要的特征权重, 确保模型关注重要特征, 从而调整特征在模型中的权重。

为提升模型对目标的检测性能, 本文设计动态深度特征增强器, 通过 Split depthwise convolution 和 Dynamic convolution<sup>[27]</sup> 的结合使用, 采用动态调整卷积核提升特征提取的灵活性和适应性, 增强模型的泛化能力。继而加入 Spatial dropout 模块随机丢弃部分特征图, 有助于减少模型在训练过程中对某些特定特征的过度依赖, 防止过拟合。同时在每一层卷积操作之后进行批量归一化与 Swish 激活函数的组合, 减少内层协方差偏移现象, 提高模型的收敛速度。Swish 激活函数可以保留负值的输入信息, 该模块显著增强了模型对于特征融合的处理能力, 使得模型在处理复杂数据时有更好的稳定。之后, 对原始验证特征进行  $1 \times 1$  卷积, 得到 MSLF 的最终输出。最后的卷积保证了前一步得到的矩阵的大小保持不变, 最终的输出如式 (7) 所示, 其中  $z$  可通过卷积操作得到。

$$MSLF_{out}(F^v, F^{r1}, F^{r2}) = F^v + z(f(\theta(F^v), \varphi_1(F^{r1})g_1(F^{r2}))). \quad (7)$$

## 2.4 注意力机制

模型无法在分组间有效地选择和增强关键特征, 同时模型更依赖于局部特征信息, 导致无法有效捕捉远程依赖关系。这种情况会导致 Encoder 的输出在传递过程中丢失部分有价值的信息, 普通的多头自注意力机制灵活性有限, 缺乏复杂交互能力, 本文引入一种高效注意力机制, 称为级联组注意力 CGA, 它为提供完整特征的不同分割, 显式地分解了头部的注意

力计算。正式地说, 这种注意力可以表述为

$$\tilde{X}_{ij} = Attn(X_{ij}W_{ij}^Q, X_{ij}W_{ij}^K, X_{ij}W_{ij}^V), \quad (8)$$

$$\tilde{X}_{i+1} = Concat[\tilde{X}_{ij}]_{j=1:h_z} W_i^P, \quad (9)$$

式中: 第  $j$  个头计算  $X_{ij}$  上的自注意力, 是输入特征  $X_i$  的第  $j$  次分割, 即  $X_i = X_{i1}, X_{i2}, \dots, X_{ih_z}$ , 且  $1 \leq j \leq h_z$ ,  $h_z$  是正面的总数;  $W_{ij}^Q$ 、 $W_{ij}^K$  和  $W_{ij}^V$  是映射将输入特征分割成不同子空间的投影层;  $Attn(\cdot)$  是通过注意力机制计算的输出;  $Concat(\cdot)$  是将向量拼接在一起;  $W_i^P$  是一个线性层, 它将连接输出特征投影回与输入一致的维度。虽然特征分割不如每个图像的完整特征更有效, 但通过鼓励  $Q$ 、 $K$ 、 $V$  层去学习具有更丰富信息的特征上的投影来继续提高其容量。以级联的方式计算每个头部的注意力机制图, 如图 5 所示, 它将每个头部的输出添加到后续头部中, 以逐步改进特征表示。

$$X'_{ij} = X_{ij} + \tilde{X}_{i(j-1)}, 1 < j \leq h_z, \quad (10)$$

式中:  $X'_{ij}$  为第  $j$  个输入分割  $X_{ij}$ , 与式 (8) 计算出的  $(j-1)$  一头部输出的  $\tilde{X}_{i(j-1)}$  相加。在计算自注意力时, 它替换  $X_{ij}$ , 作为第  $j$  个头的新输入特征。除此之外, 在  $Q$  投影之后应用交互层, 使自注意力能够共同捕获局部和全局关系, 进一步增强特征表示。将此注意力添加到此处可以提升不同模块之间特征的流动性和融合效果。将分组的注意力信息级联传递, 能够有效避免在多层卷积之后特征信息的损失, 以及提升整体的特征传递效率。加入 CGA 机制后, 可以有效避免无法有效捕捉远程依赖以及特征融合不充分的缺点, 使得模型能够更加智能地选择和聚合不同尺度和不同层次的特征, 从而在提升模型精度、减少冗余信息和增强长距离依赖关系建模等方面发挥重要作用。

