计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
[doi: 10.15888/j.cnki.csa.009856] [CSTR: 32024.14.csa.009856]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

ResMobileNet: 面向云和云影分割的主次残差双支路网络①

陆　楠[1], 朱亚楠[1], 王键翔[1], 闫飞一[2], 付　瑞[3]

[1](南京信息工程大学 自动化学院, 南京 210044)
[2](南京工业大学 经济与管理学院, 南京 211816)
[3](南京大学 高端控制与智能运维研发中心, 苏州 215163)
通信作者: 朱亚楠, E-mail: ynzhu@nuist.edu.cn

摘　要: 云和云影分割是遥感图像处理的关键任务, 传统深度学习方法常面临漏检、误检和细节丢失等问题. 为解决这些挑战, 本文提出了一种结合 ResNet34 和 MobileNetV3 的双支路架构. 首先, MobileNetV3 作为次残差支路, 进行初步特征提取, 这一步旨在减少在处理简单特征时的计算负担和参数量. 然后, 将初步特征送入主残差支路 ResNet34 中进行深层特征提取. 为避免最大池化操作带来的信息丢失, 设计了多尺度条带卷积池化模块 (multi-scale strip convolutional pooling module, MS-SCPM), 通过多种池化和条形卷积提取特征, 保留重要细节. 为融合多尺度信息并有效检测小目标, 引入了注意力动态金字塔多尺度特征提取模块 (attention-based dynamic pyramid multi-scale feature extraction module, ADPMFEM), 灵活捕捉关键特征并抑制冗余信息. 解码器部分采用了注意力特征感知重组模块 (content-aware reassembly of features with attention, CWA), 通过特征图权重优化上采样过程, 改善边缘恢复效果, 提升分割精度. 最后, 在像素分类之前引入可变形卷积进一步优化分割效果. 实验结果表明, 所提模型在 Biome 8、HRC-WHU 和 SPARCS 数据集上表现优异, *MIoU* (mean intersection over union) 分别提升至 79.19%、90.41% 和 77.89%, 优于现有技术. 该成果可应用于遥感领域中的云和云影图像分析, 如环境监测、灾害评估和农业监控等领域, 提升数据处理精度和效率.

关键词: 语义分割; 遥感; 云检测; 特征融合; 深度学习; ResNet

引用格式: 陆楠,朱亚楠,王键翔,闫飞一,付瑞.ResMobileNet: 面向云和云影分割的主次残差双支路网络.计算机系统应用. http://www.c-s-a.org.cn/1003-3254/9856.html

ResMobileNet: Primary-secondary Residual Dual-branch Network for Cloud and Cloud Shadow Segmentation

LU Nan[1], ZHU Ya-Nan[1], WANG Jian-Xiang[1], YAN Fei-Yi[2], FU Rui[3]

[1](School of Automation, Nanjing University of Information Science & Technology, Nanjing 210044, China)
[2](School of Economics & Management, Nanjing Tech University, Nanjing 211816, China)
[3](Center for Advanced Control and Smart Operations, Nanjing University, Suzhou 215163, China)

Abstract: Cloud and cloud shadow segmentation is a key task in remote sensing image processing, where traditional deep learning methods often encounter problems such as missed detection, error detection, and loss of detail. To address these challenges, this study proposes a dual-branch architecture combining ResNet34 and MobileNetV3. First, MobileNetV3 is used as the secondary residual branch for preliminary feature extraction, aiming to reduce computational burden and parameter count when processing simple features. The preliminary features are then passed into the primary residual branch, ResNet34, for deep feature extraction. To avoid the information loss caused by max pooling operations, a multi-

① 基金项目: 国家自然科学基金 (62203224, 12302032); 高校哲学社会科学研究项目 (44205270)
收稿时间: 2024-11-01; 修改时间: 2024-11-29; 采用时间: 2024-12-16; csa 在线出版时间: 2025-04-30

scale strip convolutional pooling module (MS-SCPM) is designed, which extracts features through various pooling and strip convolution methods to preserve important details. To fuse multi-scale information and effectively detect small targets, an attention-based dynamic pyramid multi-scale feature extraction module (ADPMFEM) is introduced, which flexibly captures key features while suppressing redundant information. The decoder uses a content-aware reassembly of features with attention (CWA) module, which optimizes the upsampling process through feature map weighting to improve edge recovery and enhance segmentation accuracy. Finally, deformable convolutions are introduced before pixel classification to further optimize the segmentation results. Experimental results show that the proposed model performs excellently on the Biome 8, HRC-WHU, and SPARCS datasets, with the mean intersection over union (*MIoU*) reaching 79.19%, 90.41%, and 77.89%, respectively, outperforming existing methods. This achievement can be applied to image analysis of clouds and cloud shadows in remote sensing domains, including environmental monitoring, disaster assessment, and agricultural surveillance, improving data processing accuracy and efficiency

Key words: semantic segmentation; remote sensing; cloud detection; feature fusion; deep learning; ResNet

云分割和云影分割在遥感、气象、环境监测及资源管理等领域扮演着重要角色. 云的变化不仅是气象指标, 还能反映气候变化的动态. 精准的云检测可以显著提高天气预报的可靠性, 有助于预防极端天气的发生, 并监测环境变化的趋势及其影响[1,2]. 然而, 由于大约 67% 的地球表面被云层覆盖[3], 许多自然灾害 (如洪水和森林火灾) 可能会因云层遮挡而难以被有效监测和评估, 准确的云与云影分割对于及时识别受灾区域至关重要, 有助于指导救援和应急响应. 此外, 云和云影分割还对土地覆盖评估和太阳能资源的优化分配具有重要意义, 从而推动农业和可再生能源的发展. 传统的检测方法主要由阈值分割技术[4,5]和人工特征提取方法[6]组成. 它们适用于大多数场景的云检测任务, 可以根据不同场景和需求进行灵活调整, 易于优化. 但这些方法很容易受到光线和亮度等因素的影响, 在复杂的情况下效果不佳.

近年来, 卷积神经网络 (CNN) 在计算机视觉任务中表现出色. 通过深度学习, 模型能够利用训练数据集实现端到端的训练并提升预测性能[7,8]. Long 等[9]在 2015 年提出全卷积网络 (FCN), 通过卷积层替代全连接层, 实现了像素级分类, 提升了语义分割效果. Ronneberger 等[10]开发了 U-Net, 采用对称的编码器-解码器结构, 通过跳跃连接结合高分辨率特征与上下文信息, 实现精确分割. Chen 等[11]提出的 DeepLab 通过空洞卷积捕捉多尺度上下文信息, 从而有效地增强了分割模型对物体边界的识别能力. Zhao 等[12]提出的 PSPNet 通过金字塔池模块整合多尺度特征, 增强全局信息获取能力. Yu 等[13]提出 BiSeNet 通过双边架构提取细节和语义信息, 实现了更好的分割效果. 刘云等[14]结合空间金字塔卷积和空间金字塔池化作为多尺度特征单元, 提出了轻量级分割模型 MiniNet. 龙丽红等[15]利用 Dilation 技术拓宽空洞卷积模块, 提出 U-Net 变体 D-UNet (Dliated-UNet). 尽管现有的图像分割方法在许多应用中表现出色, 但在复杂场景下, 传统卷积方法由于其固定的感受野, 常难以有效捕捉物体的多样性和细节变化. 虽然深度可分离卷积在计算效率上有所提升, 但仍然依赖于局部特征, 难以全面把握场景的全局结构信息. 同时, 虽然全局特征提取方法能够提供整体视图, 但往往忽视了局部细节的重要性. 这些挑战使得在特定任务中, 提升分割精确度和鲁棒性仍然显得尤为关键.

近年来, Transformer 在计算机视觉 (CV) 领域逐渐崭露头角, 取得了显著的进展. 最初设计用于自然语言处理任务的 Transformer, 如今在各种视觉任务中同样展现出优异的性能. 与传统的全局平均池化方法不同, Transformer 通过多头注意机制, 可以在关注重要区域的同时有效整合全局信息, 自注意机制的关键在于捕捉不同像素间的相互关系. 在计算自注意力时, 所有像素参与信息处理, 但各自的影响力不同, 从而实现了对全局信息的把握与对关键区域的关注. 这一特性使得 Transformer 在复杂场景中的应用更加灵活且高效. 最近的一些研究[16–18]对 Transformer 进行了扩展以获得全局特征. Wang 等[19]提出了金字塔式的 ViT (PVT), 该网络采用 Transformer 作为 ViT 的主干, 并将金字塔结构引入到 Transformer 中. Swin Transformer[20]引入

了分层设计和滑动窗口方法, 在处理图像时效率更高, 在处理多尺度任务时更灵活. Wu 等[21]提出卷积视觉 Transformer (convolutional vision Transformer, CvT), 在 ViT 中引入了卷积, 提高了 Transformer 的性能, 取得了最好的效果. Hu 等[22]用参数密集度较低的 EdgeViT 取代双分支网络的 ViT, 从而提高了网络的推理速度.

在云和云影的语义分割中, Chen 等[2]提出了一种多尺度条状特征注意网络, 实现了高分辨率可见光谱图像的端到端云和云阴影检测. Gu 等[23]提出一种多路径 Transformer 和 CNN 组合网络来实现可见光谱高分辨率遥感图像的云影和云影语义分割, 该体系结构在检测云和云影的小目标和分割边界方面取得了良好的效果. Lu 等[24]将 ViT 引入到 CNN 网络中, 提出了一种用于云和云影检测的双分支网络, 在达到较高准确率的同时, 也表现出了良好的鲁棒性和泛化能力. 李远禄等[25]通过结合 ViT 和 D-UNet, 提出了关注局部和全局信息的双分支遥感云影检测方法. 杨军等[26]结合自注意力机制和深度可分离卷积 (depthwise separable convolution, DSC) 提出适用于高分辨率遥感影像语义分割的线性多头自注意力 (linear multi-head self-attention, LMSA) 网络模型. 云和云影分割技术在遥感分析中具有重要的实际应用价值. 随着深度学习方法的不断发展, 这些技术在准确性和效率上都取得了显著进步.

此外, 双分支网络通过同时结合两种不同的支路, 显著提升了语义分割精度. BiSeNet[13]采用空间路径分支保留高分辨率细节, 捕捉局部特征; 采用上下文路径分支获取全局语义信息, 增强分割效果. ExtremeC3Net[27]平衡了计算效率与精度, 适合实时场景. DANet[28]通过空间和通道维度的双重注意力机制, 提升了对复杂场景的理解和精度. 这些设计使得双分支网络在分割任务中表现更好. 受上述双分支网络的启发, 本文提出了一种主次残差双分支网络, 它通过两个残差网络分支 ResNet34[29]和 MobileNet V3[30]分别进行深度特征提取和初步特征提取. 这种设计旨在结合两个网络的优势, 以提高云和云影分割的准确性和效率, 本文贡献如下.

(1) 针对 ResNet34 中最大池化层可能导致的细节和重要信息丢失等问题, 本文提出了一种多尺度条带卷积池化模块 (MS-SCPM). 该模块通过结合多种池化操作和条形卷积操作, 提供了一种更灵活和多样化的特征提取方式. 这种方法不仅能够保留关键的图像细节, 还能够增强模型对不同尺度特征的捕捉能力.

