

基于波段-位置自适应选择的多光谱 遥感图像语义分割

梁正印, 汪西莉*

陕西师范大学计算机科学学院, 陕西 西安 710000

摘要 多光谱遥感图像(MSIs)包含大量的地物信息,这些信息蕴含在图像的多个光谱波段中。不同波段或者同一波段不同空间位置所含信息量差异很大,如何从MSIs中捕获有效信息是遥感图像语义分割中一项具有挑战性的任务。基于此,提出一种基于波段-位置自适应选择的端到端语义分割网络(BLASENet)。所提网络采用编码器-解码器结构,在编码阶段,提出波段-位置自适应选择机制来自适应学习不同波段和同一波段不同空间位置权重,增强有效特征表达。为了利用MSIs的波段相关性,进一步提出三维残差块编码图像的光谱-空间特征。在解码阶段,提出自适应特征融合模块,通过网络学习自适应调整低级细节特征与高级语义特征的融合比例,并探究加法(BLASENet-A)、元素乘法(BLASENet-M)和串联(BLASENet-C)等3种融合策略对模型性能增益的影响。此外,将通道注意力扩展到三维数据上,对融合后的特征图在通道维度上进行特征重标定,得到更准确的多级交互特征图。在ISPRS Potsdam、Qinghai和Tibet Plateau等3个数据集上的实验结果证明了BLASENet的有效性。

关键词 图像处理; 语义分割; 三维卷积; 波段-位置自适应选择机制; 注意力机制

中图分类号 TP391

文献标志码 A

DOI: 10.3788/LOP222250

Semantic Segmentation of Multispectral Remote Sensing Images Based on Band-Location Adaptive Selection

Liang Zhengyin, Wang Xili*

School of Computer Science, Shaanxi Normal University, Xi'an 710000, Shaanxi, China

Abstract Multispectral remote sensing images (MSIs) provide a substantial amount of ground object information spread over various spectral bands of the image. The quantity of information contained in different bands or different spatial locations within the same band varies significantly. How to capture useful information from MSIs is a challenging task in semantic segmentation of remote sensing images. An end-to-end semantic segmentation network (BLASENet) based on band-location adaptive selection is proposed here. The proposed network adopts an encoder-decoder structure. In the coding phase, a band-location adaptive selection mechanism is proposed to adaptively learn the weights of different bands and different spatial locations within the same band, enhancing the effective features expression. The spectral-spatial features of 3D residual block-coded images are further proposed to make use of the band correlation of MSIs. During the decoding phase, an adaptive feature fusion module is proposed to adaptively adjust the fusion ratio of low-level detail features and high-level semantic features via network learning, as well as investigate the impact of three fusion strategies, namely, addition (BLASENet-A), element multiplication (BLASENet-M), and concatenation (BLASENet-C), on the model's performance gain. Furthermore, channel attention is extended to 3D data, and the fused feature map is recalibrated on the channel dimension to produce a more accurate multi-level interactive feature map. The effectiveness of BLASENet has been demonstrated by experimental results on ISPRS Potsdam, Qinghai and Tibet Plateau datasets.

Key words image processing; semantic segmentation; 3D convolution; band-location adaptive selection mechanism; attention mechanism

收稿日期: 2022-08-05; 修回日期: 2022-09-05; 录用日期: 2022-09-26; 网络首发日期: 2022-10-06

基金项目: 第二次青藏高原综合科学考察研究项目(2019QZKK0405)

通信作者: *wangxili@snnu.edu.cn

1 引言

语义分割根据不同的类别为每个像素赋予一个标签,多光谱遥感图像(MSIs)可以通过语义分割实现地物分类,在土地覆盖、环境变化等领域发挥着重要作用。传统语义分割方法主要依据目标的几何(如形状、空间位置)、纹理、颜色等属性设计特征^[1],通过阈值、聚类、分类算法对目标实现像素级分类^[2-3],从而对整个图像进行分割。传统方法性能很大程度上取决于人工设计的特征,高度依赖于领域知识,具有一定的局限性。

近年来,基于深度学习的语义分割方法引起了研究人员的广泛关注。与传统方法需要人工设计特征不同,基于深度学习的方法通过迭代训练能够自动提取到图像的多层次特征。其中,最具有代表性的是卷积神经网络(CNN)。CNN在分类、分割、检测等图像视觉任务中表现出了卓越的性能^[4-7]。在图像分割任务中,需要获得像素级别的分类图,单靠提取的高级语义信息无法实现准确的密集预测。常见的解决方案是在上采样的过程中融入浅层细节信息。如基于全卷积网络(FCN)^[8]的方法,U-Net^[9]和SegNet^[10]等。U-Net采用从浅层到深层的跳跃连接策略,拼接不同级特征来保留图像细节信息。SegNet在编码阶段记录最大池化层的池化索引,上采样时通过池化索引恢复更多的图像细节信息。RefineNet^[11]利用下采样阶段的多层信息,以远程残差连接方式实现高分辨率预测。此外,使用可训练的反卷积^[12]能够有效恢复特征图的空间分辨率。同一张图像中不同物体的大小、分布存在差异,导致不同物体分割的难度也大不相同。常见的做法是通过改进损失函数来使模型更加关注于困难物体,从而提高整个图像分割的准确度。如focal损失^[13]使用动态缩放因子来降低训练过程中易分类物体的权重,从而使模型聚焦于困难物体。交叉熵损失(WCE)^[14]在交叉熵损失函数的基础上,对困难物体施加更大的权重来提高分割准确度。