## 3 实验结果与分析

### 3.1 数据集及实验环境

WHU Building Dataset<sup>[28]</sup> 是一个高分辨率的遥感影像数据集, 旨在通过图像处理和深度学习方法提取建筑物的几何形状和空间位置。被设计用于训练和评估建筑物提取模型, 尤其是针对从遥感图像中自动识别和分割建筑物的任务。首先本文将数据集裁剪为 512 pixel  $\times$  512 pixel 大小的学习样本, 按照对应标签和图片整理成 VOC 格式数据集, 图 6 为实验数据集样本与标签图的示例。

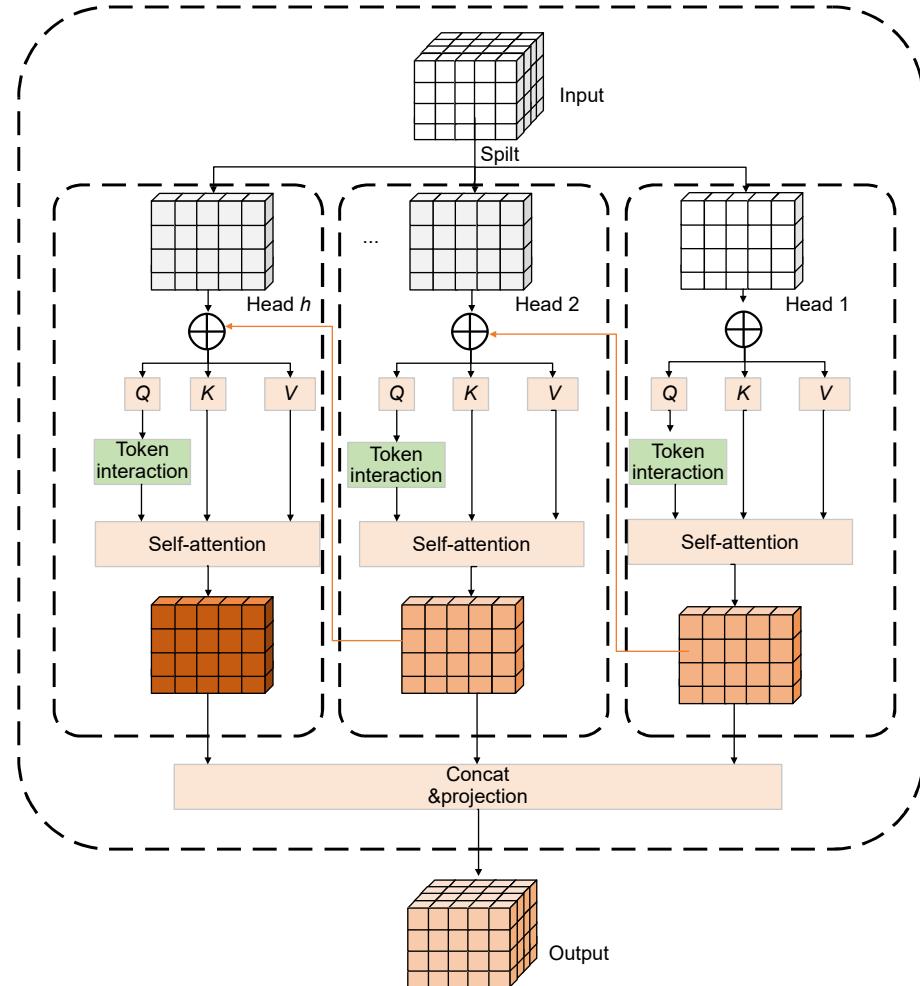


图 5 注意力机制 (CGA)

Fig. 5 Attention mechanism (CGA)

本实验环境为 Windows 11, 处理器 Intel i9-10900x, GPU 是 NVIDIA GeForce RTX 3090, 编程语言为 Python 3.8, 深度学习框架为 pytorch 2.0, 设置 100 Epoch。

### 3.2 评价指标

本实验评价指标选择精度 *Precision*、召回率 *Recall*、*mIoU* 以及模型精度和召回率的一种加权平均值  $F_1$  作为评价指标, 指标定义如下:

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

$$Recall = \frac{TP}{TP+FN}, \quad (12)$$

$$mIoU = \frac{1}{N} \sum_{i=1}^N \frac{TP_i}{TP_i + FP_i + FN_i}, \quad (13)$$

$$F_1 = \frac{2TP}{2TP+FP+FN}, \quad (14)$$

式中: *TP* 表示预测为正的正像素的数量; *FP* 表示被分类为正的负像素数; *FN* 表示被分类为负的正像素数。

### 3.3 消融实验

在 DeepLabv3+模型的基础上改进, 在 WHU 数据集中进行消融实验, 实验数据环境以及其他配置要求全部要求严格一致。实验数据结果如表 1 所示, 其中 A 表示使用 New\_GhostnetV2 为主干网络, B 表示加入 CGA, C 表示加入 MS-II module, D 表示加入 DyMSLFusion, 同时表格中“√”表示使用该改进策略, “×”表示未使用此策略。