(2) 在特征提取阶段, 为克服无法获得全局上下文信息以及难以适应不同尺度和形状的云和云影的挑战, 本文提出了一种注意力动态金字塔多尺度特征提取模块 (ADPMFEM), 该模块使得池化操作能够更好地适应输入特征图的内容和大小, 灵活捕捉多尺度信息. 通过这种方式, 模型能够减少不必要的计算, 同时减少信息丢失, 从而提高分割的准确性.

(3) 为了更好地恢复云影图像中的细节信息, 在解码阶段, 本文提出了基于注意力的特征内容感知重组模块 (CWA) 上采样方法. 这种方法是对 CARAFE 的改进, 旨在提高边界和细节的分割精度. CWA 通过注意力机制增强了特征图的权重指导, 从而在上采样过程中更有效地恢复边缘和细节, 提升分割结果的质量.

1 基础网络及相关模块

1.1 模型结构

本文提出了一种名为 ResMobileNet 的双支路架构, 结合 ResNet34 和 MobileNetV3 的优势, 旨在实现复杂场景下云和云影的精确检测. 该架构通过高效的边界处理和细节恢复, 能够有效处理各种规模的图像. 网络结构如图 1 所示, 其中 up1、up2 和 up3 分别表示第 1 层、第 2 层和第 3 层转置卷积上采样. 图 2–图 7 展示了各个模块的具体细节. 对于任意大小的输入图像, ResMobileNet 通过 MobileNetV3 提取这 3 层低级别的多尺度特征, 并通过转置卷积将这些特征上采样到与原图相同的尺寸, 随后将这 3 层低级别特征进行拼接, 从而确保计算效率的同时, 获得丰富的多尺度特征. 接着, 这些特征被传递至 ResNet34 进行深层特征提取, 充分利用 ResNet34 的深层残差结构增强表示能力, 同时避免直接将特征图送入主干网络所带来的计算和资源消耗. 为防止简单使用最大池化层时的信息丢失, 提出了多尺度条带卷积池化模块 (MS-SCPM), 通过多种池化和条形卷积操作灵活提取特征, 同时保留细节. 在深层特征提取阶段, 引入了注意力动态金字塔多尺度特征提取模块 (ADPMFEM), 该模块能够动态捕捉多样化特征, 抑制冗余信息, 防止特征丢失. 解码器部分采用基于注意力的特征内容感知重组模块 (CWA), 生成特征图权重, 提升对关键区域的关注, 避免边缘恢复不佳, 提升细节清晰度. 最后, 在像素分类之前引入可变形卷积, 进一步提升了分割的准确性.

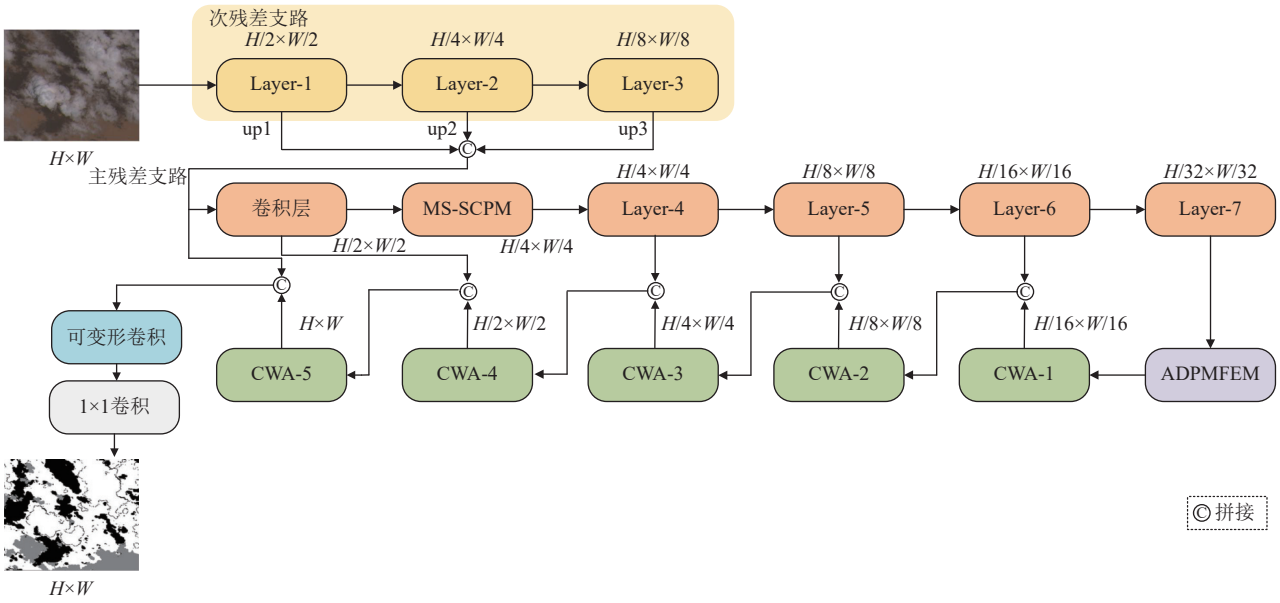


图 1 双支路网络总体架构

总体而言, ResMobileNet 结合了 MobileNetV3 的轻量化特性和 ResNet34 的深层特征学习能力, 通过多样化模块设计与特征融合策略, 构建了高效、强大的云和云影检测模型, 能够在复杂场景中提供高精度和高效率的分割结果.

1.2 骨干网络

本文中的双支路骨干网络分别采用 MobileNetV3 和 ResNet34, 两者均为残差网络. MobileNetV3 采用了最新的架构搜索技术和先进的特性 (如 H-swish 和 SE 模块), 在保持轻量级的同时, 相比 V1 和 V2 版本提供了更高的效率和性能. 因此本文选择 V3 版本作为次级支路的骨干网络, 负责初步的特征提取. 它通过深度可分离卷积来显著节约计算资源, 同时借助 squeeze-and-excitation (SE) 模块进一步提升特征提取能力. MobileNetV3 的倒立残差结构是实现高效特征提取的关键部分, 包括 3 个主要步骤: 扩展、深度可分离卷积和压缩. 这种结构不仅优化了网络的计算效率, 还增强了模型对特征的捕捉能力. MobileNetV3 的残差结构如图 2 所示, 清晰地展示了其高效的特征提取流程.

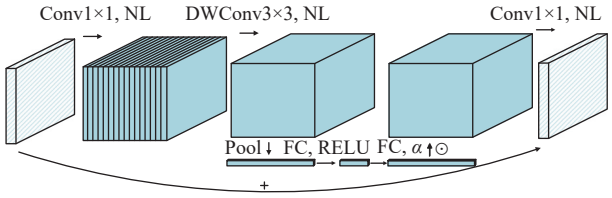


图 2 MobileNetV3 bottleneck 示意图

倒立残差结构首先对输入特征进行通道扩展, 通过一个卷积核大小为 1×1 的卷积$Conv_{1\times1}$将通道数从输入通道数C_{in}扩展到$C_e = t \cdot C_{\text{in}}$, 其中 t 是扩展因子. 该步骤的公式如下:

$$X_e = ReLU6(Conv_{1\times1}(X_{\text{in}}, C_e)) \tag{1}$$

其中, X_{in}表示输入特征图, X_e表示输出特征图, $ReLU6$表示激活函数. 接下来, 使用卷积核大小为 3×3 的深度可分离卷积对扩展后的特征图进行空间上的卷积. 深度可分离卷积的公式如下:

$$X_{\text{dw}} = ReLU6(DWConv_{k\times k}(X_e, C_e)) \tag{2}$$

其中, X_{dw}表示输出特征图, $DWConv_{k\times k}$表示卷积核大小为 $k\times k$ 的深度可分离卷积, 这里 k 取 3.

接着通过 squeeze-and-excitation 模块对通道进行重新加权. SE 模块包括以下部分.

1) Squeeze: 使用全局平均池化 (global average pooling, GAP) 压缩空间维度, 得到每个通道的全局信息:

$$z = GAP(X_{\text{dw}}) \tag{3}$$

2) Excitation: 通过两个全连接层产生通道的重新加权系数:

$$s = \alpha(W_2 \cdot ReLU(W_1 \cdot z)) \tag{4}$$

其中, α是 Sigmoid 激活函数, W_1和W_2是全连接层的权重矩阵.

3) 重新加权: 将权重应用到特征图的每个通道.

$$X_{se} = X_{dw} \odot s \tag{5}$$

其中, ⊙表示逐通道乘法.

然后使用$Conv_{1\times1}$将特征图的通道数从C_e压缩到C_{out}, 并且这个操作不改变空间维度. 压缩通道公式为:

$$X_{out} = Conv_{1\times1}(X_{se}, C_{out}) \tag{6}$$

其中, X_{out}表示输出特征图. 最后, 如果输入和输出的空间尺寸和通道数一致, 则使用跳跃连接将输入直接加到输出上. 跳跃连接的公式如下:

$$Y = X_{in} + X_{out} \tag{7}$$

否则, 输出就是压缩后的特征图:

$$Y = X_{out} \tag{8}$$

其中, Y表示最终特征图的输出.

次残差分支的具体结构如表 1 所示. 表中详细展示了第 1 个分支的 3 个阶段的参数设置, 这些设置精心设计以确保网络在保持计算效率的同时, 能够提取丰富的特征信息.

表 1 MobileNetV3 参数设置表

层名	输出尺寸	卷积类型	扩张的中间通道数	输出通道数	SE	NL	步幅大小
Layer-1	$256^2\times3$	Bneck, 3×3	—	16	—	HS	2
	$128^2\times3$	Bneck, 3×3	16	16	—	RE	1
	$128^2\times3$	Bneck, 3×3	64	24	—	RE	2
Layer-2	$56^2\times3$	Bneck, 5×5	72	24	—	RE	1
	$56^2\times3$	Bneck, 5×5	72	40	√	RE	2
	$28^2\times3$	Bneck, 5×5	120	40	√	RE	1
Layer-3	$28^2\times3$	Bneck, 5×5	120	40	√	RE	1
	$28^2\times3$	Bneck, 3×3	240	80	—	HS	2

ResNet34 比 ResNet18 具有更深的网络结构, 能够捕捉更复杂的特征, 但其计算需求和参数量仍远低于更深的 ResNet50、ResNet101 和 ResNet152 模型. 因此, 我们选择 ResNet34 作为主支路的骨干网络, 它既能提供较好的性能, 又能保持较低的计算成本和内存使用. 在第 2.2.4 节的消融实验中, 我们充分证明了选择 ResNet34 作为骨干网络的优越性. 它拥有强大的特征学习和表达能力, 通过残差连接有效克服了深层网络中常见的梯度消失问题, 使得网络能够深入地提取更丰富的特征. 为应对 ResNet34 中的最大池化层可能导致信息丢失的问题, 尤其是在处理细节丰富的图像时. 本文提出了 MS-SCPM 模块作为替代方案, 将在第 1.3 节详细讨论该模块的设计与应用. 此外, ResNet34 中传统的卷积操作通常仅限于处理具有有限感受野的局部区域, 限制了网络捕获远程特征相关性的能力. 因此, 我们将 ResNet34 残差块中的所有卷积替换为条状卷积[31]. 条状卷积通过采用非方形卷积核, 能够在特定方向上捕捉长条形、线性或带状结构的特征, 这在云图像处理中尤其有效. 云层通常以大片或带状的形式分布, 条状卷积能够在水平或垂直方向上更有效地提取这些特征. 这种改进提升了模型对云边缘、云层形态及其空间分布的感知能力, 从而增强了云和云影分割的准确性. 改进后的 ResNet34 残差结构如图 3 所示, 展示了如何通过这些创新的调整, 优化网络结构以更好地适应云和云影分割任务的需求.

