

基于双模态高效特征学习的高分辨率遥感图像分割

张银胜^{1,2}, 吉茹², 童俊毅², 杨宇龙², 胡宇翔¹, 单慧琳^{1,2}

1. 无锡学院 电子信息工程学院, 无锡 214105;

2. 南京信息工程大学 电子与信息工程学院, 南京 210044

摘要: 遥感图像因具有丰富的语义信息和空间信息, 增加了语义分割的难度。然而已有提取双模态特征的分割方法采用相同的主干网络, 没有考虑互补特征的差异, 存在特征提取、特征融合和上采样恢复细节信息不足等问题, 无法准确高效的学习高分辨率遥感图像信息。因此, 本文提出基于双模态高效特征学习的高分辨率遥感图像分割算法。首先, 针对不同模态的遥感图像设计合适的编码器, 高效的提取双模态特征, 并通过交互加强模块减少不同路径特征之间的差异。其次, 提出双模态特征聚合模块和深层特征提取模块进一步融合和提取双模态特征, 使网络能够充分学习互补信息。最后, 提出多层特征上采样模块, 利用语义信息丰富的高层特征对细节信息丰富的低层特征进行加权操作, 逐步上采样实现特征高效恢复, 提升分割性能。实验结果表明, 所提算法在 ISPRS Potsdam 和 Vaihingen 数据集上的总体精度分别达到了 94.52%、90.45%, 能够高效的提取并融合高分辨率遥感图像的双模态特征, 提高遥感图像分割的准确率。

关键词: 遥感图像分割, 高效特征提取, 交互融合, 双模态特征聚合, 深层特征提取, 多层特征上采样

中图分类号: TP391/P2

引用格式: 张银胜, 吉茹, 童俊毅, 杨宇龙, 胡宇翔, 单慧琳. 2024. 基于双模态高效特征学习的高分辨率遥感图像分割. 遥感学报, 28(2): 481-493

Zhang Y S, Ji R, Tong J Y, Yang Y L, Hu Y X and Shan H L. 2024. High resolution remote sensing image segmentation based on dual-modal efficient feature learning. National Remote Sensing Bulletin, 28(2): 481-493 [DOI:10.11834/jrs.20233162]

1 引言

遥感图像和图像处理技术广泛应用于城市规划 (Zhou 等, 2011), 变化检测 (Gerard 等, 2010), 植被检测 (Ahmed 等, 2017), 军事侦察 (王树连, 2005) 等领域。随着卷积神经网络 CNN (Convolutional Neural Networks) 的发展, 遥感图像分割研究 (Farabet 等, 2013; Mnih 和 Hinton, 2010; Paisitkriangkrai 等, 2015) 进展迅速。但是获取的遥感图像分辨率逐渐提高, 导致了不同类别之间差异变小, 相同类别之间差异变大的问题, 即异物同谱和同物异谱现象严重。

目前高分辨率遥感图像分割存在普通遥感图像分割和双模态遥感图像分割等方法。在普通遥感图像分割研究领域, Xu 等 (2018) 将 U-Net

(Ronneberger 等, 2015) 中特征提取的主干网络替换为 ResNet (Residual Networks) (He 等, 2016), 并使用导向性滤波器来加强分割效果。Wu 等 (2018) 使用了多尺度约束的损失函数, 加强了网络上采样阶段对遥感图像空间结构的恢复。张小娟和汪西莉 (2020) 设计了一个采用完全残差连接和多尺度特征融合的端到端遥感图像分割模型, 能够提取丰富的上下文信息, 同时使用叠加的 ASPP (Atrous Spatial Pyramid Pooling) (Chen 等, 2018) 结构应对目标尺度变化, 提升分割性能。Sun 和 Li (2021) 提出多尺度特征融合网络 MFFNet, 使用三支结构提取图像特征, 并通过注意力机制进行优化, 取得了较高的分割精度。Chen 等 (2021) 提出多层次特征聚合网络 MFANet, 在深度特征提取和上采样特征融合两方面进行改进, 对

收稿日期: 2023-05-18; 预印本: 2023-10-08

基金项目: 国家自然科学基金 (编号: 62071240, 62106111); 江苏省研究生科研与实践创新计划基金 (编号: SJCX23_0379)

第一作者简介: 张银胜, 研究方向为遥感图像处理、深度学习。E-mail: yorkzhang@nuist.edu.cn

通信作者简介: 吉茹, 研究方向为遥感图像分割、深度学习。E-mail: 20211249643@nuist.edu.cn

高分辨率特征图进行细化和恢复。

在双模态高分辨率遥感图像分割研究领域, 遥感图像的双模态包括近红外、红绿 IRRG (Infrared、Red、Green) 图像和数字表面模型 DSM (Digital Surface Model) 图像, 其中 IRRG 图像具有纹理、颜色、形状等语义信息, 语义信息一般指的是目标区域像素和周围区域像素之间的关联性。DSM 图像具有高度等空间信息, 空间信息是指图像中分割出来的多个目标之间的相互的空间位置或相对方向关系。Marcos 等 (2018) 将 IRRG 图像和 DSM 图像简单结合作为输入, 产生冗余信息。Marmanis 等 (2016) 设计双分支网络提取不同模态图像的特征, 并在中间层进行特征融合, 造成参数冗余。Audebert 等 (2018) 采用两个 CNN 网络对 IRRG 图像和 DSM 图像进行处理, 并通过元素相加的方式进行模态特征融合。为了进一步加强两种模态之间的融合和上采样恢复, Yang 等 (2021) 提出注意力融合网络 AFNet, 利用 FCN (Fully Convolutional Networks) (Shelhamer 等, 2017) 提取多路径特征, 并利用通道注意力结构和空间注意力结构相互约束和引导, 融合多径特征和多层次特征以逐步实现上采样。Chen 等 (2021) 提出一种高度感知多路径并行网络 HA-MPPNet, 在 DSM 图像的监督下学习高度特征来改进上下文语义信息。孙汉淇等 (2022) 提出编码器—解码器结构的网络 MMFNet 融合双模态特征, 并使用密集连接的方式加强特征的传播和复用。张文凯等 (2022) 提出 MSFAFNet, 利用高程信息强化光谱特征, 结合非局部块 (Wang 等, 2018) 和压缩激励块 (Hu 等, 2020) 的优点实现遥感图像分割。

由于单模态高分辨率遥感图像仅利用可见光图像进行分割, 无法充分获取高分辨率遥感图像特征。因此, 为了高效学习高分辨率遥感图像特征, 本文采用双模态结构。但是现有双模态分割方法没有针对不同模态的图像设计适合的特征提取网络, 且存在融合特征不充分, 上采样恢复细节信息不足等问题, 本文提出了一个基于双模态高效特征学习 DEFL (Dual-modal Efficient Feature Learning) 的高分辨遥感图像分割算法。主要包括:

(1) 双模态高效特征提取。在编码模块 CM (Code Module) 中提出了双路径结构的 IRRG 编码器对 IRRG 图像的语义信息和细节信息进行特征提取; 提出了 DSM 编码器对 DSM 图像的空间信息进

行特征提取; 并利用交互加强模块 IEM (Interactive Enhancement Module) 融合多路径特征, 减少双模态特征之间的差距。

(2) 双模态高效特征融合。在融合模块中, 提出了一个双模态特征聚合模块 DFAB (Dual-modal Feature Aggregation Block), 利用 DSM 特征动态的加强 IRRG 特征, 实现两种模态的信息融合, 有效避免噪声对分割的影响; 还提出了一个深层特征提取模块 DFEM (Deep Feature Extraction Module), 对聚合后的遥感图像特征进行深层全局特征提取, 获得更多语义信息, 提高图像分割准确率。

(3) 双模态高效上采样特征恢复。在解码模块中提出了一个多层特征上采样模块 MFUM (Multilayer Feature Upsampling Module), 用高层特征对低级特征进行校准, 实现逐步上采样特征融合, 恢复更多的细节信息, 实现遥感图像的精准分割。