实验结果表明, 未改进过的 DeepLabv3+模型在此数据集下的实验结果为: Precision 为 82.32%, Recall 为 75.58%, F1-Score 为 81.64%, mIoU 为 74.35%。将原 DeepLabv3+模型中的 backbone 替换为 New\_GhostnetV2 后, 在相同数据集下的实验结果为:

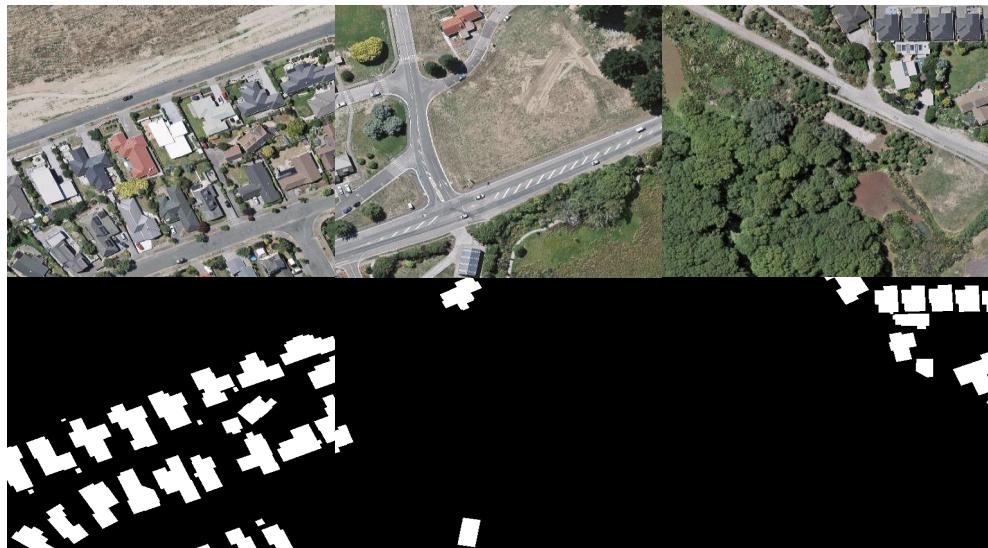


图 6 数据集样本与标签图示例

Fig. 6 Dataset sample and label chart example

表 1 消融实验结果

Table 1 Ablation experiments results

Number	A	B	C	D	Precision/%	Recall/%	$F_1$ /%	mIoU/%
1	✗	✗	✗	✗	82.32	75.58	81.64	74.35
2	✓	✗	✗	✗	84.14	77.48	83.68	77.51
3	✗	✓	✗	✗	84.45	76.96	82.94	76.23
4	✗	✗	✓	✗	85.56	78.95	83.45	78.51
5	✗	✗	✗	✓	86.54	79.45	83.96	80.48
6	✗	✗	✓	✓	87.19	82.49	84.59	81.74
7	✓	✓	✗	✗	88.94	84.53	85.61	82.48
8	✓	✗	✓	✓	91.85	83.89	86.95	81.87
9	✗	✓	✓	✓	91.25	86.57	88.91	82.36
10	✓	✓	✓	✓	<b>94.45</b>	<b>88.86</b>	<b>90.21</b>	<b>86.83</b>

Precision 为 84.14%， Recall 为 77.48%， F1-Score 为 83.68%， mIoU 为 77.51%。相较于原模型：Precision 提升了 1.82%， Recall 提升了 1.90%， F1-Score 提升了 2.04%， mIoU 提升了 3.16%。可以看出，在不影响性能的前提下，通过 new\_GhostnetV2 骨干产生了更多的特征图，且加入的 SC\_FC 进一步增强了模型的特征表达能力，细节部分处理更为精确，模型性能提升。因此，又单一引入 CGA，通过增加注意力机制使模型更精准，可以看到数据结果为：Precision 为 84.45%， Recall 为 79.96%， F1-Score 为 82.94%， mIoU 为 76.23%。可以看到，相较于原模型仍然有提升效果，分别单独引入 MS-SGF module，通过综合不同层次的特征信息来进行实验，以及单独引入 DyMSLFusion 来进行特征融合实验，实验结果相