与自然图像不同,MSIs语义分割存在两个挑战:1) MSIs成像范围广,存在同物异谱、同谱异物的特点,单靠地物的光谱特征难以实现准确分割;2) MSIs有多个光谱波段,不同波段所含信息量差异很大,捕获有效信息更为困难。对于挑战1,建模上下文信息有助于分割MSIs复杂场景。多数方法采用类似U-Net的跳跃连接将不同级特征直接融合^[15],来生成多尺度上下文特征,但其较为粗糙且没有进一步探究不同级特征的贡献程度。因此,本文提出一个自适应特征融合(ADF)模块来探究不同级特征对模型分割性能的影响。另外,注意力机制能够让网络更加关注有意义的特征,在语义分割任务中引入注意力机制能够有效捕获MSIs上下文信息^[16]。对于挑战2,现有MSIs语义分割方法多采用二维卷积处理每个波段的信息,不

考虑MSIs不同波段及其相关性对分割的影响。因此,提出波段-位置自适应选择(BLAS)机制,选择MSIs有效波段的同时捕获有意义的空间信息,进一步设计一种三维残差块(3DRB)来编码MSIs的光谱-空间特征。在上述工作的基础上,提出一个端到端的语义分割模型(BLASENet)。

2 相关工作

2.1 多光谱遥感图像语义分割

MSIs具有地物类别多、分布不均匀,且存在类内方差大、类间方差小的特点,难以准确分割。大量研究表明,建模上下文信息有利于MSIs复杂场景分割。目前已经有很多方法用于捕获图像的上下文信息。例如,Zhao等^[17]提出一种金字塔池化模块(PPM),通过拼接4种不同金字塔尺度的特征,捕获全局上下文信息。Yu等^[18]将PPM引入MSIs分割任务中,取得了令人鼓舞的效果。Ding等^[19]提出的两阶段多尺度训练策略可有效扩大网络的感受野。Mou等^[20]提出一种空间关系感知模块来建模目标间的中远程空间关系,通过逐元素点积运算生成包含空间上下文信息的关系特征图,并将生成的关系特征与原始特征串联起来,从而保留图像原始的高级语义特征。

另一个棘手的问题是,MSIs不同地物类型存在较大差异,甚至同一类型在不同空间位置上也存在尺度上的不同。具体表现在,不同场景下的房屋、道路在形状、大小上也会存在类内不一致的特点。因此,模型捕获多尺度特征的能力便显得尤为重要。例如,Chen等^[21]提出空洞空间金字塔池化(ASPP)模块,ASPP模块通过采用多个具有不同扩张率(dilation rate)的空洞卷积提取图像的多尺度特征,但空洞卷积也会造成一些细节信息的丢失。对于语义分割任务而言,能否从特征图中恢复出准确的输出分数图直接影响到分割质量。由于高层特征图仅包含高级语义信息,生成的分割图存在边缘模糊的问题。融合多层次特征能够缓解边缘模糊问题。例如,Chen等^[22]将含有不同尺度信息的高级特征与含有空间细节信息的低级特征串联,在上采样恢复特征图空间分辨率的同时,恢复目标边界信息。Li等^[23]设计了一个自适应多尺度特征融合模块,通过全连接层拟合一个线性函数来学习骨干网络中不同层的特征权重,使有效的层次特征在融合特征中更具表达性。另外,由于超像素内部的同质性和超像素之间的异质性,基于超像素的方法也能够缓解边缘模糊问题。例如,Mi等^[24]基于超像素提出一个区域损失,赋予边界像素更大的权重来准确分割图像边界轮廓。

MSIs通常包含几个到十几个光谱波段,如何从多个波段中捕获到具有代表性的地物信息是一个值得考虑的问题。对于高光谱图像(HSIs)分类,有学者基于三维卷积方法研究波段关系对分类性能的影响^[25-27]。

然而目前关于 MSIs 语义分割的研究大多关注通道或者空间维度上的关系建模^[28-29],没有考虑波段间的相关性。因此,探究 MSIs 多个波段对语义分割任务的影响是本研究的一个动机。

2.2 注意力机制

注意力机制在计算机视觉任务中取得了巨大的成功^[30]。越来越多的研究者致力于开发更加有效的注意力模块。SENet^[31]提出通道注意力来显式建模通道之间的相互依赖关系,它在通道上执行全局平均池化来聚集全局空间信息,然后使用两个全连接层来捕获全局跨通道交互信息。有学者发现,捕获所有通道之间的依赖关系是低效且不必要的。因此,ECA-Net^[32]提出一种局部跨通道交互策略,将SENet中的全连接层替换为一维卷积层,通过卷积方式实现局部跨通道交互的同时降低了模型复杂度。FcaNet^[33]从频域视角出发,提出一种多光谱通道注意力,使用离散余弦变换(DCT)为每个通道分配不同的频率分量,提取更多的通道信息。convolutional block attention module (CBAM)^[34]使用串行的方式将通道注意力和空间注意力连接起来,捕获有效通道的同时增强空间局部区域的特征表达。

对于 MSIs 语义分割,集成注意力机制的方法也得到了很大的发展。Li等^[35]将通道注意力和空间注意力整合到SegNet分割网络中,通过空间注意力建模

MSIs 空间上下文信息以提取有意义的空间特征,通道注意力在通道维度上选取有效特征。Zhao等^[36]提出一个金字塔注意力池化模块,将注意力机制引入多尺度模块中进行自适应特征细化。Li等^[37]提出一种轻量型线性注意力并将其引入U-Net跳跃连接中,进一步细化获得的多尺度特征图。无论是空间注意力还是通道注意力,这些工作都没有考虑 MSIs 多个光谱波段的相关性,也没有进一步考虑不同波段信息的冗余度。为此,提出波段-位置自适应选择(BLAS)机制,通过网络在光谱域和空间域自适应建模位置权重,选择有效波段的同时捕获有意义的空间信息。