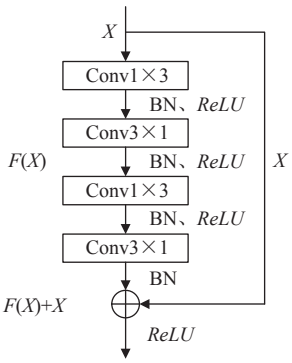


图 3 改进的 ResNet34 残差结构

改进的 ResNet34 残差结构通过一系列精心设计的卷积操作进行处理, 以提取和增强关键特征. 特征图X_{in}首先经过卷积核大小为 1×3 的条状卷积$Conv_{1\times3}$和 3×1 的条状卷积$Conv_{3\times1}$, 从而得到中间特征图X_1:

$$X_1 = ReLU(BN(Conv_{3\times1}(BN(Conv_{1\times3}(X_{in}, C_{mid}))), C_{mid})) \tag{9}$$

其中, $ReLU$表示激活函数, BN表示归一化层, C_{mid}表示中间层的通道数, 特征图再经过$Conv_{1\times3}$和$Conv_{3\times1}$得到输出特征图X_{out}:

$$X_{out} = (BN(Conv_{3\times1}(BN(Conv_{1\times3}(X_1, C_{out}))), C_{out})) \tag{10}$$

其中, C_{out}表示输出层的通道数. 如果输入输出的通道数及空间尺寸一致, 则采用跳跃连接 (identity mapping) 来进一步增强特征传递:

$$Y = ReLU(X_{in} + X_{out}) \tag{11}$$

跳跃连接允许网络直接利用输入特征, 有助于减少训练过程中的梯度消失问题, 并提高特征的利用效率. 如果通道数或空间尺寸不一致, 我们仅使用$Conv_{1\times1}$

调整X_{out}的通道数, 以匹配输入特征图的维度.

残差分支的具体结构见表 2, 详细列出了第 2 分支几个阶段的参数设置. 这些参数设置经过精心优化, 以确保网络在处理复杂图像特征时的效率和准确性.

表 2 改进的 ResNet34 参数设置表

层名	输出尺寸	参数配置
卷积层	128×128	StripConv7×7, 64, stride2
MS-SCPM	64×64	[StripConv3×3, 64 StripConv5×5, 64 StripConv7×7, 64 3×3MaxPool, 64 5×5MaxPool, 64 7×7MaxPool, 64 AvgPool, 64 MaxPool, 64]
Layer-4	64×64	[StripConv3×3, 64 StripConv3×3, 64] ×3
Layer-5	32×32	[StripConv3×3, 128 StripConv3×3, 128] ×4
Layer-6	16×16	[StripConv3×3, 256 StripConv3×3, 256] ×6
Layer-7	8×8	[StripConv3×3, 512 StripConv3×3, 512] ×3

1.3 多尺度条带卷积池化模块 MS-SCPM

在 ResNet 网络中, 最大池化层一般位于网络的前段, 作用是使网络迅速减少特征图的尺寸, 使后续残差块能够处理较小的特征图, 更专注于学习高层次特征. 然而在云和云影分割任务中, 这种方法并不总能取得理想效果, 尤其是在需要细颗粒度和高分辨率细节分类的场景下. 首先, 最大池化通过选取局部区域中的最大值来降维, 这可能导致细节信息丢失. 在云和云影分割中, 细节 (例如云的边缘、云影的形状) 至关重要, 如果最大池化层丢失这些信息, 将直接影响分割精度. 其次, 最大池化具有固定的尺度和步幅, 这限制了其适应不同大小云影区域的能力. 云和云影区域的尺度和形状各异, 而标准的最大池化层缺乏必要的灵活性来有效处理这些变化. 最后, 最大池化只保留局部区域中的最大值, 忽略了区域中的其他信息, 这在云和云影分割任务中是不利的, 因为上下文信息对于区分云影与周围背景至关重要. 为解决这些问题, 本文提出了多尺度条带卷积池化模块 MS-SCPM, 其结构详见图 4.

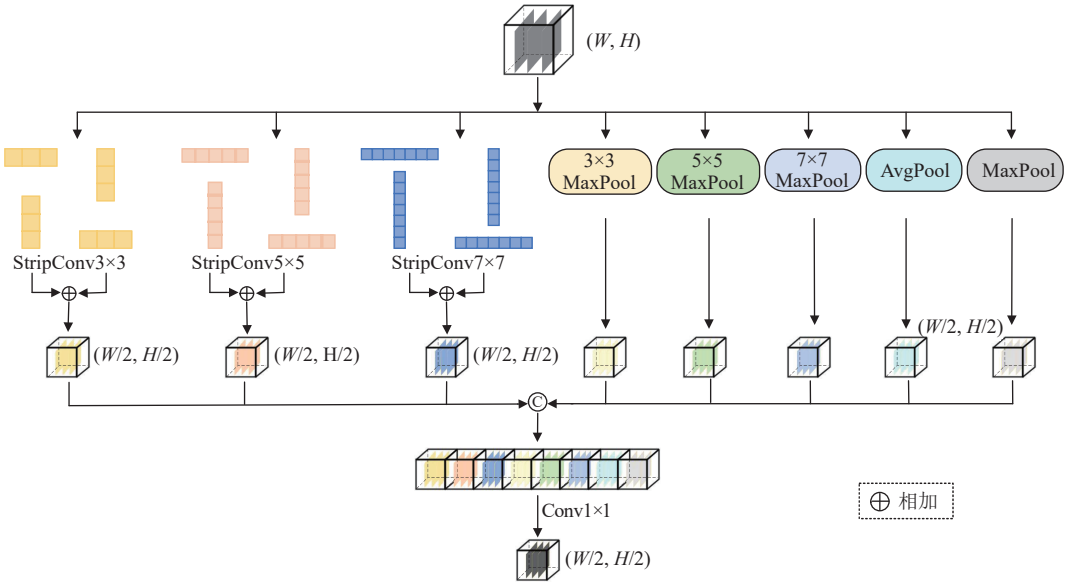


图 4 多尺度条带卷积池化模块 (MS-SCPM)

MS-SCPM 集成了 3 种不同尺度的条形卷积操作和 5 种不同尺度的池化操作, 能够在多个尺度上提取特征. 相比单一的最大池化层, MS-SCPM 能够更全面地捕捉不同尺寸的上下文信息, 并有效防止细节信息丢失. 对于云影分割任务, 不同的云影区域具有不同的空间结构, MS-SCPM 通过多尺度操作的组合, 能够更好地适应这些结构变化, 从而提高分割的精度和效率. 这种创新的模块设计, 使网络能更灵活地应对云和云影的复杂性, 为实现高精度分割提供了强有力的支持.

1.4 注意力动态金字塔多尺度特征提取模块 ADPMFEM

为了在复杂的遥感图像中提取云和云影的多尺度上下文信息, 本文深入探讨了 PSPNet[12]的金字塔池化模块. 该模块通过全局平均池化和不同大小的区域池化 (如 1×1、2×2、3×3、6×6) 来捕获多尺度信息. 然

而, 固定的池化尺度存在可能导致细节信息丢失的问题. 为了解决这一问题, 许多研究引入了自适应池化技术以提高灵活性. 尽管自适应池化在适应性上有所改进, 但其池化尺度通常仍是预设的固定值, 难以根据具体的特征图动态调整, 这可能造成信息捕获的不足或冗余. 针对这一挑战, 本文提出了动态金字塔特征提取模块 (dynamic pyramid feature extraction module, DPFEM). 该模块摒弃了传统的固定池化尺度, 而是根据输入特征图动态预测池化尺度, 从而灵活适应不同的输入内容与大小, 有效捕获更多有价值的信息. 动态预测池化尺度的过程详见图 5.

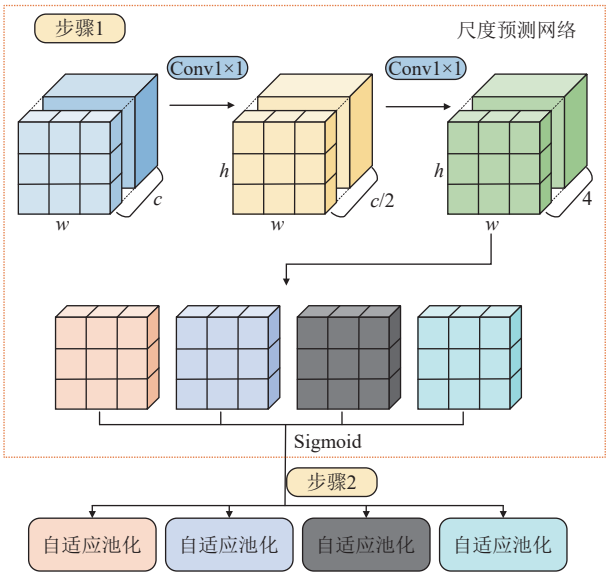


图 5 动态金字塔特征提取模块 (DPFEM)

图 5 中步骤 1 为设置池化层数量i及最大池化尺寸M和最小池化尺寸N, 例如池化层数量设置为 4, 最大池化尺寸设置为 8, 最小池化尺寸设置为 2. 假设输入特征图X的维度为 (B, C, H, W), 其中 B 为批次大小, C 为通道数, H 为高度, W 为宽度. 模型会预测每个池化尺度的大小, 具体通过一个尺度预测网络实现, 该网络由两个卷积层和一个 Sigmoid 激活函数组成, 进入网络的输入通道数是 C, 经过第 1 个卷积层降为 $C/2$, 经过第 2 个卷积层输出通道变为 i, 得到尺度预测网络的输出特征图, 维度为 (B, i, H, W), 接着对每个通道使用 Sigmoid 激活函数得到每个尺度的预测值, 公式如下:

$$s_i = \alpha(Conv_{1\times1}(Conv_{1\times1}(X))_i)\times M \tag{12}$$

其中, α表示 Sigmoid 激活函数, s_i表示第i个池化尺度的预测大小, M 是步骤 1 设置的最大池化尺寸. 我们通过步骤 2 来将预测的池化尺寸进一步约束到合理范围, 即在[N, M]之间, 这一步骤可以用以下公式表示:

$$\widetilde{s_i} = \max(N, \min(M, s_i)) \tag{13}$$

其中, $\widetilde{s_i}$是最终的池化尺寸, 确保其在[N, M]之间. 通过这种方式, 我们就动态地得到了 4 个预测池化尺度.

本文在动态金字塔特征提取模块 (DPFEM) 的基础上, 进一步融合了 DeepLabv3+[11]中的自适应空间金字塔池化 (ASPP) 模块, 以增强云和云影的分割效果. 普通卷积通常局限于处理局部区域, 而 ASPP 模块中的膨胀卷积则能够提供稳定的大尺度上下文信息. 通过结合 DPFEM 预测的自适应池化, 确保了特征提取的灵活性和多样性. 此外, 我们引入 CBAM 注意力模块[32]以实现更全面的特征提取, 从而增强整体性能. CBAM 的通道注意力机制能够增强对识别云和云影的重要特征通道的关注, 同时有效抑制无关特征, 减少噪声, 进一步提升分割精度. 在应用空间注意力后, 模型能够更准确地聚焦于关键区域, 例如, 云的边界和细节, 这不仅提高了细粒度的分割精度, 还减少了误分割的发生. 注意力动态金字塔多尺度特征提取模块的总体结构如图 6 所示.