于原 deeplabv3+模型依然有较高的提升。

将骨干网络换成 New\_GhostnetV2 同时加入 CGA，可以看到相较于原模型：Precision 提升了 6.62%， Recall 提升了 8.95%， F1-Score 提升了 3.97%， mIoU 提升了 8.13%，说明算法很好地减少了冗余信息。将 MS-SGF module 与 DyMSLFusion模块同时加入导模型中，可以看到相较于原模型：Precision 提升了 12.13%， Recall 提升了 13.28%， F1-Score 提升了 8.57%， mIoU 提升了 12.48%，说明 MS-SGF module 可以很好地保留浅层信息，保持对特征的敏感性。这种跨层次的特征整合可以捕捉细节特征。DyMSLFusion 模块在特征融合时可以很好地保留每个通道的独特信息，同时避免信息的冗余，使得最终的输出更加丰富和精确。由实验结果可知，该改进模型具有较好

的性能。

### 3.4 对比试验

为了更好的评价改进 DeepLabv3+模型的效果, 将提出的改进算法与近年流行的算法在 WHU 数据集上进行对比, 对比的网络包括 UNet<sup>[29]</sup>、SegNet<sup>[30]</sup>、DeepLabv3+<sup>[31]</sup> 和 FCN<sup>[32]</sup>, 对比实验结果如表 2 所示。

为了体现实验结果的公平性, 在采用相同数据集和环境配置情况下, 得出改进模型评价指标精度图, 如图 7 所示。Precision (蓝线): 表示模型预测为正样本的准确率。可以看出, Precision 从第一个 Epoch 开始逐渐上升, 在训练初期略有波动, 但总体趋势是逐步提高。在后期 (50 个 Epoch 以后), Precision 趋于平稳, 并维持在接近 0.94 的水平。Recall (橙线): 表示模型在实际正样本中正确预测的比例。Recall 在训练初期迅速上升, 并在中后期 (约 30 个 Epoch 之后) 波动较小, 最终稳定在 0.88 左右的水平。F1-Score (灰线): F1 分数是 Precision 和 Recall 的调和平均值, 用于综合评估模型的表现。曲线显示 F1-Score 在训练初期快速提升, 随后进入相对平稳的上升阶段, 在训

练后期接近 0.9。mIoU (黄线): mIoU (Intersection over Union) 是语义分割模型中常用的指标, 它衡量的是模型预测的结果与真实标注之间的重叠程度, 用于评估分割模型的性能。mIoU 曲线在初期有较大的波动, 但在训练的中后期逐渐趋于平稳, 最终接近 0.86 的水平。改进后的模型效果在 Precision、Recall、F1-Score 和 mIoU 指标上的效果均优于其他各类算法。

本文采用的方法以及这些经典语义分割网络在相同的验证集进行的分割效果如图 8 所示, 对不同的算法进行试验数据比较, 横行表示 FCN、Unet、SegNet、DeepLabv3+和所提算法的模型分割能力。可以看出, FCN 可以实现端到端的像素级分割, 但是对于细节恢复的能力比较弱, 对于样本空间分布不均衡的场景, 很难实现良好的分割效果。UNet 网络可以很好的保留低级特征, 但是对于大规模场景数据特征表达能力不足, 很难处理高分辨率图像。SegNet 虽然减少了计算复杂度, 但其无法处理过于复杂的网络。DeepLabv3+可以捕获多尺度上下文信息, 但是计算量大, 计算成本高。所提算法对小目标样本与复杂背景下的多尺度

表 2 各类算法精度对比结果  
Table 2 Precision comparison results of various algorithms

Arithmetic	Precision/%	Recall/%	F <sub>1</sub> /%	mIoU/%	Time/(min/Epoch)
UNet	87.85	86.87	86.95	82.64	4.0
FCN	86.46	85.76	84.63	79.62	4.5
SegNet	89.39	88.68	87.25	83.46	3.5
DeepLabv3+	82.32	75.58	81.64	74.35	5.0
Our Methods	<b>94.45</b>	<b>88.86</b>	<b>90.21</b>	<b>86.83</b>	2.5