3 所提方法

BLASeNet的结构如图1所示,由编码器和解码器两部分组成。对于输入的 MSIs,首先使用BLAS捕获 MSIs 富含信息的波段并建模空间位置权重,然后通过9个3DRB编码 MSIs 的光谱-空间特征。在每个3DRB中,加入BLAS层进行特征权重再学习。在解码阶段,通过ADF模块自适应调整不同级特征的融合比例,融合后的特征图被送入扩展的通道注意力模块(CAM)上进行特征细化(特征融合与特征细化操作执行两次)。最后,通过二维卷积将学习到的特征映射到样本类别空间,通过Softmax函数获得语义分割结果。

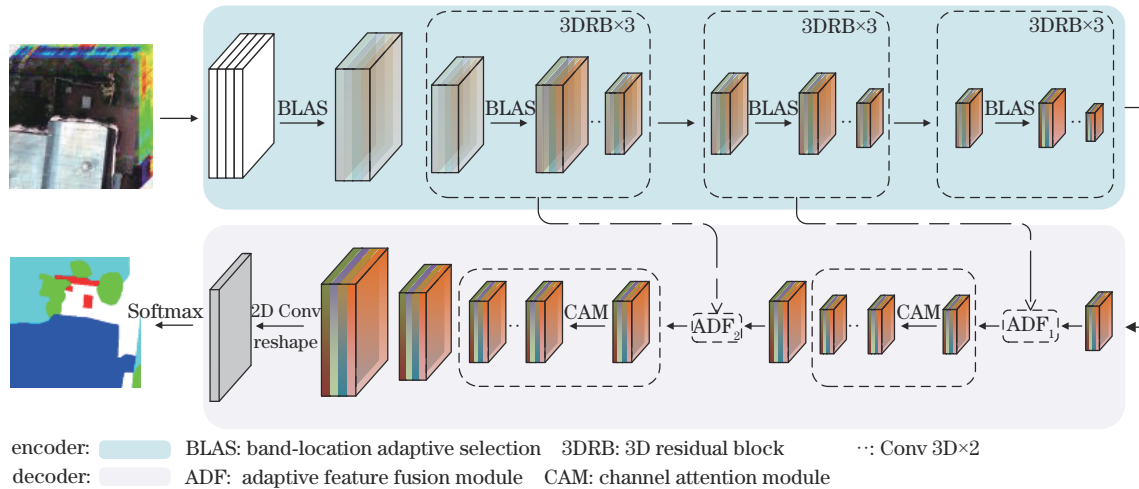


图1 BLASeNet结构示意图

Fig. 1 Structure schematic of BLASeNet

3.1 波段-位置自适应选择机制

从 MSIs 中捕捉有效信息的能力对于语义分割任务十分重要。所提 BLAS 机制如图 2 所示,输入的 MSIs 记作 $F_i \in \mathbf{R}^{N \times H \times W}$, N 、 H 和 W 分别表示为 MSIs 光谱波段的数量、高度和宽度。为了捕获富含有效信息的光谱波段,首先使用全局平均池化(GAP)为每个波段提供全局上下文信息,然后通过一维卷积学习波段权重图 $W_n \in \mathbf{R}^{N \times 1 \times 1}$ 。权重值由 Sigmoid 函数缩放至 0~1。计算公式如下:

$$W_n = \sigma \left\{ \text{Conv1d}^3 \left[\delta(F_i) \right] \right\}, \quad (1)$$

$$\delta(F_i) = \frac{1}{HW} \sum_{h=0}^{H-1} \sum_{w=0}^{W-1} F_{i(:,h,w)}, \quad (2)$$

式中: $\delta(\cdot)$ 表示逐波段全局平均池化操作; $\text{Conv1d}^3(\cdot)$ 表示卷积核大小为3的一维卷积操作; $\sigma(\cdot)$ 表示 Sigmoid 激活函数。通过网络学习,富含有效信息的光谱波段将被赋予较大的权重,而含有冗余信息的波段将被赋予较小的权重。然后,对输入图像 F_i 与波段权重图 W_n 进行哈达玛积运算,得到赋予不同波段权重的三维特

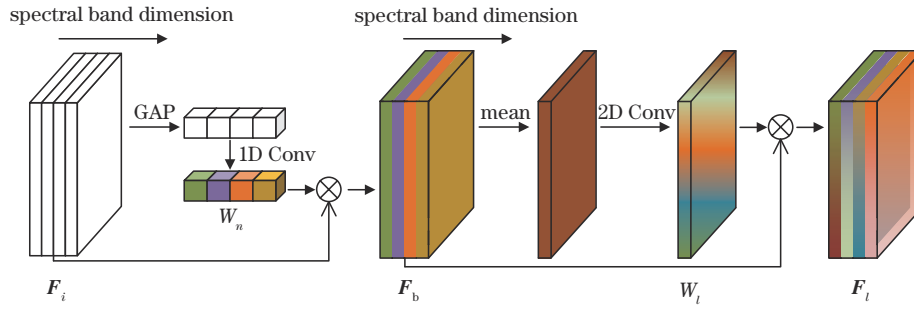


图 2 波段-位置自适应选择机制示意图

Fig. 2 Diagram of band-location adaptive selection mechanism

征图 F_b 。

$$F_b = \sum_{h=0}^{H-1} \sum_{w=0}^{W-1} F_{l(:,h,w)} \odot W_{n(l)}, \quad (3)$$