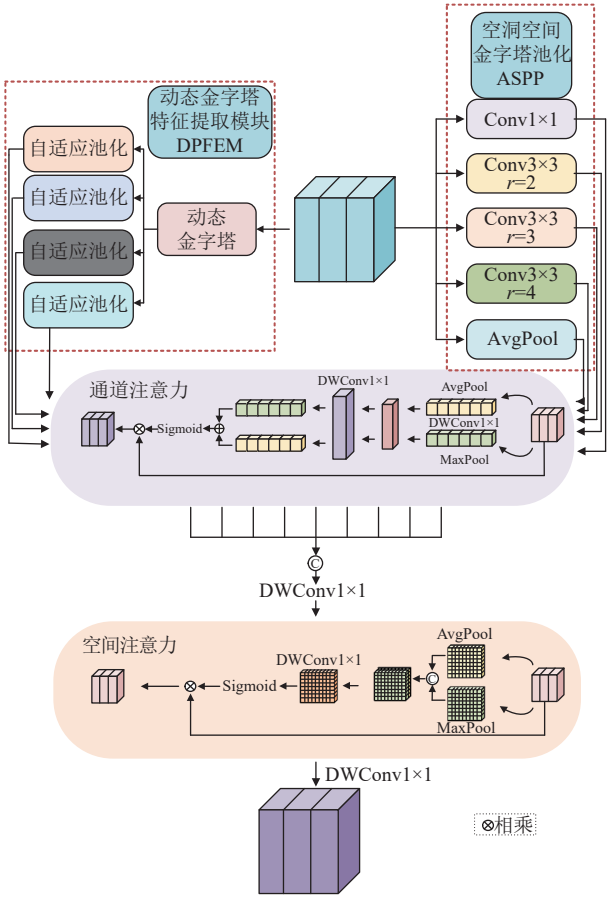


图 6 注意力动态金字塔多尺度特征提取模块 (ADPMFEM)

该结构展示了如何通过结合 DPFEM 模块、ASPP 模块和 CBAM 注意力机制, 形成一个强大的特征提取框架. 这种综合方法不仅提升了模型对局部特征的敏感度, 还增强了其对全局上下文的理解, 使得模型在处理具有复杂背景和多变特征的图像时, 能够实现卓越的分割性能.

1.5 注意力特征感知重组模块 CWA

在云和云影分割任务中, 解码器负责将低分辨率特征图恢复至与原始图像相同的分辨率, 并融合高分辨率特征以优化边界细节. 常见的上采样方法包括双线性插值和转置卷积. 双线性插值虽然简单, 但它仅依赖于相邻像素值的加权平均, 这常导致云与背景之间的边界变得模糊, 缺乏必要的细节, 并且由于其固有的非学习性质, 难以适应云和云影复杂的形状变化. 转置卷积则提供了一种自适应的解决方案, 能够根据特定任务学习上采样的参数. 然而, 它也可能带来棋盘效应、边缘模糊和不自然纹理的问题, 这些因素都可能使生成的云边界不够准确.

为了应对这些挑战, 本文提出了注意力特征感知重组模块 (content-aware reassembly of features with attention, CWA), 如图 7. CARAFE[33]的核心优势在于内容感知的上采样能力, 它通过生成特征图的权重来指导上采样过程, 有效重建和增强图像中的细节. 对于云和云影这种具有复杂边界和细节的任务, CARAFE 能够更准确地保留和恢复图像中的重要特征, 而非简单地进行插值. 此外, CWA 的注意力机制进一步提升了上采样的质量. 该机制能够自动识别并关注图像中的重要区域, 强调云和云影的关键特征, 从而提升模型整体性能. 最终, 通过在特征图的像素级别上应用权重, CWA 实现了更精细的特征重组, 这对于复杂图像结构 (如云的边缘和阴影的细节) 尤为重要. 这种精细处理方式不仅更好地保留了细节, 还显著提高了分割精度.

CWA 上采样模块首先通过特征提取器生成上采样过程中需要的权重图, 这些权重图将用于对上采样结果进行加权处理. 假设输入特征图 X 的维度为 (B, C, H, W), 其中 B 为批次大小, C 为通道数, H 为高度, W 为宽度. 特征图经过特征提取器得到权重 ϖ:

$$\varpi = DWConv_{3\times3}(DWConv_{3\times3}(DWConv_{3\times3}(X))) \quad (14)$$

得到的权重 ϖ 的维度为 (B, $C\times scale\ factor\times scale\ factor$, H, W), 然后将权重调整为 (B, C, $H\times scale\ factor$, $W\times scale\ factor$) 以适应上采样操作. 其中, *scale factor* 缩放因子, 本文中设置为 2. 接着对输入特征图进行转置卷积上采样得到上采样特征图 X', 然后将上采样特征图 X' 和权重 ϖ 进行元素级别的乘, 得到相乘后的特征图 X'':

$$X'' = X' \times \varpi \quad (15)$$

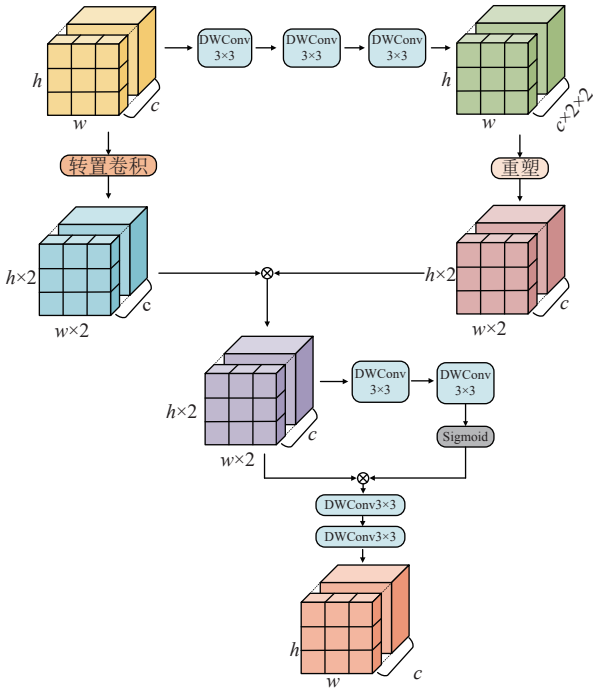


图 7 注意力特征感知重组模块 (CWA)

逐元素乘法的目的在于利用自适应的权重图 ϖ 对上采样后的特征图 X' 进行细致调整, 使模型能够更好地处理云和云影分割任务中的细节、边缘和复杂特征. 这种方法通过增强重要区域的特征表达, 有助于提高分割结果的精度, 特别是在处理复杂边界和相似特征时. 接着对 X'' 应用自适应注意力机制, 以提升特征图的质量, 它包含两个深度可分离卷积层和一个 Sigmoid 激活函数:

$$\begin{cases} att = DWConv_{3\times3}(X'') \\ att = DWConv_{3\times3}(att) \\ X''' = \alpha(att)\times X'' \end{cases} \quad (16)$$

其中, *att* 表示中间输出特征图. 自适应注意力机制通过增强特征的表达、聚焦于重要区域、抑制背景干扰、动态调节特征强度等手段, 大幅提高了模型的分割性能, 尤其是在处理复杂的云和云影形态及其相似性时, 具有显著的优势. 最后通过深度可分离卷积进行通道

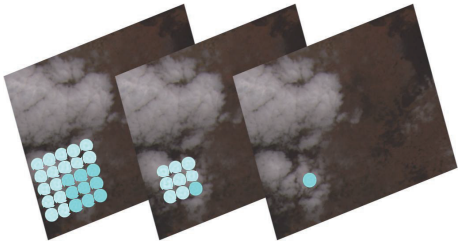
数的调整:

$$\begin{cases} X^1 = DWConv_{3\times3}(X''') \\ X^2 = DWConv_{3\times3}(X^1) \\ X^3 = GELU(X^2) \end{cases} \tag{17}$$

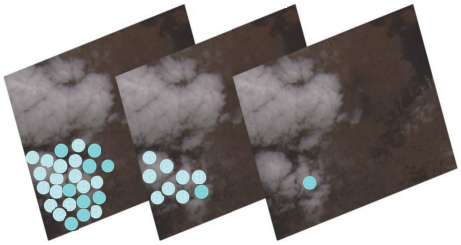
其中, $GELU$表示激活函数, X^1和X^2表示中间输出特征图, X^3表示最终输出特征图.

1.6 可变形卷积

由于云和云影的边界往往是不规则且复杂的, 云的形状、大小和结构非常多样化, 传统卷积通过固定的感受野来提取特征, 因此难以准确捕捉这些复杂的信息. 而可变形卷积[34]通过学习动态的偏移量来调整卷积核的位置和形状, 可以更好地适应这些不规则的边界, 从而在处理不同类型的云时具有更高的灵活性. 图 8 显示了普通卷积和可变形卷积的示意图.



(a) 普通卷积



(b) 可变形卷积

图 8 普通卷积和变形卷积示意图

普通卷积的操作定义简洁而经典, 其公式如下:

$$y(p_0) = \sum_{k=1}^{K} w_k \cdot X(p_0 + p_k) \tag{18}$$

其中, $y(p_0)$是输出特征图在位置p_0处的值, X是输入特征图, w_k是卷积核在位置p_k处的权重, p_k是卷积核的第k个位置对应的偏移, K是卷积核的大小. 相对于此, 可变形卷积在普通卷积的基础上引入了动态偏移量Δ_k, 公式如下:

$$y(p_0) = \sum_{k=1}^{K} w_k \cdot X(p_0 + p_k + \Delta_k) \tag{19}$$

其中, Δ_k是通过偏移量卷积层预测得到的偏移量.

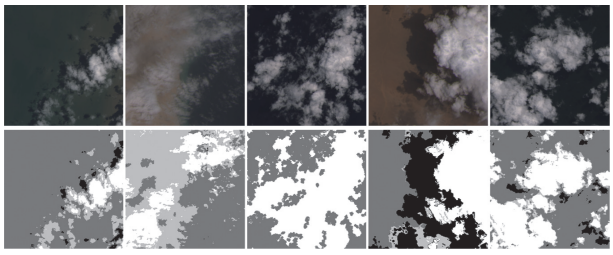
在此, 我们特别强调将可变形卷积放置在像素分类之前, 而非直接集成于卷积层中, 这一决策是基于对计算资源和特征提取效果的细致权衡. 具体而言, 在编码器中, 特征图的通道数和尺寸较大, 此时应用可变形卷积会显著增加计算负担. 相反, 在特征融合和上采样后的高级特征图中, 通道数减少, 应用可变形卷积不仅能够增强特征的空间适应性, 还能有效提升计算效率. 此外, 经过多层特征融合和上采样后, 特征图已融合了丰富的上下文信息, 在这一阶段引入可变形卷积, 能够更精准地动态调整感受野, 整合全局与局部信息, 进而提升特征表达能力, 提升像素分类精度, 并优化在复杂场景下的分割效果. 这样的配置更符合视觉系统的层级处理机制, 确保了特征处理的有效性和计算资源的合理分配性, 助力高效地解析图像.