2 DEFLNet 网络结构

本文提出的 DEFLNet 的整体框架如图 1 所示, 主要由编码模块、融合模块和解码模块 3 部分组成。其中编码模块包括 IRRG 编码器、DSM 编码器和交互加强模块, 融合模块包括双模态特征聚合模块和深层特征提取模块, 解码模块为多层特征上采样模块。

网络将遥感图像的 IRRG 图像和 DSM 图像作为双输入, 编码模块包括 IRRG 编码器和 DSM 编码器。其中 IRRG 编码器是双分支结构, 分别提取 IRRG 图像的语义信息和细节信息。同时该模块采用多尺度融合策略, 将 IRRG 编码器中提取的语义信息和细节信息、IRRG 编码器提取的 IRRG 特征和 DSM 编码器提取的 DSM 特征在特征提取的不同阶段进行多次交互融合。融合模块包括双模态特征聚合模块和深层特征提取模块。其中双模态特征聚合模块利用 DSM 特征动态的加强 IRRG 特征, 有效的融合两种模态信息。接着深层特征提取模块从全局上提取融合后的遥感特征图的深层信息, 获取更多语义信息。最后解码模块中的多层特征上采样模块利用高层信息对低层信息进行加权操作, 促进高分辨率图像的细节信息恢复, 通过高效特征学习获得最终精准的分割图。

2.1 编码模块

已有双模态分割网络采用相同的主干网络对

IRRG 图像和 DSM 图像进行特征提取, 无法准确高效的提取双模态特征。编码模块如图 2 所示, 其内部各模块的详细结构如图 3 所示。编码模块由 IRRG 编码器和 DSM 编码器组成, 分别对 IRRG 图像丰富的内容信息和 DSM 图像丰富的空间信息进行特征提取。其中 IRRG 图像的低层特征便于提取位置、边缘等细节信息, 但是语义信息较少; 高层特征便于提取丰富的语义信息, 但是细节信息

比较粗略。因此, 本文采用双分支结构, 分别提取 IRRG 图像的语义信息和细节信息, 并在内部通过 IEM 将两种信息进行融合。分别获取 IRRG 图像的语义信息、细节信息和 DSM 图像的空间信息的策略, 可以加强网络学习图像不同特征的能力, 进而不断建立不同特征之间的联系, 减少双模态特征差异, 达到高效融合。

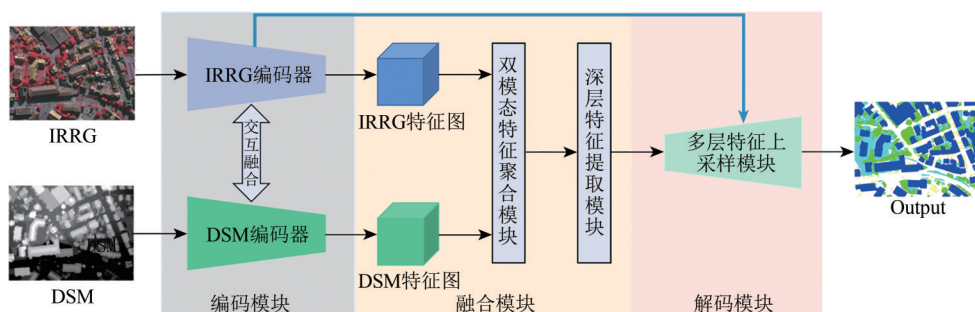


图1 DEFLNet的整体框架

Fig. 1 The overall framework of DEFLNet

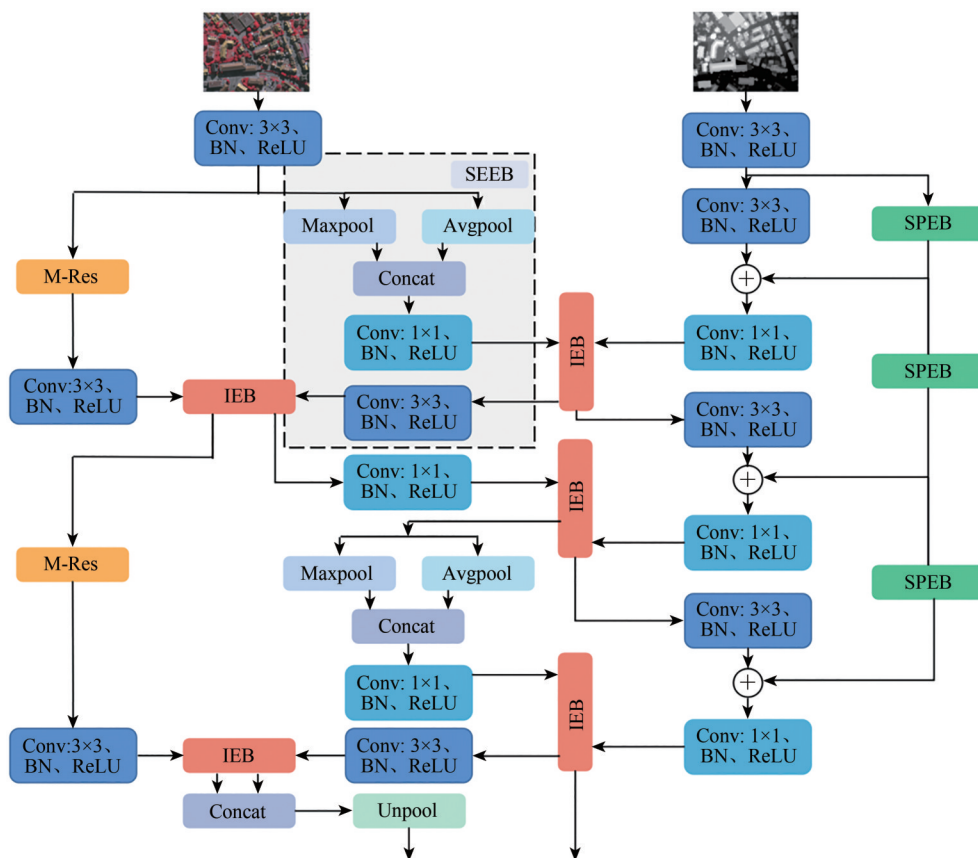


图2 编码模块结构图

Fig. 2 Code module structure diagram

IRRG 编码器的目的是提取 IRRG 图像丰富的内容信息, 如外观、颜色、纹理等。采用双分支

结构同时获取语义信息和细节信息, 可以有效增强信息提取能力。左侧的细节分支由卷积块和多

层残差块 M-Res (Multi-Residual Block) 组成。其中 M-Res 将输入特征图 X_{in} 经过最大池化层 P_m 、带有 BN 和 Sigmoid 激活函数的 1×1 卷积 $C_{1\sigma}$ 获取权重。然后对经过带有 BN 和 ReLU 激活函数的 3×3 卷积

$C_{3\sigma}$ 之后的特征图进行加权。再经过带有 BN 和 ReLU 激活函数的 1×1 卷积 $C_{1\delta}$ 建立通道联系。最后与 3×3 卷积 $C_{3\sigma}$ 构成残差结构，在后续层中补充丢失的信息。