式中: \odot 表示哈达玛积运算。

在实际应用中,MSIs 存在地面物体分布复杂,形状大小差异较大等特点,能否捕获有效的空间信息对于语义分割任务十分重要。为此,进一步对特征图 F_b 在光谱维上做均值运算,得到保留全波段信息的空间特征图,并通过二维卷积学习得到能够准确反映 MSIs 地物信息的位置权重图 $W_l \in \mathbf{R}^{1 \times H \times W}$ 。然后将位置权重图 W_l 与 F_b 在空间维上进行逐元素相乘运算,得到包含空间关系的波段-位置特征图 F_l 。计算方式如下:

$$W_l = \sigma \left\{ \text{Conv}2d^{3 \times 3} \left[\varpi(F_b) \right] \right\}, \quad (4)$$

$$\varpi(F_b) = \frac{1}{N} \sum_{n=0}^{N-1} F_{b(n,:,:)}, \quad (5)$$

$$F_l = \sum_{n=0}^{N-1} F_{b(n,:,:) \odot W_{l(l,:)}}, \quad (6)$$

式中: $\text{Conv}2d^{3 \times 3}(\bullet)$ 表示卷积核大小为 3×3 的二维卷积操作; $\varpi(\bullet)$ 表示在光谱维上进行均值运算。

3.2 三维残差块

ResNet^[38] 的残差思想缓解了令人苦恼的网络退化问题,在特征提取方面表现出卓越的性能,成为广泛使用的主干网络之一。MSIs 含有多个光谱波段,能否从 MSIs 中提取到鉴别性的光谱-空间特征,直接影响语义分割结果的质量。原始 ResNet 的残差块使用二维卷积来提取图像特征,没有考虑光谱波段间的相关性。三维卷积能够在光谱维和空间维同时进行卷积运算,在特征提取方面优于二维卷积。基于 ResNet 提出的 3DRB 如图 3 所示。与原始残差块相比,3DRB 能够在光谱域和空间域同时编码 MSIs 上下文信息。

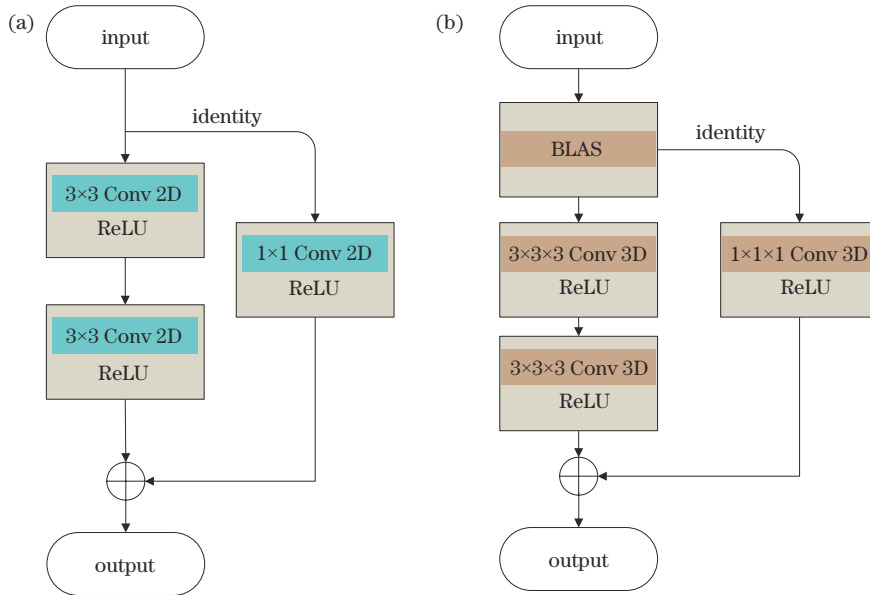


图 3 残差块。(a)原始残差块;(b)所提三维残差块

Fig. 3 Residual blocks. (a) Original residual block; (b) proposed 3D residual block

原始 ResNet 从输入图像到输出特征图进行 32 倍下采样,导致大量的像素信息被丢弃,这使得在解码阶段恢复图像细节变得更加困难。为此,在编码阶段采

用 9 个 3DRB 对输入图像进行 16 倍下采样,相较于原始 ResNet,输出特征图的分辨率大小能够提升 2 倍,有效编码 MSIs 语义信息的同时能够保留更多的细节信

息。ResNet通过恒等映射将学习到的残差与原始特征进行相加,以确保网络不会随着层数加深而产生梯度爆炸/消失的问题。但并非所有的特征都有益于目标任务,因此在每个3DRB中,加入BLAS层引导原始特征图去除冗余特征,增强有效特征表达。对于每个3DRB,计算过程表述如下:

$$\mathbf{x}_{l+1} = B(\mathbf{x}_l) + R[B(\mathbf{x}_l), \mathbf{u}_l], \quad (7)$$

式中: \mathbf{x}_l 和 \mathbf{x}_{l+1} 分别表示第 l 个3DRB的输入和输出; $B(\cdot)$ 表示BLAS层; $R(\cdot)$ 表示残差函数; \mathbf{u}_l 是第 l 个3DRB的权重矩阵。

3.3 自适应特征融合模块

低级特征包含大量的形状、边界等细节信息,高级特征包含更多能够对物体类别进行分类的语义信息。在解码阶段,采用反卷积来恢复特征图空间分辨率。仅用高级特征恢复的特征图过于粗糙且丢失了很多细节信息,模型性能将会受到很大限制。在上采样过程中,融入低级特征能够使恢复的特征图更加精细。低级细节特征与高级语义特征的融合有助于提高图像分割的准确性,但很难确定不同级特征对语义分割任务的贡献程度,即很难确定融合比例。因此,在ADF模块中引入两个可训练参数矩阵 \mathbf{W}_α 、 \mathbf{W}_β ,自适应调整不同级特征的融合比例。 \mathbf{W}_α 和 \mathbf{W}_β 的定义如下:

$$\mathbf{W}_\alpha = \begin{bmatrix} \alpha_{1,1,1} \cdots \alpha_{1,w,1} & \cdots & \alpha_{1,1,n} \cdots \alpha_{1,w,n} \\ \vdots & \vdots & \vdots \\ \alpha_{h,1,1} \cdots \alpha_{h,w,1} & \cdots & \alpha_{h,1,n} \cdots \alpha_{h,w,n} \end{bmatrix}, \quad (8)$$

$$\mathbf{W}_\beta = \begin{bmatrix} \beta_{1,1,1} \cdots \beta_{1,w,1} & \cdots & \beta_{1,1,n} \cdots \beta_{1,w,n} \\ \vdots & \vdots & \vdots \\ \beta_{h,1,1} \cdots \beta_{h,w,1} & \cdots & \beta_{h,1,n} \cdots \beta_{h,w,n} \end{bmatrix}, \quad (9)$$