2 实验结果与分析

2.1 数据集

2.1.1 Biome 8 数据集

Biome 8 数据集[35]用于验证分割模型, 包含 96 个场景和人工生成的地面真值数据 (GT), 覆盖 11 个样本. 针对云和云影分割任务, 数据集包括 32 个云影场景, 分为云、薄云、晴朗、阴影和空地 5 类图像. 由于 GPU 内存的限制, 我们将原始遥感图像数据集裁剪为 256×256 像素的尺寸, 并通过筛选 (去除空类别和类别分布不均的图像), 最终保留了 7012 张图像用于实验. 图 9 提供数据集的可视化示例, 包括荒地、森林、草地和城区等场景.



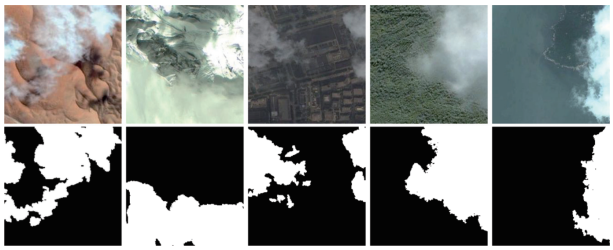
(a) 荒地 (b) 森林 (c) 草地 (d) 城市 (e) 水域

图 9 Biome 8 数据集可视化示例

2.1.2 HRC-WHU 数据集

为了更准确评估模型性能, 本文采用了 Li 等人[36]创建的 HRC-WHU (high-resolution cloud coverage validation) 数据集. 该数据集包含 150 张 1280×720 分辨率的高分辨率遥感图像, 涵盖红、绿、蓝 3 个通道, 背景包括雪地、水域、城市、沙漠和植被等多种复杂

场景. 为适应训练需求, 每张图像被裁剪为 256×256 像素, 经过严格筛选, 最终保留了 9148 张图像. 为防止过拟合, 我们对数据集进行打乱处理并应用图像增强技术, 以提高模型的泛化能力. 模型性能测试采用了云 (白色) 与背景 (黑色) 的二分类标签, 并使用五重交叉验证方法, 确保评估的一致性和可比性. 图 10 展示了数据集的可视化示例, 第 1 行为原图, 第 2 行为标签.

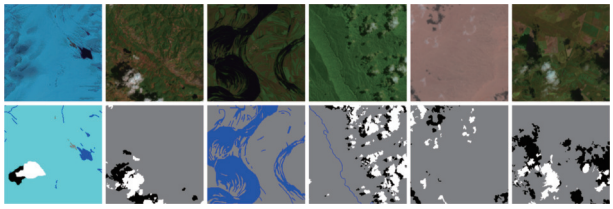


(a) 沙漠 (b) 雪地 (c) 城市 (d) 森林 (e) 水域

图 10 HRC-WHU 数据集可视化示例

2.1.3 SPARCS 数据集

我们使用 SPARCS 数据集[37]验证本文方法在多光谱场景下的性能. SPARCS 数据集包含 80 幅 1000×1000 像素的高分辨率遥感图像, 覆盖多种复杂场景. 为确保实验一致性, 我们将图像裁剪为 256×256 像素, 并通过数据增强 (包括垂直、水平翻转和随机旋转) 生成 1280 张图像. 数据集按 8:2 比例划分为训练集和验证集, 包含山丘、林地、雪地、水域、田野、沙漠等场景, 标签包括云、云影、雪/冰、水和陆地 5 类, 并使用五重交叉验证评估模型性能, 图 11 展示了典型样本, 包括不同场景的原图和标签.



(a) 雪地 (b) 山地 (c) 河流 (d) 草地 (e) 沙漠 (f) 田野

图 11 SPARCS 数据集可视化示例

2.2 评价指标和实验设置

2.2.1 优化方法

实验是在配备 RTX 4070Ti Super GPU 的系统上使用 PyTorch 框架进行的. 本研究采用了随机梯度下降 (SGD) 优化算法, 并为动量项设置了 0.9 的系数, 以实现更稳定的训练过程. 在训练中, 我们采用了聚学习率 (Poly *LR*) 策略, 初始学习率 (*LR*) 设定为 0.001, 而聚幂设置为 2. 整个模型共进行了 300 个训练周期.

为了有效预防过拟合, 我们在训练过程中引入了 L2 正则化技术, 并将权重衰减系数设置为 0.01, 这有助于引导模型学习更加泛化的特征表示. 每个训练周期的学习率 (*LR*) 按照以下衰减规律进行调整:

$$LR = 0.001 \times \left(1 - \frac{E}{300}\right)^{\gamma} \tag{20}$$

其中, E代表训练周期的序号, γ为 2. 这种学习率衰减策略有助于在训练初期快速收敛, 在训练后期则通过降低学习率来细化模型参数, 以获得更好的性能.

2.2.2 评价指标

为了全面评估本文方法在云和云影分割任务中的性能, 选用了一系列广泛认可的评价指标, 包括精度 (*P*)、召回率 (*R*)、*F*1 分数 (*BF*)、像素精度 (*PA*)、平均像素精度 (*MPA*)、平均交并比 (*MIoU*) 和频率加权交并交集 (*FWIoU*). 各评价指标的计算公式如下:

$$P = \frac{TP}{TP + FP} \tag{21}$$

$$R = \frac{TP}{TP + FN} \tag{22}$$

$$BF = 2 \times \frac{P \times R}{P + R} \tag{23}$$

$$PA = \frac{TP + TN}{TP + FP + FN + TN} \tag{24}$$

$$MPA = \frac{1}{N} \sum_{i=0}^{N} \frac{TP_i}{TP_i + FP_i} \tag{25}$$

$$MIoU = \frac{1}{N} \sum_{i=0}^{N} \frac{TP_i}{TP_i + FP_i + FN_i} \tag{26}$$

$$FWIoU = \sum_{i=0}^{N} \frac{TP_i + FN_i}{TP + TN + FP + FN} \times \frac{TP_i}{TP_i + FP_i + FN_i} \tag{27}$$

其中, 真正例 (*TP*) 表示预测为云或阴影且实际为云或阴影的像素数; 假正例 (*FP*) 表示预测为云或阴影, 但实际不是云或阴影的像素数; 真负例 (*TN*) 表示预测为背景且实际为背景的像素数. 假负例 (*FN*) 表示实际为云或阴影, 但模型预测为背景的像素数. TP_i、FP_i 和 FN_i 分别表示第 i 类别 (如云、阴影等) 的真正例、假正例和假负例, N 表示类别的数量。

2.2.3 损失函数

本文实验过程中的损失函数为交叉熵损失函数, 对于多分类问题 (如云、云影、背景等多类别的分割), 交叉熵损失函数的公式为:

$$\Psi = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} y_{i,c}\log\left(p_{i,c}\right) \tag{28}$$

其中, N是图像的像素总数, C是类别的总数, $y_{i,c}$是第i个像素的真实标签, 属于第c类时为 1, 否则为 0. $p_{i,c}$是第i个像素属于第c类的预测概率.

交叉熵损失函数能够有效处理像素级的分类问题, 它通过衡量预测的类别概率分布与真实标签间的差异来适应多分类任务. 结合 Softmax 激活函数, 交叉熵能够将模型的输出转化为每个类别的概率, 并且在训练过程中通过最小化损失优化模型的预测精度. 在云和云影的分割任务中, 交叉熵损失函数有助于提高分类的准确性, 特别是在处理复杂的遥感图像时. 通过使用交叉熵损失函数训练 Biome 8 数据集, 得到的损失收敛曲线如图 12(a) 和 (b) 所示, 分别为训练和测试损失收敛趋势. 通过细致观察损失曲线, 可以发现模型在训练过程中展现出了优异的收敛特性. 损失值在初始几轮训练中迅速降低, 随后逐渐趋于稳定, 这表明模型已经在大多数训练样本上实现了有效的拟合.

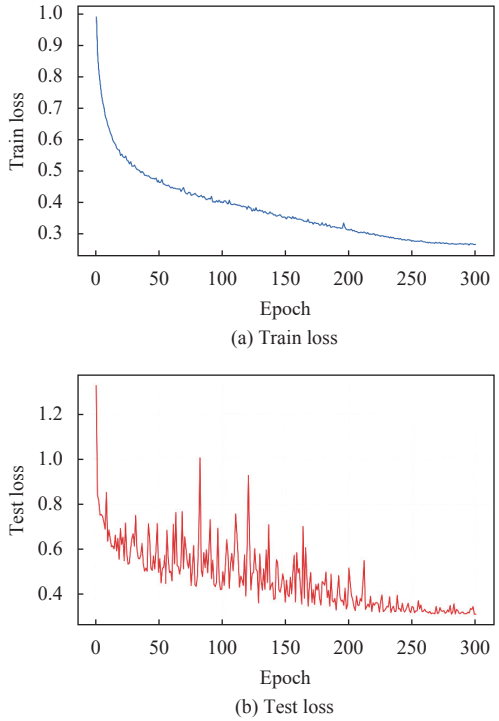


图 12 训练 Biome 8 数据集得到的损失收敛曲线

2.2.4 消融实验

(1) 主骨干网络的消融实验

选择 ResNet34 作为主支路的骨干网络而不选择 ResNet18、ResNet50、ResNet101 或 ResNet152 的原因在于 ResNet34 在模型深度和计算效率间提供了一个优秀的折中方案, 它比 ResNet18 拥有更多的层, 这意味着它能够学习更加复杂的数据特征, 识别精度更高. 同时, 与更深的 ResNet50、ResNet101 和 ResNet152 相比, ResNet34 的参数更少, 运算需求也更低, 适合在资源受限的情况下使用, 确保既能达到较好性能, 也不会过度消耗计算资源. 实验结果见表 3, 以 ResNet18 为骨干网络尽管参数最少 (38.4M), 但其 *MIoU* 性能为 77.52%, 是所有选项中最低的, 表明它对于较复杂的分割任务可能不够有效. 相比之下, 以 ResNet34 为骨干网络以稍高的参数量 (45.1M) 实现了更高的 *MIoU* 值 (79.19%), 能够实现更好的性能和计算效率的平衡. 以 ResNet50 为骨干网络虽然性能略优 (79.49%), 但参数量增加至 66.6M, 可能会导致更高的计算成本. 更深层的 ResNet101 和 ResNet152 虽然理论上能学习更复杂的特征, 但在实际性能上却未必提升, 如 ResNet101 的 *MIoU* 为 76.84%, ResNet152 的 *MIoU* 为 78.54%, 同时它们的高参数量 (分别为 82.2M 和 94.2M) 也显著增加了计算资源的需求. 因此, ResNet34 在性能、资源消耗及适应性之间提供了最合理的折中方案, 使其成为不同深度 ResNet 模型中相对较优的选择, 表 3 中用*标注了该最合理的折中方案.

表 3 针对模型中不同骨干网络的消融实验

方法	参数量 (M)	*MIoU*(%)
ResMobileNet (ResNet18)	38.4	77.52
* ResMobileNet (ResNet34)	45.1	79.19
ResMobileNet (ResNet50)	66.6	79.49
ResMobileNet (ResNet101)	82.2	76.84
ResMobileNet (ResNet152)	94.2	78.54

(2) 不同模块的消融实验

我们首先使用原始 ResNet34 作为基准模型, 并实施一个简单的上采样过程, 将最后一层的上采样结果与前一层的特征图进行拼接, 然后重复这一过程, 直到恢复原始图像的尺寸. 随后, 我们逐步将提出的模块添加到模型中, 验证每个模块和整个模型的可行性, 图 13 展现了通过逐步增加模块如何逐步提升分割效果的精细化程度. 实验结果如表 4 所示, 主要使用 *MIoU* 作为评价指标, 清晰地显示出包含所有模块的模型在性能上达到了最佳效果.