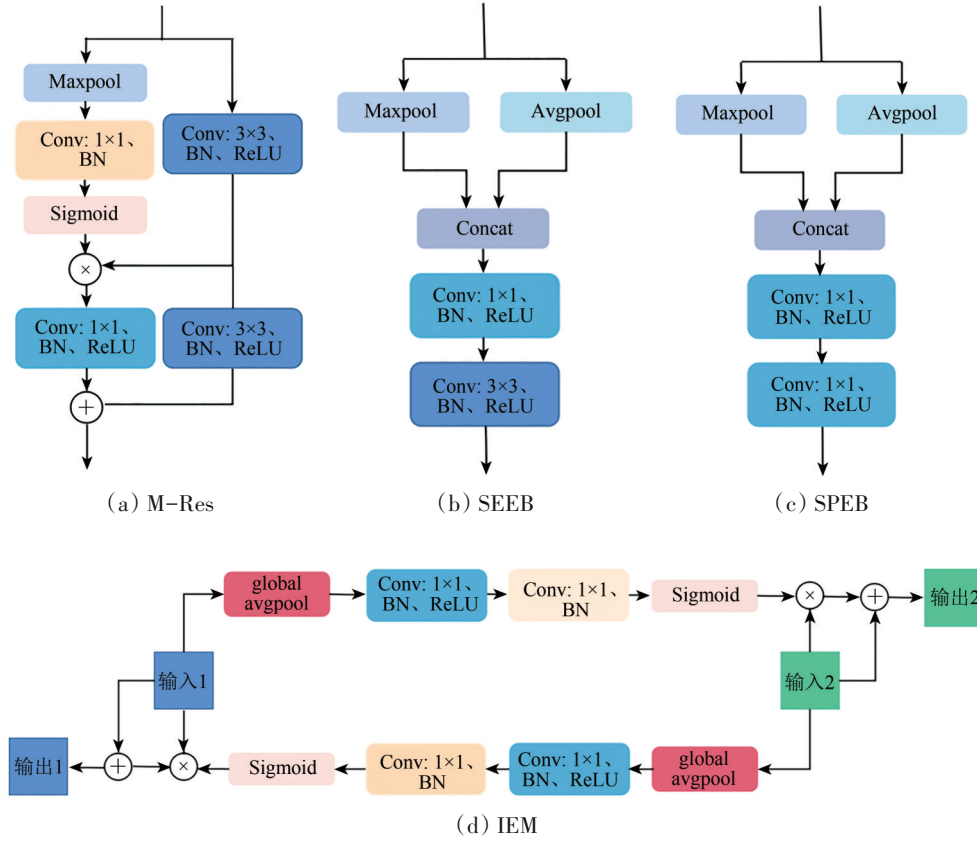


图3 各模块详细结构

Fig. 3 Detailed structure of each module

M-Res 的过程可表示为

$$M = C_{1\delta} \left(C_{1\sigma} \left(P_m(X_{in}) \right) \otimes C_{3\sigma}(X_{in}) \right) \oplus C_{3\sigma}(X_{in}) \quad (1)$$

右侧语义分支通过快速下采样增大感受野，获得 IRRG 图像丰富的语义信息。语义提取块 SEEB (Semantic Extraction Block) 采用 P_m 和平均池化层 P_a 同时对特征图进行降维，然后经过级联 *Concat* 和 $C_{1\delta}$ 建立跨通道信息联系，最后经过 $C_{3\sigma}$ 再次降维。该结构达到了一个快速下采样的效果，不断扩大感受野，同时通过卷积不断对语义信息进行提取。SEEB 的过程可表示为

$$SE = C_{3\sigma} \left(C_{1\delta} \left(\text{Concat} \left(P_m(X_{se}), P_a(X_{se}) \right) \right) \right) \quad (2)$$

式中， X_{se} 表示 SEEB 模块的输入特征。在 IRRG 编码器的双分支特征提取过程中加入 IEM，它是一个双输入双输出的结构，利用注意力机制的思想实现两种信息的特征交互学习。第一个特征图 X_{ie1} 通

过全局平均池化层 P_{ga} 保留显著特征，然后经过两个 1×1 卷积 $C_{1\delta}$ 、 $C_{1\sigma}$ 实现跨通道信息整合、获取各通道权重，对第二个特征图 X_{ie2} 进行加权、求和，得到第一个输出 X_{out2} ，同时第二个特征图也对第一个特征图进行相同的操作得到第二个输出 X_{out1} 。IEM 的过程可表示为

$$\begin{cases} X_{out2} = \left(C_{1\sigma} \left(C_{1\delta} \left(P_{ga}(X_{ie1}) \right) \right) \otimes X_{ie2} \right) \oplus X_{ie2} \\ X_{out1} = \left(C_{1\sigma} \left(C_{1\delta} \left(P_{ga}(X_{ie2}) \right) \right) \otimes X_{ie1} \right) \oplus X_{ie1} \end{cases} \quad (3)$$

DSM 编码器的目的是提取 DSM 图像的空间信息，由卷积 $C_{1\delta}$ 、 $C_{3\sigma}$ 和空间提取块 SPEB (Spatial Extraction Block) 构成的残差结构组成。SPEB 中的 P_m 和 P_a 并列提取特征并 *Concat*，以关注重点特征学习。然后经过两个 1×1 卷积 $C_{1\delta}$ 建立通道联系，并利用残差结构逐步获取更多空间信息。SPEB 的过程可表示为

$$SP = C_{16} \left(C_{16} \left(\text{Concat} \left(P_m(X_{sp}), P_a(X_{sp}) \right) \right) \right) \quad (4)$$

式中, X_{sp} 表示 SPEB 模块的输入特征。在 DSM 编码器进行空间信息提取的同时加入 IEM 与 IRRG 编码器提取的内容信息进行交互, 减小两种模态特征之间的差异, 避免双模态特征聚合时产生噪声对图像分割造成影响。

2.2 融合模块

2.2.1 双模态特征聚合模块

随着遥感图像的分辨率逐渐提高, IRRG 图像

所包含的内容信息愈加丰富, DSM 图像也包含丰富的空间信息, 但是两种模态特征之间存在差异。对此, 提出一个新的双模态特征聚合模块 DFAB 如图 4 所示, 进一步融合编码模块输出的 DSM 特征和 IRRG 特征, 提高遥感图像分割准确率。DFAB 利用两种模态特征之间的映射关系, 即利用 DSM 特征动态的对 IRRG 特征从全局上进行特征重校准。DFAB 有效的将 DSM 特征融入到 IRRG 通道中, 同时避免冗余特征和过多噪声对图像分割产生影响, 提高遥感图像分割效率。