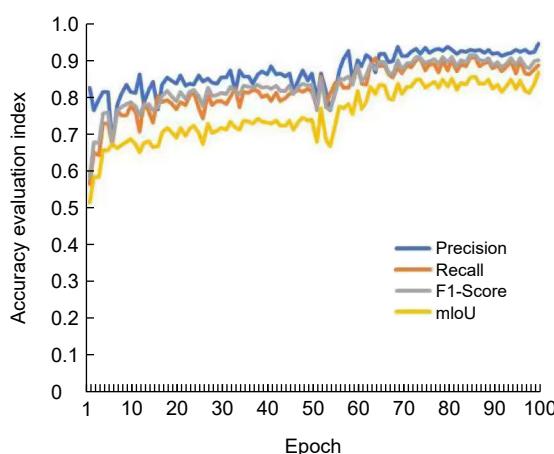


图 7 改进模型的评价指标精度图

Fig. 7 Accuracy diagram of evaluation index of the improved model

样本均表现出保持良好的分割能力以及连通性, 即使在目标差异大、边界模糊和样本分布不均衡的场景下依然表现良好。可以看出, 所提方法取得了最佳的成果, mIoU 提升了 12.48%, 相较于其他算法提升了 4.19%~12.48%。实验结果表明, 本文方法优于其他方法。

如图 8 所示, 即使建筑物周围包含复杂场景依旧可以得到良好的分割结果。由图 8 红框所示, FCN、Unet 对于目标尺度差异大的情况下存在漏检现象, SegNet 和 DeepLabv3+在此情况下检测效果并不理想, 本文采用 New\_GhostNetv2 网络增强模型对特征信息的捕捉能力, 使得其在目标初读差异大时依然可以保持良好分割。如图 8 蓝框所示, 在样本分布不均衡和

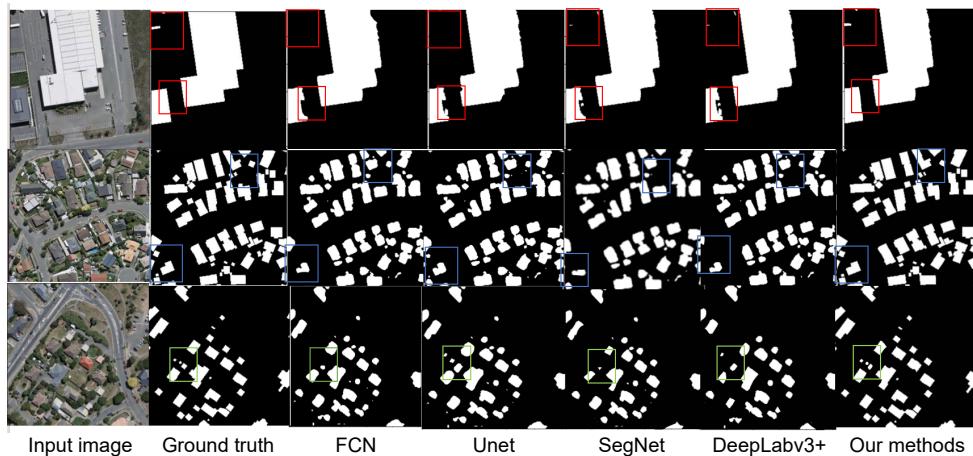


图 8 不同模型分割效果对比图

Fig. 8 Comparison of segmentation effects of different models

地物边界模糊的情况下, FCN、Unet、SegNet 以及 DeepLabv3+对于建筑物分割均存在漏检以及检测模糊现象, 本文算法采用了 Ghost 卷积结合跳跃连接和特征分支策略, 构建多层级信息增强模块, 有效减少边界模糊以及分割误差现象, 提升模型在处理多样化地物形态时的适应性。如图 8 绿框所示, 在场景区域跨度大以及背景差异大的情况下, FCN 的分割结果对目标边界的处理模糊, 而 Unet 通过跳跃连接有效保留了部分边缘信息, 使得分割结果更接近 Ground truth, SegNet 以及 DeepLabv3+检测效果仍不尽人意, 本文设计动态特征增强器, 依据特征空间的相关性对特征图进行逐通道跨尺度自适应融合, 进一步增强了模型对全局关键特征和局部细粒度特征的捕获能力。在连接的建筑物以及场景区域跨度大的情况下, 所提方法在确保边缘分割效果良好的同时, 能够对非建筑物部分进行有效区分和分割, 相对其他模型都有更好的效果。

## 4 结 论

为解决语义分割中的对于建筑物细节提取不够精准以及特征融合效率的问题, 本文提出了改进的 DeepLabv3+算法, 通过改进主干网络 New\_GhostNetV2, 使用 Ghost module 减少计算量, 同时保证特征图的效率, 改善了分割失误的问题。构建信息整合模块, 避免缺失上下文信息导致的特征信息流断裂与失准问题。其次, 引入注意力机制使其更好的关注不同地方的特征组, 避免信息整合时产生过多的冗余信息, 影响模型的性能。最后, 在特征融合模块中, 搭建动态深度特征增强器, 增强模型对多尺度及