改进的 ResNet34 网络采用 MS-SCPM 模块来替代原最大池化层, 并且引入多尺度条形卷积操作, 增强了对不同尺度特征的提取能力, 以适应云和云影的复杂形态变化. 条状卷积的应用有效减少了不相关区域的干扰, 使 *MIoU* 提升至 74.76%. MobileNetV3 支路利用深度可分离卷积和 squeeze-and-excitation (SE) 模块,

减少了计算量并提高了特征提取能力, 使 *MIoU* 提高 1.08%. ADPMFEM 模块通过动态预测池化尺度, 灵活捕获多尺度信息, 比 ASPP 和 PPM 模块表现更佳, *MIoU* 提升 1.09%. CWA 通过内容感知上采样和内置注意力机制, 优化图像细节重建, 尤其在处理边界复杂的云和云影任务时表现突出, *MIoU* 达 79.05%. 最后, 变形卷积进一步修复边缘细节, 适应不规则云边界, 使 *MIoU* 提升至 79.19%.

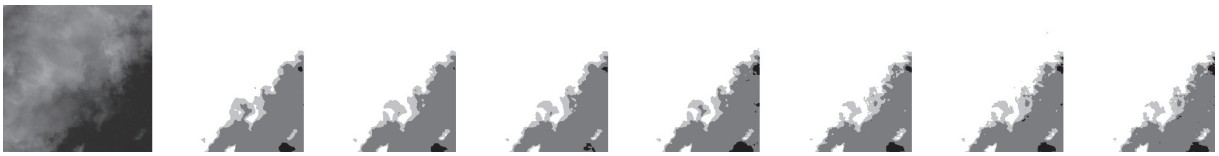


图 13 Biome 8 数据集上的消融实验分割结果对比

表 4 针对模型中不同模块的消融实验

模型	描述	*MIoU* (%)
模型1	R	71.45
模型2	R+MS-SCPM	**72.47 (1.02↑)**
模型3	R+MS-SCPM+SC	**74.76 (2.29↑)**
模型4	R+M+SC+MS-SCPM	**75.84 (1.08↑)**
模型5	R+M+SC+MS-SCPM+PPM	76.21
模型6	R+M+SC+MS-SCPM+ASPP	76.39
模型7	R+M+SC+MS-SCPM+ADPMFEM	**76.93 (1.09↑)**
模型8	R+M+SC+MS-SCPM+ADPMFEM+CWA	**79.05 (2.12↑)**
模型9	R+M+SC+MS-SCPM+ADPMFEM+CWA+DC	**79.19 (0.14↑)**

注: R: ResNet34; SC: 条形卷积; M: MobileNetV3; DC: 可变形卷积

2.3 与其他模型的对比实验结果分析

2.3.1 在 Biome 8 数据集上的对比结果分析

本节与现有优秀模型 (如 PSPNet、DeepLabv3+、BiSeNet V2、CGNet、HRNet、PVT、CvT、DBNet) 进行对比. PSPNet 通过池化层提取多尺度语义信息; DeepLabv3+使用 ASPP 模块进行特征融合; BiSeNet V2 优化了推理速度和精度; CGNet 通过上下文引导模块提升分割效率; HRNet 保持了高分辨率特征图; 基于 Transformer 的 PVT 和 CvT 结合卷积和 Transformer 的优势; DBNet 通过双支路模型提升精度.

在 Biome 8 数据集上设计了对比实验, 评估不同模型在云和云影分割任务中的表现, 为保证实验的客观性, 所有实验参数均设置为默认值. 实验结果显示, 本文方法在云检测、薄云检测、云影检测中均取得了最佳精度, 明显优于其他对比模型 (如 FCN、PSPNet、BiSeNet V2 等). 表 5 和表 6 记录了各模型在精度 (*P*)、召回率 (*R*)、*F*1 分数 (*BF*) 等指标上的表现. 我们所提出的模型在这些任务中表现突出, 尤其在云和云影的精确分割和边界恢复上具有明显优势.

表 5 不同模型在 Biome 8 数据集上的各类别评估指标比较 (%)

Method	Cloud			Thin cloud			Cloud shadow		
	P	*R*	*BF*	*P*	*R*	*BF*	*P*	*R*	*BF*
FCN[9]	77.64	80.97	78.82	51.66	50.61	50.42	70.73	44.04	53.21
PSPNet[12]	85.19	85.56	85.27	62.42	65.59	63.76	78.18	64.53	70.42
ABCNet[38]	85.85	86.57	86.03	62.40	65.66	63.67	78.93	67.72	72.50
BiSeNet V2[13]	89.58	85.64	87.45	63.57	71.70	67.10	83.80	68.63	75.08
CGNet[39]	88.42	89.57	88.91	69.20	69.51	69.20	78.86	75.24	76.78
LinkNet[40]	93.03	87.49	90.08	67.10	78.34	72.02	86.84	72.77	78.91
ExtremeC3Net[27]	93.17	88.65	90.74	71.57	72.69	71.84	79.68	77.10	78.11
CMT[41]	90.04	93.24	91.54	74.73	70.34	72.15	83.56	76.19	79.40
HRNet[42]	90.78	91.53	91.08	73.67	74.58	73.93	83.33	75.67	79.10
A2-FPN[43]	92.12	90.17	91.07	71.68	75.52	73.27	83.68	78.34	80.68
DeepLabv3+[11]	92.18	92.09	92.08	74.35	76.29	75.09	84.89	78.02	81.06
PVT[19]	93.48	91.39	92.36	74.75	78.70	76.46	85.55	78.40	81.60
CMTF[44]	91.48	93.15	92.21	77.19	77.65	77.13	86.71	80.28	83.16
ENet[45]	93.20	92.92	93.01	75.67	79.64	77.38	86.31	80.82	83.30
Swin Unet[46]	93.60	92.36	92.89	78.61	77.65	77.93	84.79	81.71	83.09
DBNet[24]	94.37	91.51	92.82	74.64	82.60	78.16	87.57	79.72	83.21
Ours	**95.69**	**94.56**	**95.08**	**81.27**	**84.47**	**82.65**	**89.07**	**83.35**	**85.96**

表 6 不同模型在 Biome 8 数据集上的总体评估指标比较 (%)

Method	*P*	*BF*	*R*	*MPA*	*FWIoU*	*MIoU*
FCN	68.60	67.20	67.73	62.47	52.09	47.84
PSPNet	77.26	76.83	76.86	73.78	63.30	60.29
ABCNet	77.77	77.22	77.19	74.65	63.83	61.13
BiSeNet V2	80.40	79.49	79.31	76.87	66.84	64.25
CGNet	81.26	81.09	81.15	79.05	69.01	66.43
LinkNet	83.60	82.69	82.46	80.53	71.24	68.94
ExtremeC3Net	83.25	82.81	82.76	81.01	71.50	68.95
CMT	83.88	83.58	83.72	81.45	72.61	70.07
HRNet	84.14	83.90	83.96	81.68	72.98	70.43
A2-FPN	84.27	83.85	83.77	82.22	72.91	70.71
DeepLabv3+	85.09	84.82	84.83	82.99	74.34	72.07
PVT	85.78	85.41	85.36	83.66	75.18	72.98
CMTF	86.29	85.98	85.99	84.35	76.03	74.07
ENet	86.54	86.24	86.20	84.76	76.42	74.42
Swin-Unet	86.74	86.54	86.57	84.96	76.85	74.75
DBNet	87.09	86.56	86.43	85.05	76.90	74.87
Ours	**89.46**	**89.20**	**89.18**	**87.84**	**80.99**	**79.19**

表 5 记录了对云、薄云和云影 3 个类别进行检测的 *P*、*R* 和 *BF* 数据, 可以看出, 在云检测、薄云检测和云影检测中, 本文模型的 *P*、*R* 和 *BF* 均取得了最佳效果. 表 6 展示了 Biome 8 数据集上不同模型在 *P*、*BF*、*R*、*MPA*、*FWIoU*、*MIoU* 指标上的对比结果. 从实验数据上看, FCN 的表现最差, 其次是 PSPNet、ABCNet、BiSeNet V2、CGNet、LinkNet 和 ExtremeC3Net. CMT、HRNet、A2-FPN、DeepLabv3+、PVT、CMTF、Swin-Unet、ENet 和 DBNet 的表现则更好, 但仍不及本文模型.

为更直观展示模型预测效果, 将几组预测对比图进行可视化, 如图 14. 这些对比图清晰展示了本文模型在细节捕捉和边界处理上的优势, 尤其是在云和云影分割任务中. 通过比较可以清楚地看到本文方法在处理复杂遥感图像方面的卓越性能.

图 14 精选了在 Biome 8 数据集上表现优异的 6 种方法进行预测效果可视化比较. 前 3 行分别呈现了荒地、森林和草地场景. 可以看出, 本文模型在恢复边界信息上具有明显优势, 且在处理小目标云和薄云时表现出色. 表 5 所示, 本文模型在云和薄云检测中分别达到 95.69% 和 81.27% 的精度. 第 4 行为灌木丛场景, 其中上框体现了模型对云和薄云混合场景的分割能力, 下框反映了模型对大范围不规则薄云的分割能力. 本文模型在云和薄云的 *BF* 指标上分别达到了 95.08% 和 82.65%, 在所有对比模型中均排名第 1. 第 5 行是雪场景, 由于雪与云、薄云的视觉效果极为相似, 因此这一场景极易发生漏检误检. 得益于多尺度条带卷积池化模块和注意力动态金字塔多尺度特征提取模块的贡献, 我们能够更好地利用上下文语义信息, 增强对全局信息的感知能力, 因此在雪场景中同样表现出色. 在第 6 行的城市场景测试了模型对大范围零碎目标的检测能力. 示例显示了模型在分割零碎云和阴影方面展现出卓越的性能. 表 5 中, 本文模型在云影检测各项指标上均表现最佳, 精度达到 89.07%, 召回率达到 83.35%, *BF* 达到 85.96%. 第 7 行为水场景, 通过 CWA 模块的上采样过程, 本文模型更好地恢复了细节信息, 精化了边界. 从示例图可以明显看出, 模型对细节信息的检测能力显著超越其他模型. 第 8 行是湿地场景, 其背景相对复杂, 但我们依然成功完成了对薄云、阴影和云的精确分割, 有效解决了云的阴影定位问题及边缘分割问题. 在该数据集中, 与其他模型相比, 本文模型在大规模云、薄云和阴影的定位能力方面, 以及在细节信息和边界信息的恢复能力方面均展现出显著优势.