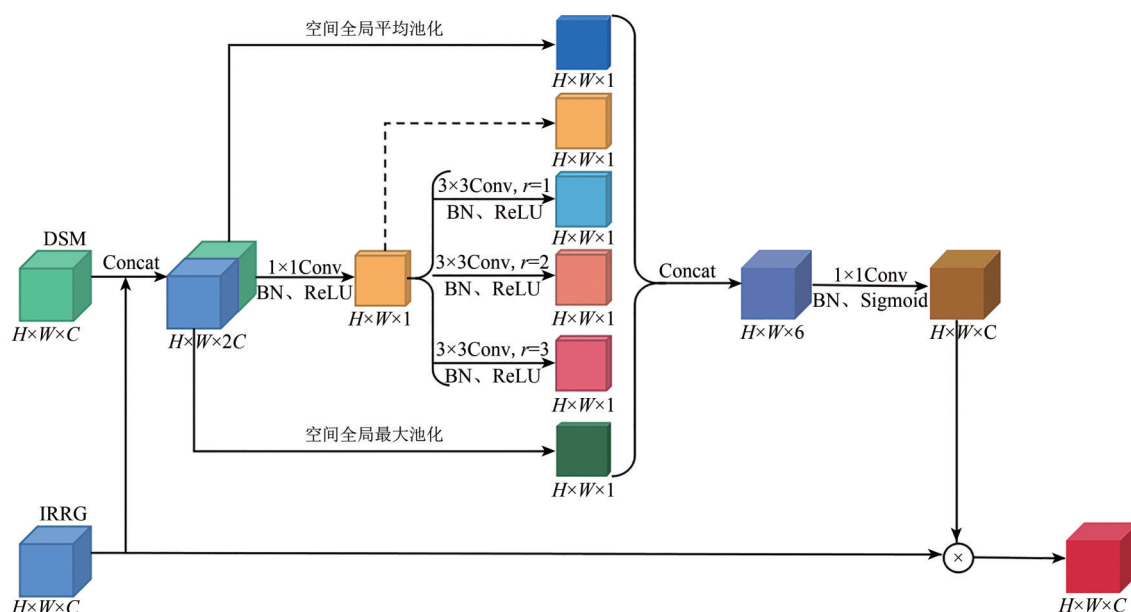


图4 双模态特征聚合模块结构图

Fig. 4 Dual-modal feature aggregation Block structure diagram

编码模块输出的 DSM 特征 X_a 和 IRRG 特征 X_b 经过 *Concat* 级联, 然后分别经过 1×1 卷积 C_{16} 、空间全局平均池化层 P_{sga} 和空间全局最大池化层 P_{sgm} 得到一维特征图 $T_1 = C_{16}(\text{Concat}(X^a, X^b))$ 、 $T_2 = C_{16}(P_{sga}(X^a, X^b))$ 和 $T_3 = C_{16}(P_{sgm}(X^a, X^b))$ 。接着将 T_1 经过空洞率 r 为 1、2、3 的并行空洞卷积 $C_{\delta}^{r=i}$, $i = 1, 2, 3$, 得到具有不同感受野的一维特征图 $T_4 = C_{\delta}^{r=1}(T_1)$ 、 $T_5 = C_{\delta}^{r=2}(T_1)$ 和 $T_6 = C_{\delta}^{r=3}(T_1)$ 。然后对所有的一维特征图进行 *Concat* 级联, 经过 C_{16} 建立通道之间的联系并得到范围固定在 0—1 之间的特征权重 $s = C_{16}(\text{Concat}(T_1, T_2, T_3, T_4, T_5, T_6))$ 。最后利用权重对 X^b 进行加权, 得到最后的双模态聚合的特征图 $X^c = s \otimes X^b$ 。

2.2.2 深层特征提取模块

由于遥感图像尺度差异较大, 因此提取多尺度信息有利于遥感图像的精准分割。SENet 经过全局平均池化层对特征图进行压缩。对像素级别的分割而言, SENet 仅考虑通道信息, 导致像素点之间的远程依赖较低, 无法获得全局之间的关系。因此本文引入了注意力机制的思想, 提出一个新的深层特征提取模块 DFEM 如图 5 所示。该模块从全局角度进一步提取深层语义信息, 同时过滤双模态融合产生的冗余信息, 进一步提高遥感图像分割准确率。

DFEM 去除了 SENet 中的全局平均池化层, 直接通过两个 3×3 卷积 C_{36} 、 C_{36} 获取全局上下文之间的依赖性, 对每个元素进行加权操作, 重新获得具有全局信息的深层特征图 $X^d = C_{36}(C_{36}(X^c)) \otimes X^c$ 。

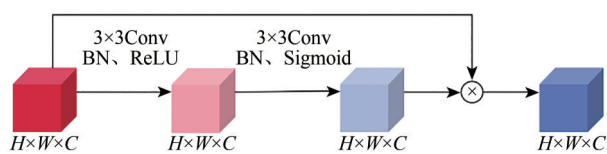


图5 深层特征提取模块结构图

Fig. 5 Deep feature extraction module structure diagram

2.3 解码模块

随着对遥感图像多层次的特征提取，图像分辨率逐渐降低，丢失很多细节信息。图像分割效果取决于上采样对高分辨率图像的信息恢复程度，因此为加强上采样细节信息的恢复，提出多层特征上采样模块 MFUM 如图 6 所示。通过卷积、池化、反池化等操作对特征图进行两次加权组合，使低层特征在上采样过程中不断获得高层特征丰富的指导信息，充分恢复上采样的细节信息，提高遥感图像分割的准确率。

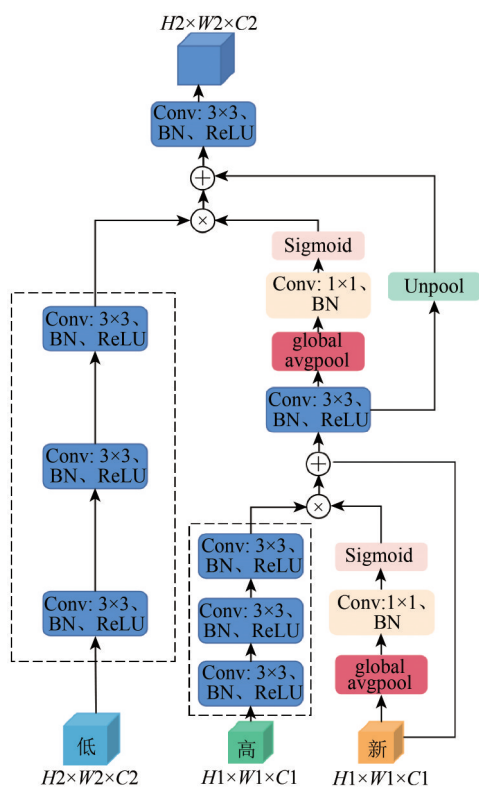


图6 多层特征上采样模块结构图

Fig. 6 Multilayer feature upsampling module structure diagram

该上采样模块有 3 个输入，其中两个输入为编码模块中 IRRG 编码器的语义分支经过第一个 3×3 卷积或 IEM 加强之后输出的相邻特征；第 3 个输入为 DFEM 或上一个 MFUM 输出的特征，分别称为低层特征、高层特征和最新特征。在 MFUM 中，

首先用最新特征对高层特征进行加权，得到新的特征，其次用新的特征对低层特征进行加权，实现上采样。执行 4 次 MFUM 后，经过 3×3 卷积输出遥感分割结果图。

虚线框表示对高层特征和低层特征均执行 3 次具有 BN 层和 ReLU 激活函数的 3×3 卷积操作，且保持通道数和尺寸大小不变，即前一个 MFUM 的低层特征也可作为后一个 MFUM 的高层特征。因此除了第一个上采样模块中对低层特征和高层特征进行卷积操作，其余 3 个模块均只对低层特征进行卷积操作。高层特征卷积后的特征图直接使用前一个模块的低层特征卷积后的特征图，在不影响上采样效果的同时大幅度降低了参数量。

DEFLNet 的解码部分重复使用了 4 次 MFUM，其运算过程可表示为

$$New_{i+1} = M(New_i, B_{5-i}, B_{(5-i)-1}), i = 0, 1, 2, 3 \quad (5)$$

式中， $M(\cdot)$ 表示 MFUM 的运算过程， New_0 为 DFEM 的输出特征图， $New_i (i = 1, 2, 3)$ 为第 i 个 MFUM 的输出特征图，均为最新特征； B_i 为 IRRG 编码器中语义分支输出的第 i 个尺寸的特征图， New_4 为精准的分割结果图。

3 数据结果处理与分析

本文使用随机梯度下降法 SGD (Stochastic Gradient Descent) 对模型进行训练，初始学习率设为 0.01，动量为 0.9，权重衰减为 0.0005，批量大小为 16。实验环境见表 1。

表 1 实验环境

Table 1 Experimental environment

软硬件	配置
操作系统	Windows10
CPU	12th Gen intel(R) Core(TM) i9-12900KF @ 3.19 GHz
GPU	NVIDIA GeForce RTX 3090 GPU
运行内存	32 G
显存	24 G
语言及框架	Python3.8.13+Pytorch1.12.1

3.1 数据集介绍

为了验证提出的 DEFLNet 对高分辨率遥感图像的分割性能，本文在 ISPRS Potsdam 和 Vaihingen 数据集上进行了实验。Potsdam 数据集包含 38 幅高分辨率遥感图像，Vaihingen 数据集包含 33 幅高分辨率遥感图像。每个数据集都分类为 5 个常见的土地覆盖类型：不透明表面、建筑物、低矮植被、