式中: $\alpha_{h,w,n}$ 和 $\beta_{h,w,n}$ 分别表示矩阵 \mathbf{W}_α 、 \mathbf{W}_β 位置 (h, w, n) 处的参数值。 \mathbf{W}_α 和 \mathbf{W}_β 分别控制低级特征与高级特征的融合权重。低级特征图 \mathbf{l}_{low} 和高级特征图 \mathbf{l}_{high} 的定义如下:

$$\mathbf{l}_{low} = \begin{bmatrix} l_{1,1,1} \cdots l_{1,w,1} & \cdots & l_{1,1,n} \cdots l_{1,w,n} \\ \vdots & \vdots & \vdots \\ l_{h,1,1} \cdots l_{h,w,1} & \cdots & l_{h,1,n} \cdots l_{h,w,n} \end{bmatrix}, \quad (10)$$

$$\mathbf{l}_{high} = \begin{bmatrix} l_{1,1,1} \cdots l_{1,w,1} & \cdots & l_{1,1,n} \cdots l_{1,w,n} \\ \vdots & \vdots & \vdots \\ l_{h,1,1} \cdots l_{h,w,1} & \cdots & l_{h,1,n} \cdots l_{h,w,n} \end{bmatrix}, \quad (11)$$

式中: $l_{h,w,n}$ 表示特征图 l 位置 (h, w, n) 处的特征值。上采样后的 \mathbf{l}_{high} 和 \mathbf{l}_{low} 保持相同大小。有多种特征融合策略,但不同的融合策略在具体目标任务中的性能增益大不相同。为此,探究了加法、元素乘法和串联等3种融合策略对模型分割性能的影响。加法融合将浅层提取的低级细节特征直接融入高级特征图中,弥补细节信息的丢失。加法融合会增加高级特征图每一维的信息量,从而使解码阶段恢复的特征图更加准确。元素乘法融合类似于依据低级特征分布来增强高级特征。

串联融合将低级特征与高级特征拼接在一起,会造成特征图(通道)数量的增加。不同策略融合后的特征图 \mathbf{U} 表述如下:

$$\mathbf{U}_\oplus = \mathbf{W}_\alpha \mathbf{l}_{low} \oplus \mathbf{W}_\beta \mathbf{l}_{high}, \quad (12)$$

$$\mathbf{U}_\odot = \mathbf{W}_\alpha \mathbf{l}_{low} \odot \mathbf{W}_\beta \mathbf{l}_{high}, \quad (13)$$

$$\mathbf{U}_{[\cdot, \cdot]} = [\mathbf{W}_\alpha \mathbf{l}_{low}, \mathbf{W}_\beta \mathbf{l}_{high}], \quad (14)$$

式中: \mathbf{U}_\oplus 、 \mathbf{U}_\odot 和 $\mathbf{U}_{[\cdot, \cdot]}$ 分别表示加法、元素乘法和串联策略融合后的特征图 \mathbf{U} 。参数矩阵 \mathbf{W}_α 和 \mathbf{W}_β 随着网络学习更新相应位置的参数值。 \mathbf{W}_α 的初始值 α 设置为0.5。 \mathbf{W}_β 的参数值 β 被限制为 $|1 - |\alpha||$ 。通过反向梯度传播算法更新 \mathbf{W}_α 和 \mathbf{W}_β 。 \mathbf{W}_α 、 \mathbf{W}_β 更新规则如下:

$$\bar{\mathbf{W}}_\alpha = \mathbf{W}_\alpha - \eta \Delta_{(\mathbf{W}_\alpha)}^{(L)}, \quad (15)$$

$$\bar{\mathbf{W}}_\beta = \mathbf{I}_{(h,w,n)} - \bar{\mathbf{W}}_\alpha, \quad (16)$$

式中: η 表示学习率; $\Delta_{(\mathbf{W}_\alpha)}^{(L)}$ 表示交叉熵损失函数 L 对 \mathbf{W}_α 的梯度; $\mathbf{I}_{(h,w,n)}$ 表示大小为 (h, w, n) 的单位矩阵。

对于语义分割任务,不同通道(即不同特征图)所表达信息的重要程度是不同的。融合后的特征图 $\mathbf{U} \in \mathbf{R}^{C \times N \times W \times H}$,其中, C 表示通道个数。由于三维卷积相较于二维卷积,生成的特征图多一个光谱维度,无法直接利用现有的通道注意力^[31]。因此,先对特征图 \mathbf{U} 进行逐通道压缩,之后利用一维卷积学习通道权重向量,增强对目标任务有贡献的通道,抑制无效通道:

$$\mathbf{U}_{CAM} = \sigma[G(\mathbf{U}) \cdot \mathbf{W}_u], \quad (17)$$

式中: $G(\cdot)$ 表示逐通道对特征图 \mathbf{U} 进行压缩; \mathbf{W}_u 为权重矩阵; $[\cdot]$ 表示点积操作; \mathbf{U}_{CAM} 为学习到的通道权重向量。