小目标特征自适应能力, 优化算法处理复杂数据时的稳定性与泛化能力, 降低模型参数量。在 WHU 数据集中的实验结果表明, 各项数值均有明显提升, 相比于其他算法有明显优势, 通过本文改进, 进一步提升了模型的在实际中应用的价值。在未来的研究中, 将进一步考虑算法精度以及复杂度之间的平衡, 在边缘算法中实现更好的效果, 进一步提升算法实际价值。

## 参考文献

- [1] Xu S J, Jing Y, Li H T, et al. Progressive multi-granularity ResNet vehicle recognition network[J]. *Opto-Electron Eng*, 2023, **50**(7): 230052.  
徐胜军, 荆扬, 李海涛, 等. 渐进式多粒度 ResNet 车型识别网络[J]. *光电工程*, 2023, **50**(7): 230052.
- [2] Pan L L, Shao J F. Multi-resolution point cloud completion fusing graph attention[J]. *Laser Technol*, 2023, **47**(5): 700–707.  
潘李琳, 邵剑飞. 融合图注意力的多分辨率点云补全[J]. *激光技术*, 2023, **47**(5): 700–707.
- [3] Wang C, Wu X H, Zhang Y Q, et al. Recognition and segmentation of maize seedlings in field based on dual attention semantic segmentation network[J]. *Trans Chin Soc Agric Eng*, 2021, **37**(9): 211–221.  
王璨, 武新慧, 张燕青, 等. 基于双注意力语义分割网络的田间苗期玉米识别与分割[J]. *农业工程学报*, 2021, **37**(9): 211–221.
- [4] Fu Q F, Lu M, Zhang Z Y, et al. Water level monitoring method based on semantic segmentation[J]. *Laser Optoelectron Prog*, 2022, **59**(4): 0410004.  
傅启凡, 路茗, 张质懿, 等. 基于语义分割的水位监测方法研究[J]. *激光与光电子学进展*, 2022, **59**(4): 0410004.
- [5] Meng J X, Zhang L, Cao Y, et al. Optimization of image semantic segmentation algorithms based on Deeplab v3+[J]. *Laser Optoelectron Prog*, 2022, **59**(16): 1610009.  
孟俊熙, 张莉, 曹洋, 等. 基于 Deeplab v3+的图像语义分割算法优化研究[J]. *激光与光电子学进展*, 2022, **59**(16): 1610009.
- [6] Pan S, Li J W, Jiang J W. A street view semantic