树木和汽车。ISPRS Potsdam 和 Vaihingen 数据集的样例如图 7、8 所示，分别为正射影像、DSM 图像和标签图。

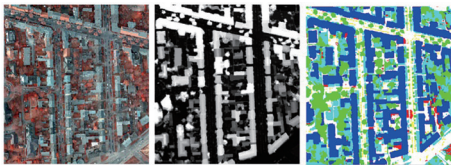


图 7 ISPRS Potsdam 数据集的样例
Fig. 7 Examples of ISPRS Potsdam dataset

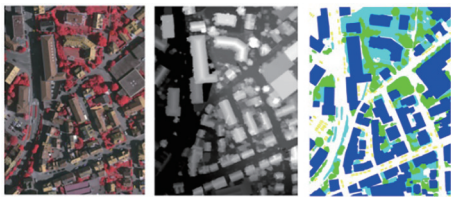


图 8 ISPRS Vaihingen 数据集的样例
Fig. 8 Examples of ISPRS Vaihingen dataset

Potsdam 数据集的尺寸达到 6000×6000，本文选取 24 幅用于训练，14 幅用于测试。Vaihingen 数据集的尺寸约为 2500×2500，本文选取 20 幅用于训练，13 幅用于测试。在训练网络之前，将数据集对应的 IRRG 图像、DSM 图像和标签图进行数据增强并随机裁剪为 224×224 大小。

3.2 实验结果

为了验证 DEFLNet 对遥感图像分割的有效性，本文在 Potsdam 数据集上进行实验。首先，为了验证提出的各个模块的性能，分别对主干网络特征提取、双模态融合方式、上采样方式和深层特征提取方式进行对比实验。其次，为了评估提出的各个模块的重要性，还对其进行消融实验。最后，将提出的 DEFLNet 与其他已有的模型进行对比实验。本文采用准确率、精准率、召回率、F1 指数和交并比 IoU（Intersection over Union）作为模型的评价指标来衡量遥感图像分割效果。

3.2.1 主干网络特征提取对比实验分析

为了验证 DEFLNet 的编码模块对 IRRG 图像和 DSM 图像的特征提取效果，将其与 MFFNet 的 3 分支特征提取结构和 AFNet 的多路径特征结构进行对比，实验结果如表 2 所示。

由表 2 可知，本文提出编码模块的准确率为 89.42%，比 MFFNet 的特征提取模块高 2.8%，比

AFNet 的特征提取模块高 0.98%。实验结果表明，本文提出的编码模块具有高效的特征提取能力和很好的分割性能。

表 2 主干网络对比 (Potsdam)
Table 2 Comparison of backbone network (Potsdam)
/%

主干网络	准确率	精确率	召回率	F1 指数	交并比
MFFNet	86.62	87.23	74.86	80.32	77.78
AFNet	88.44	88.95	75.67	81.57	78.13
DEFLNet	89.42	89.84	76.35	81.98	78.76

注：粗体表示最优值。

3.2.2 双模态融合方式对比实验分析

为了验证提出的 DFAB 的融合效果，本文将其与其他融合方式进行实验对比。以本文提出的编码模块为基础，融合方式分别采用 Fusetnet（Hazirbas 等，2016）中将编码过程的两种模态进行元素相加融合的 Add 操作；V-Fusetnet 在 Fusetnet 的基础上添加了虚拟路径 VP（Virtual Path）的操作；将两种模态特征进行级联融合的 Concat 操作；本文提出的 DFAB 操作。4 种不同融合方式的实验结果如表 3 所示。

表 3 双模态融合方式对比 (Potsdam)
Table 3 Comparison of dual-modal fusion methods (Potsdam)
/%

融合	准确率	精确率	召回率	F1 指数	交并比
Add	89.42	89.84	76.35	81.98	78.76
Add+VP	89.82	90.44	80.16	85.04	78.95
Concat	90.91	88.10	88.78	89.83	81.91
DFAB	92.22	92.74	90.43	91.63	85.94

注：粗体表示最优值。

由表 3 可知，本文提出的 DFAB 的准确率、精确率、召回率、F1 指数和 IoU 等评价指标都优于其他融合网络。这表明 DFAB 具有高效的特征融合能力，可以有效提升分割准确性。

3.2.3 上采样方式对比实验分析

已有的分割网络主要通过下采样操作对多尺度、多层次的图像进行特征提取，但该方法容易丢失较多高分辨率图像的细节信息，因此后续信息恢复效果将直接影响遥感图像的分割效果。本文利用 MFUM 对图像特征进行上采样操作，与已有的上采样方式反池化（unpool）、Unet 中的双线

性插值操作和全局注意力上采样 GAU (Global Attention Upsample) (Li等, 2018) 进行实验对比, 4种不同上采样方式的实验结果如表4所示。

表4 上采样方式对比(Potsdam)
Table 4 Comparison of Upsampling Methods (Potsdam)

上采样	准确率	精确率	召回率	F1 指数	交并比
unpool	92.22	92.74	90.43	91.63	85.94
Unet	92.94	93.63	92.01	92.47	86.39
GAU	93.91	93.10	92.98	93.44	87.91
MFUM	94.01	93.78	93.21	93.61	89.54

注: 粗体表示最优值。

由表4可知, 本文提出的MFUM的各项指标均优于其他上采样模块。在相同的双模态融合方式下, 与unpool上采样、Unet上采样和GAU上采样相比, MFUM的准确率分别提升了1.79%, 1.07%和0.1%; 精确率分别提升了1.04%, 0.15%和0.68%; 召回率分别提升了2.78%, 1.2%和0.23%; F1指数分别提升了2.28%, 1.14%和0.17%; IoU分别提升了3.6%, 3.15%和1.63%。实验结果表明MFUM具有高效的上采样细节信息恢复能力, 可以实现精准分割。

3.2.4 深层特征提取方式对比实验分析

随着遥感图像分辨率逐渐降低, 感受野逐渐变大, 特征图从细节信息逐渐往语义信息转变。语义信息可以强化网络对各分割类别的理解, 因此有效提取图像的深层特征有利于进一步改善遥感图像分割效果。本文在融合模块中增加DFEM, 同时构建了DEFLNet, 与添加SE通道注意力进行对比, 实验结果如表5所示。

表5 深层特征提取方式对比(Potsdam)
Table 5 Comparison of deep feature extraction methods (Potsdam)

特征提取	准确率	精确率	召回率	F1 指数	交并比
SE	93.88	94.06	93.78	93.83	89.98
DFEM	94.64	94.52	94.54	94.59	90.30

注: 粗体表示最优值。

由表5可知, 与SE相比, 本文提出的DFEM分割效果更好, 准确率提升了0.76%, 精确率提升了0.46%, 召回率提升了0.76%, F1指数提升了0.76%, IoU提升了0.32%。

3.2.5 模块消融实验

为了评估提出的各个模型的重要性, 本节将逐步分解DEFLNet以揭示每个模块的效果。消融实验中基准网络(Baseline)的主干网络采用本文提出的CM去除IEM后的结构, 融合多路径特征和聚合双模态特征均采用元素相加的方式, 上采样采用unpool的操作。

由表6可知, 融合多路径特征时, 使用IEM替换基准网络中元素相加的方式, 使得遥感图像分割的各项指标均最高。与基准网络相比, 准确率提升了1.28%, 精确率提升了1.19%, 召回率提升了1.19%, F1指数提升了0.87%, 交并比提升了0.88%。因此, IEM是DEFLNet中提升性能最重要的模块, MFUM次之。

表6 各个模块的消融实验(Potsdam)
Table 6 Ablation Experiments of Each Module (Potsdam)

方法	准确率	精确率	召回率	F1 指数	交并比
Baseline	88.14	88.65	75.16	81.12	77.88
Baseline+IEM	89.42	89.84	76.35	81.99	78.76
Baseline+DFAB	88.93	89.37	75.97	81.93	78.39
Baseline+MFUM	89.12	89.43	76.03	81.98	78.47
Baseline+DFEM	88.56	88.83	75.64	81.60	78.04

注: 粗体表示最优值。

3.2.6 DEFLNet与其他模型对比实验分析

为了证明本文提出的DEFLNet的分割性能, 将其与已有先进遥感图像分割网络MFFNet、MFANet, 已有先进的双模态遥感图像分割网络HA-MPPNet、AFNet、MMFNet和MSFAFNet进行对比。各类别的F1指数和总体精确率的实验结果如表7所示。