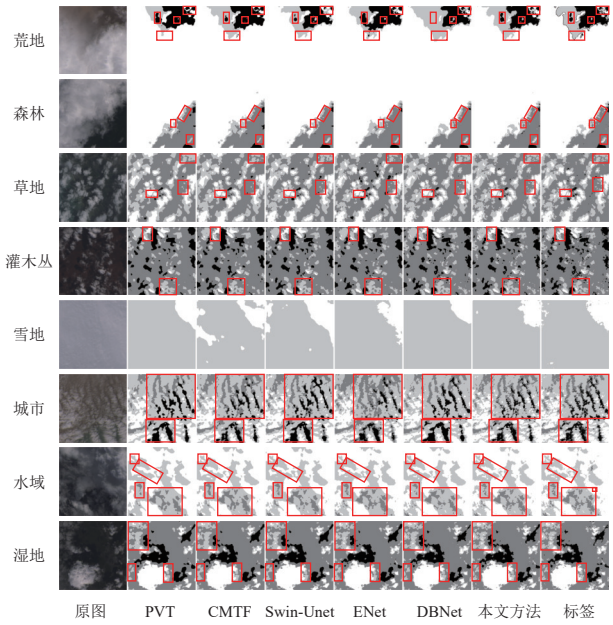


图 14 Biome 8 数据集中不同模型的分割比较

图 15 中, 我们对模型进行了全面的性能测试, 评估其准确性和抗干扰能力. 测试场景涵盖了易出现误检的环境及肉眼难以区分的云与云影. 雪地场景中, 尽管云与雪色相近, 我们仍成功避免了误检, 精确识别阴影区域, 而其他模型错误地将雪识别为云. 水域场景中, 海岛和薄云相似, 只有本文模型准确识别出薄云, 其他模型存在误判. 山地场景中, 深色山谷和光滑山体易被误判为阴影或云, 其他模型出现误检, 而本文模型几乎

无误. 沙漠场景中, 深色沙坑和沙滩易被识别为阴影或薄云, 但本文模型在准确识别小目标的同时避免了这些错误. 性能测试结果显示, 我们的模型在 *P* (89.46%)、*MPA* (87.84%)、*BF* (89.20%) 和 *FWIoU* (80.99%) 等指标上表现最佳, 展示出极强的抗干扰能力. 这得益于 CWA 上采样模块, 它通过感知内容并引导上采样过程, 有效提升了分割性能. 整体而言, 我们的模型在各种复杂场景下表现出卓越的云与云影分割能力, 验证了方法的有效性与先进性.

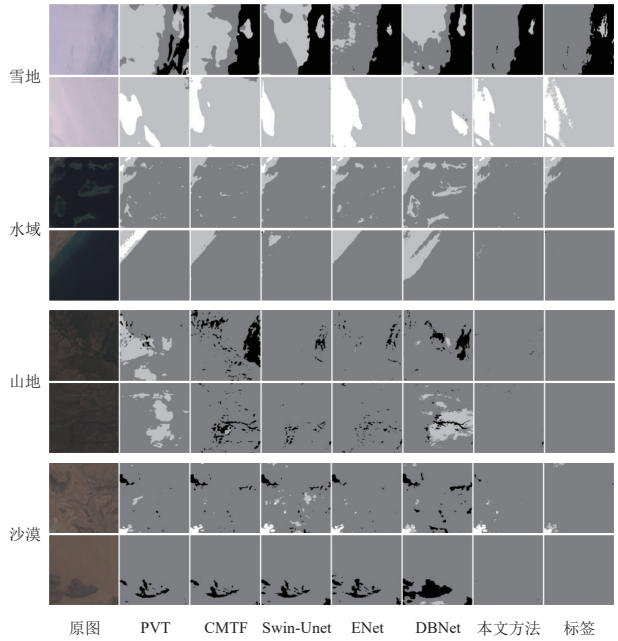


图 15 Biome 8 数据集中不同模型抗干扰性分割比较

2.3.2 在 HRC-WHU 数据集上的对比结果分析

为了验证我们模型的有效性, 我们对 HRC-WHU 数据集进行了概化对比测试, 所有模型的参数均设置为默认值以保证公平性. 实验结果如表 7 所示. HRC-WHU 数据集为二分类任务, 大多数模型表现良好, 特别是基于 Transformer 的模型及其与 CNN 结合的变体在该数据集上表现出色. 然而, 在所有对比网络中, 我们的模型在 *P*、*MPA*、*R*、*BF* 和 *MIoU* 等指标上均取得了最高分, 充分证明了其优异的泛化能力.

这些结果不仅表明本文模型在云和云影分割任务中的优势, 还展示了其在不同数据集和复杂图像上的泛化能力. 图 16 展示了本文模型的预测结果, 并与 5 个在 HRC-WHU 数据集上表现最好的模型进行比较.

实验选取了包括沙漠、林地、城市、雪景等多样化场景的图像进行评估. 在沙漠场景中, 薄云的透明度高、背景对比度低, 分割较为困难, 但本文模型在分割薄云方面表现优越. 表 7 显示, 本文模型在召回率上达到了 95.01%. 在雪地场景中, 雪与云的颜色接近, 增加了分割难度. 我们的模型在云与雪的边界分割及小目标云的检测上表现更好, *MPA* 最高达到 94.87%. 在城市场景中, 处理大规模云团和小目标云团时表现出色, 尤其在减少漏检和误检方面优势明显. 在林地场景中, 能够准确识别零碎的目标云和云影, 减少漏检和误检. 在水域场景中, 进一步验证了本文模型对小目标云和薄云的检测能力. 综上所述, 我们的模型在小目标识别、薄云检测和边界细化等方面优于其他模型, 证明了其在复杂遥感图像中云和云影分割任务中的优越性和可靠性.

表 7 不同模型在 HRC-WHU 数据集上的总体评估指标比较 (%)

Method	*P*	*MPA*	*R*	*BF*	*MIoU*
FCN	89.33	88.97	89.14	89.16	79.97
PVT	91.60	91.32	91.48	91.49	83.90
CMT	91.97	91.71	91.86	91.87	84.55
BiSeNet V2	92.25	92.15	91.98	92.01	84.84
ExtremeC3Net	92.54	92.21	92.46	92.46	85.58
CGNet	92.72	92.41	92.65	92.65	85.91
ABCNet	93.45	93.33	93.36	93.37	87.22
LinkNet	93.68	93.47	93.61	93.62	87.65
HRNet	94.28	94.07	94.24	94.01	88.77
A2-FPN	94.41	94.24	94.36	94.24	89.02
PSPNet	94.58	94.45	94.53	94.40	89.33
CMTF	94.68	94.57	94.62	94.10	89.51
DBNet	94.71	94.57	94.66	94.28	89.58
DeepLabv3+	94.74	94.60	94.69	94.09	89.63
ENet	94.82	94.65	94.79	94.49	89.79
Ours	**95.05**	**94.87**	**95.01**	**94.89**	**90.41**

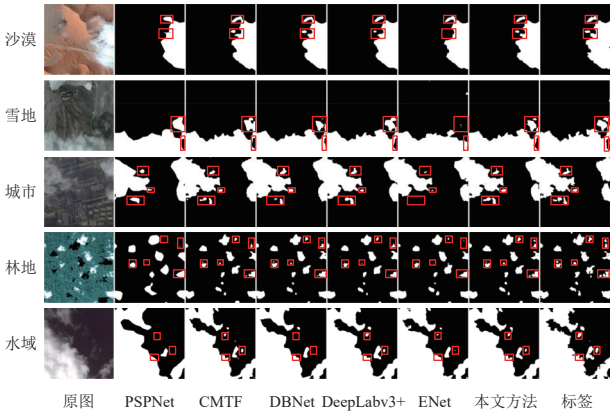


图 16 HRC-WHU 数据集中不同模型的分割比较

2.3.3 在 SPARCE 数据集上的对比结果分析

为了验证本文方法在多光谱遥感图像分割任务中

的性能, 我们在 SPARCS 数据集上进行对比实验, 结果汇总在表 8. 实验表明, 本文模型在 *MIoU*、*P*、*MPA*、*R*、*BF* 和 *FWIoU* 等多个关键指标上达到了最高值, 分别为 77.89%、90.22%、89.94%、90.10%、89.34% 和 81.57%. 这些实验结果证明了本文模型中的多尺度条带卷积池化模块 (MS-SCPM)、注意力动态金字塔多尺度特征提取模块 (ADPMFEM) 以及注意力特征感知重组模块 (CWA) 在特征提取和融合方面的出色表现.

表 8　不同模型在 SPARCE 数据集上的各类别评估指标比较 (%)

	Class Precision					Overall Result					
Method	Cloud	Shadow	Snow/Ice	Water	Land	*P*	*MPA*	*R*	*BF*	*FWIoU*	*MIoU*
FCN	81.14	71.95	80.76	67.82	82.76	79.88	72.74	79.67	79.31	66.85	61.08
ABCNet	88.60	73.55	86.42	86.85	87.04	84.27	76.44	84.12	83.99	73.20	68.01
CMT	87.31	81.03	79.30	81.39	84.69	84.78	78.16	84.74	84.33	73.71	68.57
PSPNet	85.92	78.22	84.36	79.05	88.02	85.12	79.42	85.24	85.00	74.70	70.00
BiSeNet V2	87.30	78.50	86.91	82.34	88.35	85.92	79.32	85.91	85.73	75.73	70.89
ExtremeC3Net	85.05	79.35	82.23	80.14	90.11	86.20	80.41	86.23	86.09	76.24	70.97
CGNet	86.12	78.26	84.32	79.70	91.30	86.65	80.38	86.62	86.54	76.92	71.29
LinkNet	89.05	76.19	82.04	82.45	89.12	85.93	80.34	85.79	85.73	75.70	71.36
PVT	87.30	81.10	83.09	83.06	89.13	86.78	81.77	86.80	86.62	77.01	72.65
A2-FPN	88.03	79.13	87.29	82.20	90.51	87.09	81.79	87.06	86.98	77.57	73.15
HRNet	88.34	80.62	86.08	86.80	90.64	87.66	81.32	87.68	87.55	78.46	73.51
DeepLabv3+	89.13	80.05	86.20	85.24	91.04	87.90	82.20	87.85	87.78	78.80	74.31
DBNet	87.94	82.26	85.98	85.19	91.17	88.28	83.39	88.32	88.18	79.44	75.21
ENet	88.99	82.65	85.63	86.15	92.50	88.44	83.64	88.45	88.36	79.69	75.31
CMTF	89.52	82.29	87.95	**86.91**	**92.55**	89.31	83.70	89.28	89.22	81.04	76.41
Ours	**89.73**	**83.29**	**88.00**	86.80	92.30	**90.22**	**83.94**	**90.10**	**89.34**	**81.57**	**77.89**

图 17 展示了不同场景下各模型的分割效果. 在薄云分割中, HRNet 和 DeepLabv3+存在误检和漏检, 而我们的模型在细节和边界分割上表现更佳. 对于小目标云和云影, 模型的准确率指标分别达到了 89.73% 和 83.29%, 表现优越. 在水域分割中, 我们的模型避免了误检深黑色水域为阴影的问题, 准确率为 86.80%. 在雪区背景分割中, 由于雪与云色相近, 我们的模型准确率高达 88.00%, 领先其他模型. 总体而言, 我们的模型在复杂场景下, 特别是在水域和雪区等容易误检的区域, 展现了强大的分割能力.

3　结论

本文提出了一种基于 ResNet34 和 MobileNetV3 的双支路网络, 用于实现可见光和多光谱高分辨率遥感图像的云和云影端到端分割. 该方法通过模块化架构, 分别利用 MobileNetV3 和 ResNet34 进行初步特征提取和深层特征提取, 以有效应对遥感图像中的多尺度、多形状特征. 为了更好地捕捉不同尺度的上下文信息并避免信息丢失, 我们在 ResNet34 中采用了 MS-SCPM 替代传统的最大池化层, 并通过条形卷积替换普通卷积, 进一步提升了特征提取精度. 在深层特征提取阶段, 引入了 ADPMFEM 模块, 以灵活捕获多尺度信息, 减少信息丢失的可能性. 在解码阶段, CWA 模块生成权重指导上采样过程, 既保留了精确的分类信息, 又修复了粗糙的分割边界. 最后, 在像素分类之前通过可变形卷积实现了对云和云影复杂形状的精准分割.