4 实 验

在3个数据集上进行实验来定量评估BLASeNet的有效性:ISPRS Potsdam、Qinghai和Tibet Plateau数据集。

4.1 数据集介绍

ISPRS Potsdam数据集包含38幅图像,每幅图像的尺寸为 6000×6000 ,地面采样距离为5 cm。数据集包含红、绿、蓝和红外波段等4个光谱波段,不透水表面(impervious surfaces)、建筑物(building)、低矮植被(low vegetation)、树木(tree)、汽车(car)和杂物/背景(clutter/background)等6个地物类别。使用编号为2_11、2_13、2_14、3_11、3_13、3_14、4_11、4_13、4_14、4_15、5_11、5_13、5_14、5_15、6_8、6_11、6_13、6_14、6_15、7_8、7_11、7_13的22幅图像作为训练集,其余16幅作为测试集。

Qinghai数据集是由Landsat 8卫星捕获的青海湖地区影像图,地面采样距离为30 m。图像尺寸为 5549×5004 ,包含海岸波段、蓝、绿、红、近红外、短波红外1和短波红外2等7个光谱波段。数据集含有草地(grass)、水体(water)、云(cloud)和裸地(baresoil)等

4 个类别。图像被裁剪成 2530 张大小为 256×256 的图像块,其中 50% 用于训练,其余用于测试。

Tibet Plateau 数据集覆盖青藏高原南部区域,包含 12 幅图像,每幅图像大小为 5529×5022 ,地面采样距离为 30 m,包含海岸波段、蓝、绿、红、近红外、短波红外 1 和短波红外 2 等 7 个光谱波段。数据集包含裸地 (bare ground)、建设用地 (construction land)、冰雪/云 (ice/clouds)、水体 (water)、草地 (grass) 和林木/灌木 (woodland/shrubs) 等 6 个类别。使用编号为 1、3、5、7、9、11 的 6 幅图像作为训练集,其余 6 幅作为测试集。

4.2 评价指标

通过总体精度 (A_{OA})、每类 F1 分数 (S_{F1})、平均 F1 分数 (mean S_{F1}) 和平均交并比 (R_{MIOU}) 来综合评估模型性能,计算公式分别为

$$A_{OA} = \frac{\sum_{c=0}^{C-1} N_{TP_c}}{\sum_{c=0}^{C-1} N_{TP_c} + N_{FP_c} + N_{TN_c} + N_{FN_c}} \quad (18)$$

$$S_{F1} = 2 \cdot \frac{R_{precision} \cdot R_{recall}}{R_{precision} + R_{recall}} \quad (19)$$

$$R_{MIOU} = \frac{1}{N} \sum_{c=0}^{N-1} \frac{N_{TP_c}}{N_{TP_c} + N_{FP_c} + N_{FN_c}} \quad (20)$$

$$\begin{cases} R_{precision} = \frac{N_{TP}}{N_{TP} + N_{FP}} \\ R_{recall} = \frac{N_{TP}}{N_{TP} + N_{FN}} \end{cases}, \quad (21)$$

式中: N_{TP_c} 、 N_{FP_c} 、 N_{TN_c} 和 N_{FN_c} 分别代表 c 类的真阳性、假阳性、真阴性和假阴性的数量; C 表示类别数。

A_{OA} 表示分类正确的像素个数占所有像素个数的比例,但 A_{OA} 不能很好地表示模型分割小目标物体的能力。 S_{F1} 是每类的精确率和召回率的加权调和平均,用来综合评估模型查准和查全的能力。 R_{MIOU} 表示模型对所有类别的预测结果和真实值的交集与并集的比值的平均值,能更好地衡量模型准确分割物体的能力。

4.3 实验设置

模型运行平台为 PyTorch,采用 Adam 优化器,权重衰减设置为 0.0005。初始学习率设置为 0.0003,采用学习率衰减策略,每 30 个 epoch 衰减一次,总共训练 60 个 epoch,批大小设置为 32。

与一些经典的以及最新的深度学习方法进行比较,包括 U-Net^[9]、ResNet34^[37]、MAResU-Net^[36] 和 DeepLabv3+^[21]。ResNet34 采用与 U-Net 相同的解码结构来实现图像分割任务。DeepLabv3+ 的特征提取主干是 ResNet-101。MAResU-Net 采用 ResNet-34 作为特征提取主干。表 1 列出了不同方法的结构对比,其中,主干网络统计的是卷积层和池化层的数量。所有方法都是在配备 Intel Xeon Silver 4214 CPU、128 GB 运行内存和 RTX3090 GPU 的计算机上实现的,且所有方法都基于相同的训练集和测试集进行训练和测试评估模型性能。

表 1 不同方法的结构对比

Table 1 Structure comparison of different methods

Method	Backbone (layer)	Convolution type	Attention	Feature fusion	Band selection
U-Net ^[9]	14 layers	2D Conv		✓	
ResNet34 ^[38]	ResNet-34 (34 layers)	2D Conv		✓	
MAResU-Net ^[37]	ResNet-34 (34 layers)	2D Conv	✓	✓	
DeepLabv3+ ^[22]	ResNet-101 (101 layers)	2D Conv		✓	
BLASeNet-A	20 layers	3D Conv	✓	✓	✓
BLASeNet-M	20 layers	3D Conv	✓	✓	✓
BLASeNet-C	20 layers	3D Conv	✓	✓	✓

4.4 实验结果和分析

表 2 为不同方法在 Potsdam 数据集上的实验结果,图 4 显示了不同方法的分割结果图。ResNet34 的 A_{OA} 相比 U-Net 提高了 0.73 个百分点,说明残差块能够减少卷积过程中的信息损失,提高模型的分割精度。MAResU-Net 在跳跃连接中引入注意力机制,改善了分割结果,表明注意力机制的应用能够有效建模上下文信息,可应对 MSIs 复杂场景分割。BLASeNet 不同融合策略大体上获得了相近的实验结果。总的来看,BLASeNet-A 的结果更优。与 MAResU-Net 相比, A_{OA} 、mean S_{F1} 、 R_{MIOU} 分别提高了 1.33 个百分点、2.02 个百分点、1.80 个百分点,表明 BLASeNet 有能力捕获 MSIs 中富含有效信息的光谱波段,同时能够提取到具有鉴别性