由表7可知, 本文的DEFLNet比已有的遥感图像分割网络拥有更好的分割性能。DEFLNet在不透明表面、建筑物、低矮植被、树木和汽车等类别达到了最好的分割性能, 各类别的F1指数均最高, 分割精度高达94.52%。与AFNet相比, 各类别的F1指数分别提升了0.02%, 0.26%, 1.22%, 0.48%和0.06%, 总体精确率提升了2.07%。

同时, 本文在Vaihingen数据集上进行了相同的对比实验, 结果如表8所示。

为了进一步验证所提出的DEFLNet分割的优越性, 将该模型与MMFNet、MSFAFNet、HA-MPPNet和AFNet在Potsdam和Vaihingen数据集进

行了泛化对比实验。图 9、图 10 和图 11 对 Potsdam 数据集中不同类别的特征分别进行分析，主要分为光谱特征非常相似的低矮植被和树木、光谱特征非常相似的道路和建筑物、较小且易被遮挡的汽车。图 12 是 Vaihingen 数据集分割结果对比。

表 7 DEFLNet 与其他模型对比 (Potsdam)
Table 7 Comparison of DEFLNet with other models (Potsdam)

模型	不透明表面	建筑物	低矮植被	树木	汽车	总体精度
MFFNet	88.25	92.93	86.56	86.13	89.78	87.88
MFANet	90.54	94.21	87.75	87.78	90.86	89.63
MMFNet	91.46	96.23	88.43	88.24	91.88	90.83
MSFAFNet	92.85	96.45	89.07	89.72	93.76	91.63
HA-MPPNet	93.34	96.76	90.12	90.85	93.94	91.76
AFNet	93.95	96.68	89.65	91.36	94.08	92.45
DEFLNet	93.97	96.94	90.87	91.84	94.14	94.52

注：粗体表示最优值。

表 8 DEFLNet 与其他模型对比 (Vaihingen)
Table 8 Comparison of DEFLNet with other models (Vaihingen)

模型	不透明表面	建筑物	低矮植被	树木	汽车	总体精度
MFFNet	87.53	90.32	84.07	87.63	76.57	86.13
MFANet	89.16	92.83	84.59	88.47	77.86	87.37
MMFNet	90.04	94.95	85.77	91.13	78.34	89.03
MSFAFNet	90.53	95.34	86.98	92.89	79.84	89.74
HA-MPPNet	90.87	95.51	87.27	92.94	80.67	89.56
AFNet	91.15	95.67	86.88	93.17	81.31	90.22
DEFLNet	91.46	95.88	87.35	93.68	81.88	90.45

注：粗体表示最优值。

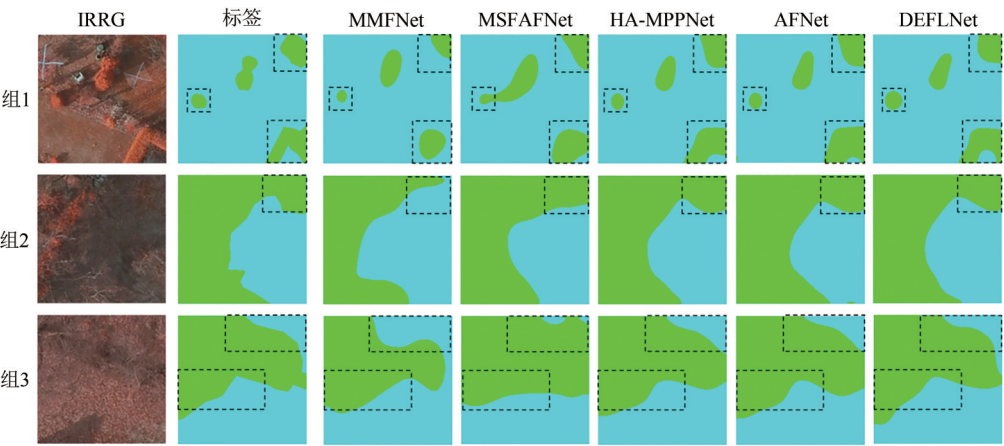


图 9 Potsdam 中低矮植被和树木分割图
Fig. 9 Segmentation of low vegetation and trees in Potsdam

图 9 表示分割网络对低矮植被和树木的分割结果，低矮植被和树木具有非常相似的光谱特征，分割难度较大。以第一组图片为例，图中树木较少且分布较分散。MMFNet 的右上方区域存在大量将树木分割为低矮植被的情况，且右下方区域存在严重的粗略分割的情况；MSFAFNet 右上方区域错误分割较少，但是在左中的交界区域仍存在大量粗略分割，错将低矮植被分割为树木；HA-MPPNet

和AFNet在错误分割方面有所改善，但是右下方区域边缘分割仍然不够清晰；DEFLNet的分割效果最佳，分割错误的区域最少，左中区域的小树木与标签图最符合，且右下方区域的边缘分割最清晰。