- segmentation algorithm based on DeeplabV3+ architecture[J]. *Proc SPIE*, 2023, **12717**: 127172D.
- [7] Harkat H, Nascimento J M P, Bernardino A, et al. Assessing the impact of the loss function and encoder architecture for fire aerial images segmentation using deeplabv3+[J]. *Remote Sens*, 2022, **14**(9): 2023.
- [8] Das S, Fime A A, Siddique N, et al. Estimation of road boundary for intelligent vehicles based on DeepLabV3+ architecture[J]. *IEEE Access*, 2021, **9**: 121060–121075.
- [9] Su Y A, Lin Y, Fang X B, et al. Improved DeepLabV3+ network segmentation method for urban road scenes[C]//2022 IEEE 10th Joint International Information Technology and Artificial Intelligence Conference (ITAIC), 2022: 1274–1280. <https://doi.org/10.1109/ITAIC54216.2022.9836922>.
- [10] Guo L L, Zhou J Z. A lightweight network for semantic segmentation of road images based on improved DeepLabv3+[C]//2022 5th International Conference on Pattern Recognition and Artificial Intelligence (PRAI), 2022: 832–837. <https://doi.org/10.1109/PRAI55851.2022.9904092>.
- [11] Li Z Y, Wang R, Zhang W, et al. Multiscale features supported DeepLabV3+ optimization scheme for accurate water semantic segmentation[J]. *IEEE Access*, 2019, **7**: 155787–155804.
- [12] Liu S W, Cui Z Y, Li D Y. Multi-task learning for building object semantic segmentation of remote sensing image based on Unet network[J]. *Remote Sens Land Resour*, 2020, **32**(4): 74–83.  
刘尚旺, 崔智勇, 李道义. 基于 Unet 网络多任务学习的遥感图像建筑地物语义分割[J]. *国土资源遥感*, 2020, **32**(4): 74–83.
- [13] Wang J, Zhang X L, Yan T H, et al. DPNet: dual-pyramid semantic segmentation network based on improved deeplabv3 plus[J]. *Electronics*, 2023, **12**(14): 3161.
- [14] Kwak J, Sung Y. DeepLabV3-refiner-based semantic segmentation model for dense 3D point clouds[J]. *Remote Sens*, 2021, **13**(8): 1565.
- [15] Dong W K, Gong S C. Hazy images segmentation method based on improved DeeplabV3[J]. *Acad J Comput Inf Sci*, 2023, **6**(5): 21–29.
- [16] Basir S, Aziz N A A, Abiddin N Z. Semantic segmentation of paddy parcels using deep neural networks based on DeepLabV3[C]//Proceedings of the 10th International Conference on Geographical Information Systems Theory, Applications and Management, 2024, **1**: 173–180. <https://doi.org/10.5220/0012698200003696>.
- [17] Nie Z, Xu J C, Zhang S C. Analysis on DeepLabV3+ performance for automatic steel defects detection[Z]. arXiv: 2004.04822, 2020. <https://doi.org/10.48550/arXiv.2004.04822>.
- [18] Cabrera I, Zhou Y X, Ngo E, et al. Image segmentation using transfer learning with DeepLabv3 to facilitate photogrammetric limb scanning[Z]. <https://doi.org/10.36227/techrxiv.19742488.v1>.
- [19] Ikeda R, Hotta K. Feature sharing cooperative network for semantic segmentation[C]//Proceedings of the 16th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, 2021, **5**: 577–584. <https://doi.org/10.5220/0010312505770584>.
- [20] Quan B, Liu B Y, Fu D C, et al. Improved Deeplabv3 for better road segmentation in remote sensing images[C]//2021 International Conference on Computer Engineering and Artificial Intelligence (ICCEAI), 2021: 331–334. <https://doi.org/10.1109/ICCEAI52939.2021.00066>.
- [21] Tang M C S, Teoh S S, Ibrahim H. Retinal vessel segmentation from fundus images using DeepLabv3+[C]//2022 IEEE 18th International Colloquium on Signal Processing & Applications (CSPA), 2022: 377–381. <https://doi.org/10.1109/CSPA55076.2022.9781891>.
- [22] Harkat H, Nascimento J M P, Bernardino A. Fire detection using residual deeplabv3+ model[C]//2021 Telecoms Conference (ConfTELE), 2021: 1–6. <https://doi.org/10.1109/ConfTELE50222.2021.9435459>.
- [23] Tang Y H, Han K, Guo J Y, et al. GhostNetv2: enhance cheap operation with long-range attention[C]//Proceedings of the 36th International Conference on Neural Information Processing Systems, 2022: 724. <https://doi.org/10.5555/3600270.3600994>.
- [24] Liu X Y, Peng H W, Zheng N X, et al. EfficientViT: Memory efficient vision transformer with cascaded group attention[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 14420–14430. <https://doi.org/10.1109/CVPR52729.2023.01386>.
- [25] Gu Z J, Jamison K W, Sabuncu M R, et al. Heritability and interindividual variability of regional structure-function coupling[J]. *Nat Commun*, 2021, **12**(1): 4894.
- [26] Deng G H, Wu D Z, Chen W Z. Attention guided food recognition via multi-stage local feature fusion[J]. *Comput Mater Continua*, 2024, **80**(2): 1985–2003.
- [27] Chen Y P, Dai X Y, Liu M C, et al. Dynamic convolution: Attention over convolution kernels[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 11027–11036. <https://doi.org/10.1109/CVPR42600.2020.01104>.
- [28] Ji S P, Wei S Q, Lu M. Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery data set[J]. *IEEE Trans Geosci Remote Sens*, 2019, **57**(1): 574–586.
- [29] Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation[C]//18th International Conference on Medical Image Computing and Computer-Assisted Intervention, 2015: 234–241. [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28).
- [30] Badrinarayanan V, Kendall A, Cipolla R. SegNet: A deep convolutional encoder-decoder architecture for image segmentation[J]. *IEEE Trans Pattern Anal Mach Intell*, 2017, **39**(12): 2481–2495.
- [31] Chen L C, Zhu Y K, Papandreou G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]//Proceedings of the 15th European Conference on Computer Vision (ECCV), 2018: 833–851. [https://doi.org/10.1007/978-3-030-01234-2\\_49](https://doi.org/10.1007/978-3-030-01234-2_49).
- [32] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 3431–3440. <https://doi.org/10.1109/CVPR.2015.7298965>.