的光谱-空间特征。对于 clutter/background 类, BLASeNet-A 获得了 39.77% 的 S_{F1} ,相较于 U-Net 提高了 12.84 个百分点。进一步探究发现, clutter/background 类的空间分布较为分散且实例尺寸较小(如图 5 所示),证明所提 BLASeNet 能够有效分割小目标。

为了进一步探究所提 BLAS 机制的有效性,对 BLASeNet-A 不同层特征图进行可视化研究。为方便起见,将每个 BLAS 层之前的特征图分别记为 layer_1、layer_2、layer_3 和 layer_4,之后的特征图分别记为 BLAS_1、BLAS_2、BLAS_3 和 BLAS_4(图 1)。为便于观察,将 layer_4/BLAS_4、layer_3/BLAS_3 和 layer_2/BLAS_2 特征图分别上采样 8 倍、4 倍和 2 倍,与原始图像保持相同空间分辨率。可视化结果如图 6

表 2 Potsdam 数据集的分割结果

Table 2 Segmentation results on Potsdam dataset

unit: %

Class name	clutter/ background	impervious surfaces	building	low vegetation	tree	car	A_{OA}	mean S_{F1}	R_{MIoU}
U-Net ^[9]	26.93	81.82	84.81	76.26	74.81	79.72	75.71	70.73	57.68
ResNet34 ^[38]	24.94	82.28	86.90	76.73	74.25	79.63	76.44	70.79	58.07
MAResU-Net ^[37]	32.82	83.43	87.79	78.82	76.96	82.07	78.31	73.45	61.10
DeepLabv3+ ^[22]	38.98	82.56	87.47	77.44	76.53	81.23	78.23	74.03	60.96
BLASeNet-A	39.77	84.16	88.40	79.33	79.50	81.64	79.64	75.47	62.90
BLASeNet-M	38.81	83.96	88.19	79.76	79.46	82.21	79.70	75.40	62.89
BLASeNet-C	38.84	84.17	88.17	79.44	77.92	82.00	79.35	75.09	62.47

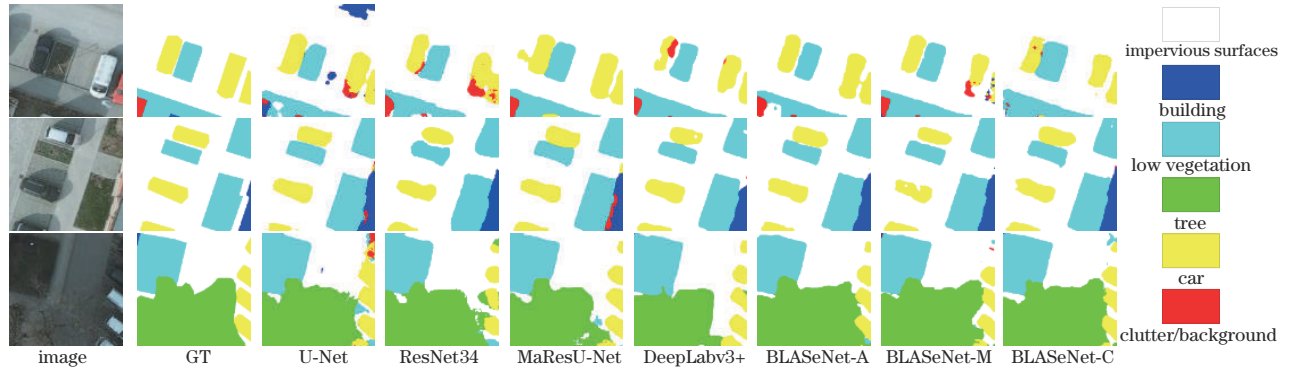


图 4 Potsdam 数据集上不同方法获得的分割结果图

Fig. 4 Segmentation result maps obtained by different methods on the Potsdam dataset