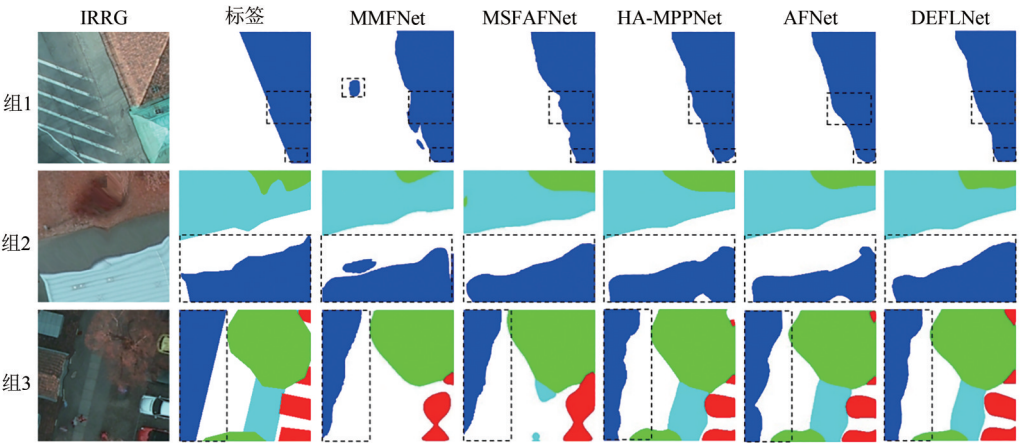


图 10 Potsdam 中建筑物和道路分割图
Fig. 10 Segmentation of buildings and roads in Potsdam

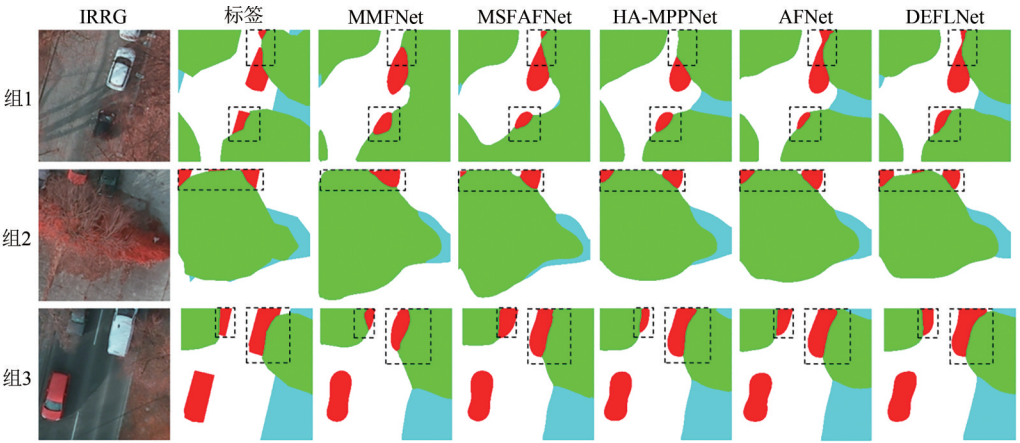


图 11 Potsdam 中汽车分割图
Fig. 11 Segmentation of cars in Potsdam

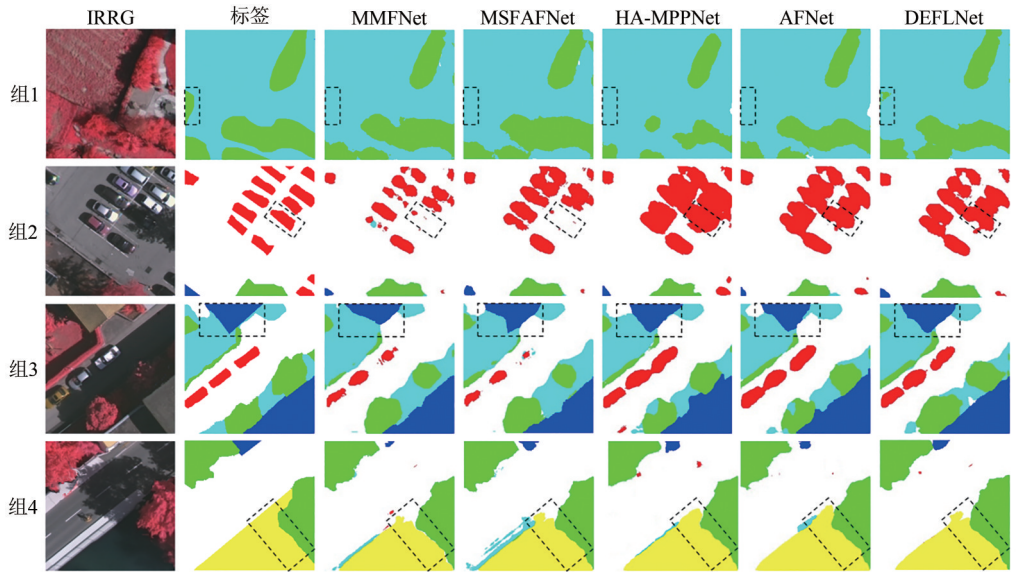


图 12 Vaihingen 分割图
Fig. 12 Segmentation in Vaihingen

图10表示分割网络对建筑物和背景（道路）的分割结果，建筑物和道路具有非常相似的光谱特征，分割难度较大。以第一组图片为例，图中建筑物和道路斜对角分布，道路区域较多。MMFNet右边区域和左中区域存在多处将道路分割为建筑物的情况；MSFAFNet右中区域存在较多将建筑物分割为道路的情况，且分割边缘模糊；HA-MPPNet右中区域分割错误区域较少，但是右下区域存在小部分将建筑物分割为道路的情况；AFNet分割错误区域有所改善，但是边缘分割较粗糙；DEFLNet中建筑物和道路分割最精准，且两者的边缘分割最清晰。

图11表示分割网络对汽车的分割结果，汽车尺寸较小且容易被遮挡，实现精准分割难度较大。第一组图片中汽车被大量遮挡，MMFNet和HA-MPPNet的右上方区域均存在将汽车检测为树木的情况；MSFAFNet和AFNet对汽车的分割效果较好，但是MSFAFNet中下方区域边缘分割模糊，AFNet中下区域分割错误较多；DEFLNet对汽车的分割效果最佳，中下方区域汽车的错误分割的部分最少。

图12表示分割网络对Vaihingen数据集的分割结果，第一组以树木和低矮植被为主，对左边区域较少的树木进行分割时，其余4个网络均没有检测出树木，DEFLNet识别出部分树木，分割最精准。第二组以汽车为主，对特征不明显的汽车进行分割时，MMFNet和MSFAFNet将其错误分割为道路，HA-MPPNet和AFNet能识别汽车，但是边缘分割模糊，DEFLNet能够完整分割出汽车，准确度最高且边缘清晰度有所改善。第三组以建筑物为主，本文模型分割建筑物边缘最清晰，与标签图最符合。第四组以不透明表面为主，其余4个网络均存在将不透明表面分割为低矮植被的情况，而DEFLNet的分割效果最好。

4 结 论

针对已有双模态网络在遥感图像分割上存在特征提取不充分、特征融合不合理、上采样特征恢复不足等问题，本文提出了一个基于双模态网络的遥感图像分割算法DEFLNet。首先，提出针对IRRG图像和DSM图像特点的编码模块。该模块分别提取IRRG图像的细节信息、语义信息和DSM图像的空间信息，并在提取的过程中不断交互融合。其次，提出双模态特征聚合模块动态调整两个模态特征之间的映射关系，有效融合两种模态

的信息。此外，提出深层特征提取模块从全局角度通过注意力获取深层语义信息，进一步提高特征提取的充分性。最后，提出多层特征上采样模块对低分辨率图像进行多层特征融合。在上采样过程中，高层特征为低层特征不断提供指导信息，实现特征的有效恢复。本文在ISPRS Potsdam和Vaihingen数据集上进行实验，结果表明与其他分割模型相比，DEFLNet更适用于分割光谱特征非常相似的低矮植被和树木、建筑物和道路。不仅如此，对汽车等小目标也能实现精准分割。

本文提出的DEFLNet达到显著提升分割准确度的效果，但是需要进一步降低模型的复杂度，且准确度仍有一定的提升空间。后续可以考虑融合两种以上的模态特征，获取更多特征信息，设计出更好的分割网络以实现更加精准的遥感图像分割。

参考文献 (References)

- Ahmed O S, Shemrock A, Chabot D, Dillon C, Williams G, Wasson R and Franklin S E. 2017. Hierarchical land cover and vegetation classification using multispectral data acquired from an unmanned aerial vehicle. *International Journal of Remote Sensing*, 38(8/10): 2037-2052 [DOI: 10.1080/01431161.2017.1294781]
- Audebert N, Le Saux B and Lefèvre S. 2018. Beyond RGB: very high resolution urban remote sensing with multimodal deep networks. *ISPRS Journal of Photogrammetry and Remote Sensing*, 140: 20-32 [DOI: 10.1016/j.isprsjprs.2017.11.011]
- Chen B Y, Xia M and Huang J Q. 2021a. MFANet: a multi-level feature aggregation network for semantic segmentation of land cover. *Remote Sensing*, 13(4): 731 [DOI: 10.3390/rs13040731]
- Chen L C, Papandreou G, Kokkinos I, Murphy K and Yuille A L. 2018. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(4): 834-848 [DOI: 10.1109/TPAMI.2017.2699184]
- Chen S T, Wu C Q, Mukherjee M and Zheng Y J. 2021b. HA-MPPNet: height aware-multi path parallel network for high spatial resolution remote sensing image semantic segmentation. *ISPRS International Journal of Geo-Information*, 10(10): 672 [DOI: 10.3390/ijgi10100672]
- Farabet C, Couprie C, Najman L and LeCun Y. 2013. Learning hierarchical features for scene labeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8): 1915-1929 [DOI: 10.1109/TPAMI.2012.231]
- Gerard F, Petit S, Smith G, Thomson A, Brown N, Manchester S, Wadsworth R, Bugar G, Halada L, Bezák P, Boltiziar M, De Badts E, Halabuk A, Moyses M, Petrovic F, Gregor M, Hazeu G, Múcher C A, Wachowicz M, Huitu H, Tuominen S, Köhler R, Olschofsky