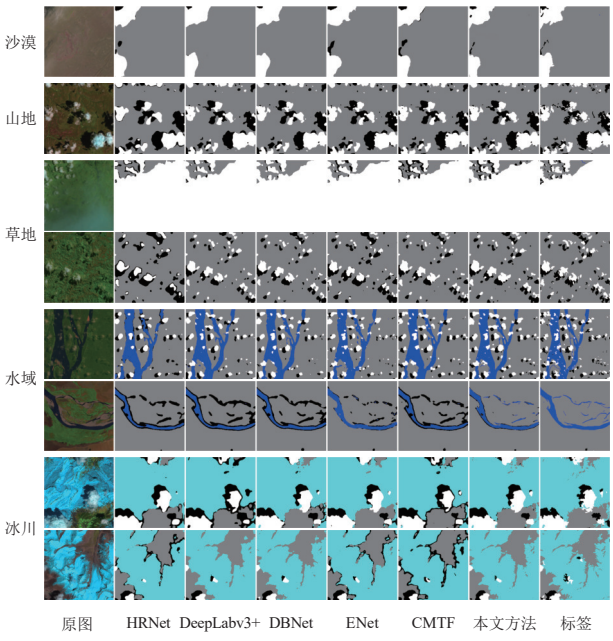


图 17　SPARCS 数据集中不同模型的分割比较

本文提出的方法在遥感图像分割中展现了显著的特异性, 主要体现在其对遥感图像中特有的多尺度、

多形状特征的处理能力, 以及在高分辨率下保持分割细节和准确性的优势. 特别地, 该方法能够有效应对遥感图像中的挑战, 如复杂地物的高分辨率分割和多尺度信息的充分利用. 同时, 得益于其模块化和灵活架构, 该方法具备了广泛的跨领域应用潜力, 能够高效处理多种图像分割任务. 实验结果验证了该方法的有效性, 与现有技术相比, 模型在精度和处理复杂场景的能力上均有显著提升. 尽管当前模型已取得优异性能, 但仍存在进一步优化的空间, 未来的工作将聚焦于减少模型参数数量, 在不牺牲分割精度的前提下提高推理速度, 从而使模型在实际应用中更加高效和实用.

参考文献

1 Ceppi P, Nowack P. Observational evidence that cloud feedback amplifies global warming. Proceedings of the National Academy of Sciences of the United States of America, 2021, 118(30): e2026290118.

2 Chen K, Dai X, Xia M, *et al*. MSFANet: Multi-scale strip feature attention network for cloud and cloud shadow segmentation. Remote Sensing, 2023, 15(19): 4853. [doi: 10.3390/rs15194853]

3 Zhang YC, Rossow WB, Lacis AA, *et al*. Calculation of radiative fluxes from the surface to top of atmosphere based on ISCCP and other global data sets: Refinements of the radiative transfer model and the input data. Journal of Geophysical Research: Atmospheres, 2004, 109(D19): D19105.

4 Li QY, Lu WT, Yang J. A hybrid thresholding algorithm for cloud detection on ground-based color images. Journal of Atmospheric and Oceanic Technology, 2011, 28(10): 1286–1296. [doi: 10.1175/JTECH-D-11-00009.1]

5 刘希, 许健民, 杜秉玉. 用双通道动态阈值对 GMS-5 图像进行自动云检测. 应用气象学报, 2005, 16(4): 434–444. [doi: 10.3969/j.issn.1001-7313.2005.04.003]

6 Tapakis R, Charalambides AG. Equipment and methodologies for cloud detection and classification: A review. Solar Energy, 2013, 95: 392–430. [doi: 10.1016/j.solener.2012.11.015]

7 Cheng GL, Wang Y, Xu SB, *et al*. Automatic road detection and centerline extraction via cascaded end-to-end convolutional neural network. IEEE Transactions on Geoscience and Remote Sensing, 2017, 55(6): 3322–3337. [doi: 10.1109/TGRS.2017.2669341]

8 Song L, Xia M, Jin JL, *et al*. SUACDNet: Attentional change detection network based on siamese U-shaped structure. International Journal of Applied Earth Observation and Geoinformation, 2021, 105: 102597. [doi: 10.1016/j.jag.2021.102597]

9 Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE, 2015. 3431–3440.

10 Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. Proceedings of the 18th International Conference on Medical Image Computing and Computer-assisted Intervention. Munich: Springer, 2015. 234–241.

11 Chen LC, Zhu YK, Papandreou G, *et al*. Encoder-decoder with atrous separable convolution for semantic image segmentation. Proceedings of the 15th European Conference on Computer Vision (ECCV). Munich: Springer, 2018. 833–851.

12 Zhao HS, Shi JP, Qi XJ, *et al*. Pyramid scene parsing network. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017. 6230–6239.

13 Yu CQ, Wang JB, Peng C, *et al*. BiSeNet: Bilateral segmentation network for real-time semantic segmentation. Proceedings of the 15th European Conference on Computer Vision (ECCV). Munich: Springer, 2018. 334–349.

14 刘云, 陆承泽, 李仕杰, 等. 基于高效的多尺度特征提取的轻量级语义分割. 计算机学报, 2022, 45(7): 1517–1528. [doi: 10.11897/SP.J.1016.2022.01517]

15 龙丽红, 朱宇霆, 闫敬文, 等. 新型语义分割 D-UNet 的建筑物提取. 遥感学报, 2023, 27(11): 2593–2602.

16 Carion N, Massa F, Synnaeve G, *et al*. End-to-end object detection with Transformers. Proceedings of the 16th European Conference on Computer Vision. Glasgow: Springer, 2020. 213–229.

17 Engel N, Belagiannis V, Dietmayer K. Point Transformer. IEEE Access, 2021, 9: 134826–134840. [doi: 10.1109/ACCESS.2021.3116304]

18 Chen HT, Wang YH, Guo TY, *et al*. Pre-trained image processing Transformer. Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021. 12294–12305.

19 Wang WH, Xie EZ, Li X, *et al*. Pyramid vision Transformer: A versatile backbone for dense prediction without convolutions. Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision. Montreal: IEEE, 2021. 548–558.

20 Liu Z, Lin YT, Cao Y, *et al*. Swin Transformer: Hierarchical

vision Transformer using shifted windows. Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision. Montreal: IEEE, 2021. 9992–10002.

21 Wu HP, Xiao B, Codella N, *et al*. CvT: Introducing convolutions to vision Transformers. Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision. Montreal: IEEE, 2021. 22–31.

22 Hu K, Zhang EW, Xia M, *et al*. MCANet: A multi-branch network for cloud/snow segmentation in high-resolution remote sensing images. Remote Sensing, 2023, 15(4): 1055. [doi: 10.3390/rs15041055]

23 Gu GW, Weng LG, Xia M, *et al*. Multipath multiscale attention network for cloud and cloud shadow segmentation. IEEE Transactions on Geoscience and Remote Sensing, 2024, 62: 5404215.

24 Lu C, Xia M, Qian M, *et al*. Dual-branch network for cloud and cloud shadow segmentation. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 5410012.

25 李远禄, 王键翔, 范小婷, 等. 基于 ViT-D-UNet 的双分支遥感云影检测网络. 计算机系统应用, 2024, 33(8): 68–77. [doi: 10.15888/j.cnki.csa.009596]

26 杨军, 张金影, 康玥. 基于自注意力机制的高分遥感影像语义分割. 哈尔滨工程大学学报, 2025, 46(2): 344–354. [doi: 10.11990/jheu.202211028]

27 Park H, Sjösund LL, Yoo Y, *et al*. ExtremeC3Net: Extreme lightweight portrait segmentation networks using advanced C3-modules. arXiv:1908.03093. 2019.

28 Fu J, Liu J, Tian HJ, *et al*. Dual attention network for scene segmentation. Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019. 3141–3149.

29 He KM, Zhang XY, Ren SQ, *et al*. Deep residual learning for image recognition. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 770–778.

30 Howard A, Sandler M, Chen B, *et al*. Searching for MobileNetV3. Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019. 1314–1324.

31 Guo MH, Lu CZ, Hou QB, *et al*. SegNeXt: Rethinking convolutional attention design for semantic segmentation. Proceedings of the 36th International Conference on Neural Information Processing Systems. New Orleans: Curran Associates Inc., 2022. 84.

32 Woo S, Park J, Lee JY, *et al*. CBAM: Convolutional block attention module. Proceedings of the 15th European Conference on Computer Vision (ECCV). Munich: Springer, 2018. 3–19.

33 Wang JQ, Chen K, Xu R, *et al*. CARAFE: Content-aware reassembly of features. Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019. 3007–3016.

34 Dai JF, Qi HZ, Xiong YW, *et al*. Deformable convolutional networks. Proceedings of the 2017 IEEE International Conference on Computer Vision. Venice: IEEE, 2017. 764–773.

35 Vermote E, Justice C, Claverie M, *et al*. Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. Remote Sensing of Environment, 2016, 185: 46–56. [doi: 10.1016/j.rse.2016.04.008]

36 Li ZW, Shen HF, Cheng Q, *et al*. Deep learning based cloud detection for medium and high resolution remote sensing images of different sensors. ISPRS Journal of Photogrammetry and Remote Sensing, 2019, 150: 197–212. [doi: 10.1016/j.isprsjprs.2019.02.017]

37 Hughes MJ, Hayes DJ. Automated detection of cloud and cloud shadow in single-date Landsat imagery using neural networks and spatial post-processing. Remote Sensing, 2014, 6(6): 4907–4926. [doi: 10.3390/rs6064907]

38 Li R, Zheng SY, Zhang C, *et al*. ABCNet: Attentive bilateral contextual network for efficient semantic segmentation of fine-resolution remotely sensed imagery. ISPRS Journal of Photogrammetry and Remote Sensing, 2021, 181: 84–98. [doi: 10.1016/j.isprsjprs.2021.09.005]

39 Wu TY, Tang S, Zhang R, *et al*. CGNet: A light-weight context guided network for semantic segmentation. IEEE Transactions on Image Processing, 2021, 30: 1169–1179. [doi: 10.1109/TIP.2020.3042065]

40 Chaurasia A, Culurciello E. LinkNet: Exploiting encoder representations for efficient semantic segmentation. Proceedings of the 2017 IEEE Visual Communications and Image Processing (VCIP). St. Petersburg: IEEE, 2017. 1–4.

41 Guo JY, Han K, Wu H, *et al*. CMT: Convolutional neural networks meet vision Transformers. Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022. 12165–12175.

42 Yu CQ, Xiao B, Gao CX, *et al*. Lite-HRNet: A lightweight high-resolution network. Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021. 10435–10445.

43 Li R, Wang LB, Zhang C, *et al*. A^2-FPN for semantic segmentation of fine-resolution remotely sensed images.

International Journal of Remote Sensing, 2022, 43(3): 1131–1155. [doi: 10.1080/01431161.2022.2030071]

44 Wu HL, Huang P, Zhang M, *et al*. CMTFNet: CNN and multiscale Transformer fusion network for remote-sensing image semantic segmentation. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 2004612.

45 Paszke A, Chaurasia A, Kim S, *et al*. ENet: A deep neural network architecture for real-time semantic segmentation. arXiv:1606.02147, 2016.

46 Cao H, Wang YY, Chen J, *et al*. Swin-Unet: Unet-like pure Transformer for medical image segmentation. Proceedings of the 2022 European Conference on Computer Vision. Tel Aviv: Springer, 2022. 205–218.

(校对责编: 王欣欣)