## 作者简介



肖振久 (1968-), 男, 副教授, 硕士生导师, 主要从事机器学习和图像与视觉信息计算方面的研究。

E-mail: [xiaozhenjiu@lntu.edu.cn](mailto:xiaozhenjiu@lntu.edu.cn)



张杰浩 (2000-), 女, 硕士研究生, 主要从事遥感图像目标检测方面的研究。

E-mail: [zjhaohao0409@163.com](mailto:zjhaohao0409@163.com)



【通信作者】田昊 (2000-), 男, 硕士研究生, 主要从事遥感图像语义分割方面的研究。

E-mail: [t18846338065@163.com](mailto:t18846338065@163.com)



曲海成 (1981-), 副教授, 硕士生导师, 副院长, CCF 会员, 主要从事遥感影像高性能计算、视觉信息计算、目标检测与识别研究。

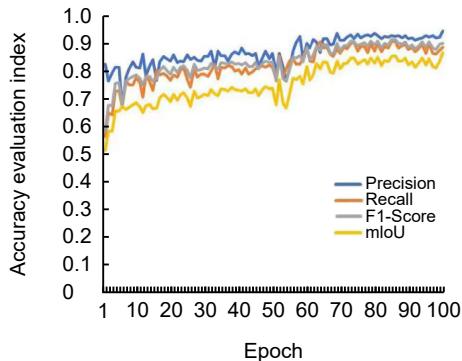
E-mail: [quhaicheng@lntu.edu.cn](mailto:quhaicheng@lntu.edu.cn)



扫描二维码, 获取PDF全文

# Fusion of dynamic features enhances remote sensing building segmentation

Xiao Zhenjiu, Tian hao\*, Zhang Jiehao, Qu Haicheng



Accuracy diagram of evaluation index of the improved model

**Overview:** Due to the complex and variable characteristics of objects in remote sensing images, such as shape, size, texture, etc., objects may overlap with each other, and at the same time, they are affected by environmental factors such as atmospheric conditions, cloud cover, and changes in lighting, which leads to a decline in image quality and increases the difficulty of accurate segmentation. Modern deep learning technology has enabled semantic segmentation models to show strong robustness and recognition ability in complex scenes, but due to the differences in regions and application scenarios, there still exist problems such as insufficient capture of complex scene details, insufficient capture of long-distance dependencies, and difficulty in integrating multi-scale features, which makes the research in the direction of semantic segmentation very important. Aiming at the poor segmentation effect caused by large scale difference of objects, uneven spatial distribution of samples, fuzzy boundaries of objects and large span of scene area, this paper proposes a high-precision remote sensing building segmentation algorithm enhanced by integrating dynamic features. Firstly, the New\_GhostNetV2 network is constructed, and the adaptive context-aware convolution is used to strengthen the discriminative ability of geometric deformation and the recognition ability of strong correlation features, improve the algorithm's ability to capture the local spatial features of samples and the global long-term dependence relationship, and realize the preliminary learning of sample images. Secondly, multi-level information enhancement modules are designed using Ghost Convolution combined with skip connections and feature branching strategies to solve the problem of key information loss and feature ambiguity caused by subsampling, enhance feature interaction and integration, and effectively reduce boundary ambiguity and segmentation errors. Then, the feature fusion module is constructed by the dynamic depth feature enhancer. According to the spatial position correlation of the feature, the feature mapping is self-adapted by channel and cross-scale to further strengthen the model's ability to mine and capture global key features and local fine-grained features, and improve the algorithm's attention to small objects. Finally, a cascade grouping attention mechanism is introduced to adjust the proportion of low-level and high-level features layer by layer, effectively suppressing background interference, and gradually optimize the segmentation results, so that the model can better cope with the diversified features in the image. Experimental results on the WHU dataset show that, compared with the baseline model, the improved algorithm is 8.57% higher than F1-Score, 12.48% higher than mIoU, 13.28% higher than Recall and 12.13% higher than precision. Compared with other mainstream semantic segmentation models, the improved DeepLabv3+ has better segmentation accuracy, effectively improving semantic segmentation performance.

Xiao Z J, Tian H, Zhang J H, et al. Fusion of dynamic features enhances remote sensing building segmentation[J]. *Opto-Electron Eng*, 2025, 52(3): 240231; DOI: [10.12086/oee.2025.240231](https://doi.org/10.12086/oee.2025.240231)

Foundation item: Basic Scientific Research Project of Liaoning Provincial Universities (LJKM20220699)

School of Software, Liaoning University of Engineering and Technology, Huludao, Liaoning 125105, China

\* E-mail: t18846338065@163.com