- K, Ziese H, Kolar J, Sustera J, Luque S, Pino J, Pons X, Roda F, Roscher M and Feranec J. 2010. Land cover change in Europe between 1950 and 2000 determined employing aerial photography. *Progress in Physical Geography: Earth and Environment*, 34(2): 183-205 [DOI: 10.1177/0309133309360141]
- Hazirbas C, Ma L N, Domokos C and Cremers D. 2016. FuseNet: incorporating depth into semantic segmentation via fusion-based CNN architecture//*Proceedings of the 13th Asian Conference on Computer Vision*. Taipei, China: Springer: 213-228 [DOI: 10.1007/978-3-319-54181-5_14]
- He K M, Zhang X Y, Ren S Q and Sun J. 2016. Deep residual learning for image recognition//*Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas: IEEE: 770-778 [DOI: 10.1109/CVPR.2016.90]
- Hu J, Shen L, Albanie S, Sun G and Wu E H. 2020. Squeeze-and-excitation networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(8): 2011-2023 [DOI: 10.1109/TPA-MI.2019.2913372]
- Li H C, Xiong P F, An J and Wang L X. 2018. Pyramid attention network for semantic segmentation. *arXiv preprint arXiv: 1805.10180*
- Marcos D, Volpi M, Kellenberger B and Tuia D. 2018. Land cover mapping at very high resolution with rotation equivariant CNNs: towards small yet accurate models. *ISPRS Journal of Photogrammetry and Remote Sensing*, 145: 96-107 [DOI: 10.1016/j.isprsjrs.2018.01.021]
- Marmanis D, Wegner J D, Galliani S, Schindler K, Datcu M and Stilla U. 2016. Semantic segmentation of aerial images with an ensemble of CNNs. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, III-3: 473-480 [DOI: 10.5194/isprs-annals-III-3-473-2016]
- Mnih V and Hinton G E. 2010. Learning to detect roads in high-resolution aerial images//*Proceedings of the 11th European Conference on Computer Vision*. Heraklion: Springer: 210-223 [DOI: 10.1007/978-3-642-15567-3_16]
- Paisitkriangkrai S, Sherrah J, Janney P and Van-Den Hengel A. 2015. Effective semantic pixel labelling with convolutional networks and Conditional Random Fields//*Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. Boston: IEEE: 36-43 [DOI: 10.1109/CVPRW.2015.7301381]
- Ronneberger O, Fischer P and Brox T. 2015. U-net: convolutional networks for biomedical image segmentation//*Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention*. Munich: Springer: 234-241 [DOI: 10.1007/978-3-319-24574-4_28]
- Shelhamer E, Long J and Darrell T. 2017. Fully convolutional networks for semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(4): 640-651 [DOI: 10.1109/TPAMI.2016.2572683]
- Sun H Q, Pan C, He L M and Xu Z J. 2022. Remote sensing image semantic segmentation network based on multimodal feature fusion. *Computer Engineering and Applications*, 58(24): 256-264 (孙汉淇, 潘晨, 何灵敏, 胥智杰. 2022. 多模态特征融合的遥感图像语义分割网络. *计算机工程与应用*, 58(24): 256-264) [DOI: 10.3778/j.issn.1002-8331.2207-0010]
- Sun J X and Li Y J. 2021. Multi-feature fusion network for road scene semantic segmentation. *Computers and Electrical Engineering*, 92: 107155 [DOI: 10.1016/j.compeleceng.2021.107155]
- Wang S L. 2005. The magic eye of modern war-military application of satellite remote sensing image. *China Surveying and Mapping*, (1): 34-37 (王树连. 2005. 现代战争的神奇天眼——卫星遥感图像的军事应用. *中国测绘*, (1): 34-37)
- Wang X L, Girshick R, Gupta A and He K M. 2018. Non-local neural networks//*Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City: IEEE: 7794-7803 [DOI: 10.1109/CVPR.2018.00813]
- Wu G M, Shao X W, Guo Z L, Chen Q, Yuan W, Shi X D, Xu Y W and Shibasaki R. 2018. Automatic building segmentation of aerial imagery using multi-constraint fully convolutional networks. *Remote Sensing*, 10(3): 407 [DOI: 10.3390/rs10030407]
- Xu Y Y, Wu L, Xie Z and Chen Z L. 2018. Building extraction in very high resolution remote sensing imagery using deep learning and guided filters. *Remote Sensing*, 10(1): 144 [DOI: 10.3390/rs10010144]
- Yang X, Li S S, Chen Z C, Chanussot J, Jia X P, Zhang B, Li B P and Chen P. 2021. An attention-fused network for semantic segmentation of very-high-resolution remote sensing imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 177: 238-262 [DOI: 10.1016/J.ISPRSJPRS.2021.05.004]
- Zhang W K, Liu W J, Sun X, Xu G L and Fu K. 2022. Multi-source features adaptation fusion network for semantic segmentation in high-resolution remote sensing images. *Journal of Image and Graphics*, 27(8): 2516-2526 (张文凯, 刘文杰, 孙显, 许光鑫, 付琨. 2022. 多源特征自适应融合网络的高分遥感影像语义分割. *中国图象图形学报*, 27(8): 2516-2526) [DOI: 10.11834/jig.210054]
- Zhang X J and Wang X L. 2020. Image segmentation models of remote sensing using full residual connection and multiscale feature fusion. *Journal of Remote Sensing (Chinese)*, 24(9): 1120-1133 (张小娟, 汪西莉. 2020. 完全残差连接与多尺度特征融合遥感图像分割. *遥感学报*, 24(9): 1120-1133) [DOI: 10.11834/jrs.20208365]
- Zhou W Q, Huang G L and Cadenasso M L. 2011. Does spatial configuration matter? Understanding the effects of land cover pattern on land surface temperature in urban landscapes. *Landscape and Urban Planning*, 102(1): 54-63 [DOI: 10.1016/j.landurbplan.2011.03.009]

High resolution remote sensing image segmentation based on dual-modal efficient feature learning

ZHANG Yinsheng^{1,2}, JI Ru², TONG Junyi², YANG Yulong², HU Yuxiang¹, SHAN Huilin^{1,2}

1.School of Electronic and Information Engineering, Wuxi University, Wuxi 214105, China;

2.School of Electronic and Information Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China

Abstract: With the rapid development of spatial technology, the resolution of remote sensing images gradually improves. The detailed information and spatial information contained in remote-sensing images are also richer. The ensuing problems are that the difference between various categories becomes and the difference between the same categories becomes larger, i.e., the phenomenon of the same spectrum of foreign objects and the different spectrum of the same objects is serious. However, the existing dual-modal segmentation methods do not extract the dual-modal feature information of remote-sensing images separately, and the fusion features are insufficient. The details of upsampling recovery are also insufficient, resulting in the inability to accurately and efficiently learn remote-sensing image information, thereby resulting in segmentation errors, edge blur, and other problems.

This study proposes a high resolution remote-sensing image segmentation based on dual-modal efficient feature learning. The algorithm designs appropriate encoders for different modal remote sensing images, efficiently extracts dual-modal features, and reduces the differences between different path features through interactive reinforcement modules. Then, the dual-modal feature aggregation module and the deep feature-extraction module are proposed to further fuse and extract the dual-modal features. As a result, the network can fully learn the complementary information of the dual-modal. Finally, a multi-layer feature upsampling module is proposed, which uses high-level features with rich semantic information to weight the low-level features with rich detail information. Gradual upsampling is then conducted to achieve efficient feature recovery and improve segmentation performance.

In this paper, experiments on the Potsdam and Vaihingen datasets demonstrate that the overall accuracy reaches 94.52% and 90.45%, respectively. Experimental results show that the segmentation effect of the proposed algorithm is better than that of existing algorithms. The proposed algorithm can efficiently extract and fuse the multi-modal complementary features of high resolution remote-sensing images and improve the segmentation accuracy of remote-sensing images.

This study proposes a high-resolution remote-sensing image segmentation based on dual-modal efficient feature learning. Experiments on the ISPRS Potsdam and Vaihingen datasets show that the proposed model is more suitable for segmenting low vegetation and trees, buildings, and roads with very similar spectral features. It can also achieve the accurate segmentation of small targets, such as cars. However, the complexity of the model needs to be further reduced, and much room for improvement in accuracy remains. In the future, a better segmentation network will be designed to fuse more than two modal features and thus obtain more feature information to achieve more accurate remote sensing image segmentation.

Key words: remote sensing image segmentation, efficient feature extraction, integration, dual-modal feature aggregation, deep feature extraction, multilayer feature upsampling

Supported by National Natural Science Foundation of China (No. 62071240, 62106111); Postgraduate Research and Practice Innovation Program of Jiangsu Province (No. SJCX23_0379)