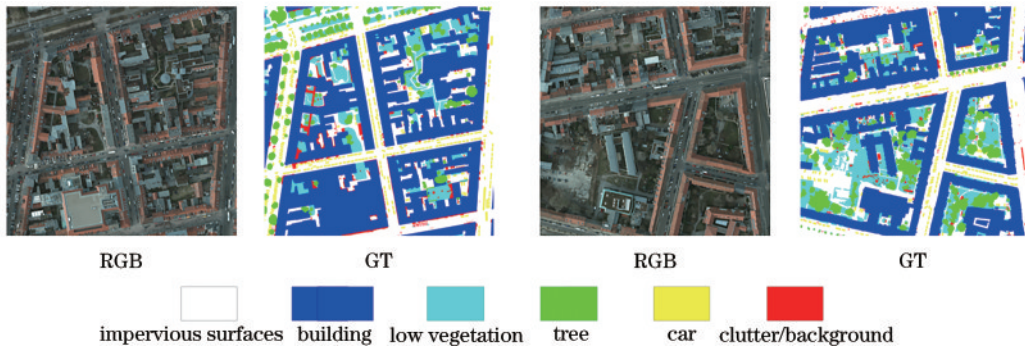


图 5 Potsdam 数据集类别空间分布图

Fig. 5 Category space distribution map of Potsdam dataset

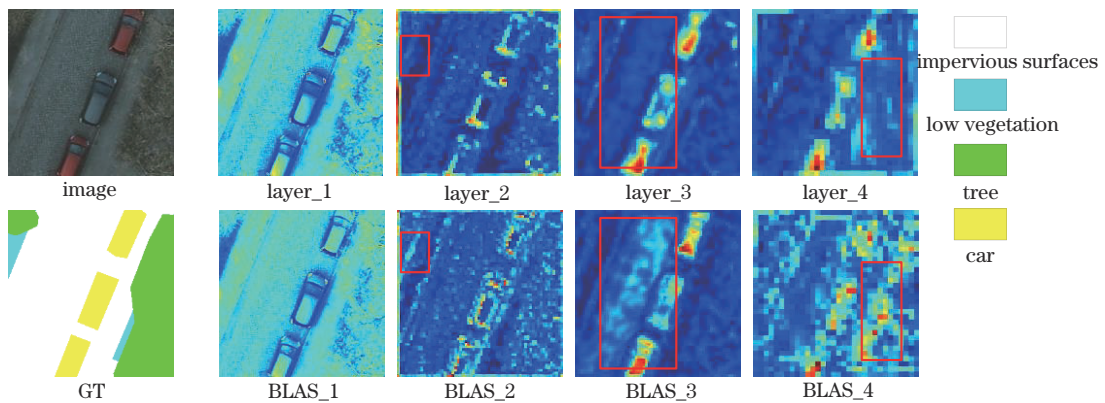


图 6 Potsdam 数据集上不同层特征图的可视化结果

Fig. 6 Visualisation results of different layer feature maps on the Potsdam dataset

所示,可以明显观察到,使用BLAS机制后的特征图边界轮廓更加清晰、完整。BLAS_2能够准确捕获左上角 low vegetation 和 impervious surfaces 的边界特征。由于 car 类特征相较于其他 3 类有着明显的区分度,因此模型能够很容易提取到显著类 car 的特征,但同时很容易丢失相似类或称困难类的特征。例如,layer_3 虽然很好捕获到了 car 类特征,但特征图过于平滑,无法有效区分其他 3 类。BLAS_3 提取到 car 类特征的同时,捕获到了 impervious surfaces 类特征,说明 BLAS 机制通过建模波段-空间位置权重,使模型能够关注于困难类,实现困难类特征的提取。

在 Qinghai 数据集的实验结果如表 3 所示,图 7 展示了不同方法的分割结果图。在 cloud 类上,BLASENet-A

获得了 77.08% 的 S_{F1} ,与第二名相比,提高了 7.53 个百分点,表明 BLASENet 通过同时对光谱域和空间域中的上下文信息进行建模,对于一些分布分散的小目标,也能够较为准确识别。 R_{MIOU} 用来评估模型能否准确预测出目标物体的形状,因此, R_{MIOU} 能更好衡量模型准确分割物体的能力。BLASENet-A 的 R_{MIOU} 分数达到了 77.71%,相较于 MAREsU-Net 提高了 3.42 个百分点。从图 7 的分割结果来看,基于 BLASENet 产生的分割图轮廓更加清晰。图 8 展示了 BLASENet-A 中 ADF 模块的可训练参数矩阵 W_α 和 W_β 的参数值变化情况(用均值表示),可以看出,随着迭代次数的增加,网络更倾向于分配给高级特征图一个更大的权重,说明高级特征图中包含着对于分割任务更加重要的语义信息。

表 3 Qinghai 数据集的分割结果
Table 3 Segmentation results on the Qinghai dataset

Class name	grass	water	cloud	bare soil	A_{OA}	mean S_{F1}	R_{MIOU}
U-Net ^[9]	82.37	96.71	64.91	86.05	86.14	82.51	71.80
ResNet34 ^[38]	82.99	97.41	69.55	86.69	86.89	84.16	73.92
MAREsU-Net ^[37]	84.06	96.91	69.39	87.34	87.47	84.43	74.29
DeepLabv3+ ^[22]	84.09	97.02	67.68	86.98	87.29	83.94	73.72
BLASENet-A	85.37	97.57	77.08	87.89	88.39	86.98	77.71
BLASENet-M	86.18	97.44	73.28	88.17	88.67	86.27	76.85
BLASENet-C	85.89	97.37	72.45	88.35	88.63	86.02	76.52

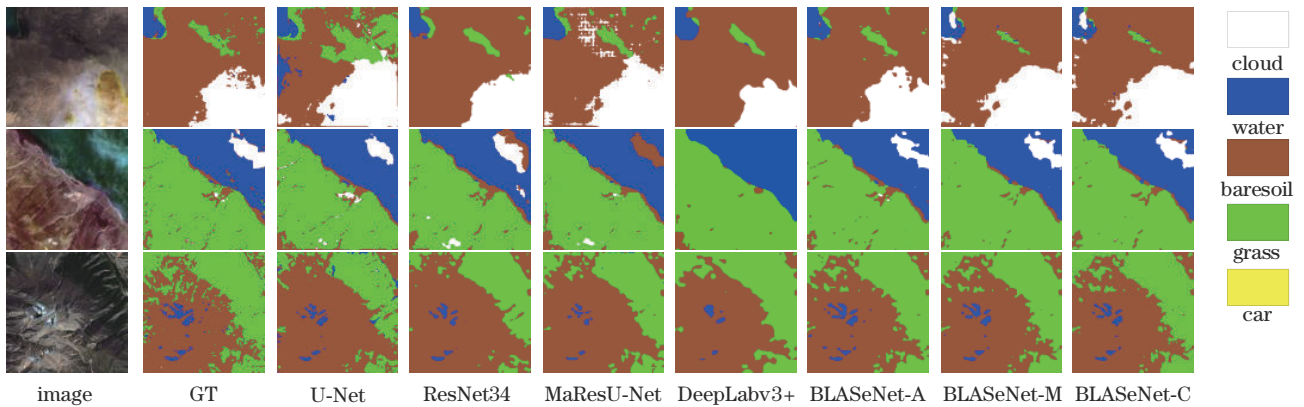


图 7 Qinghai 数据集上不同方法获得的分割结果图

Fig. 7 Segmentation result maps obtained by different methods on the Qinghai dataset

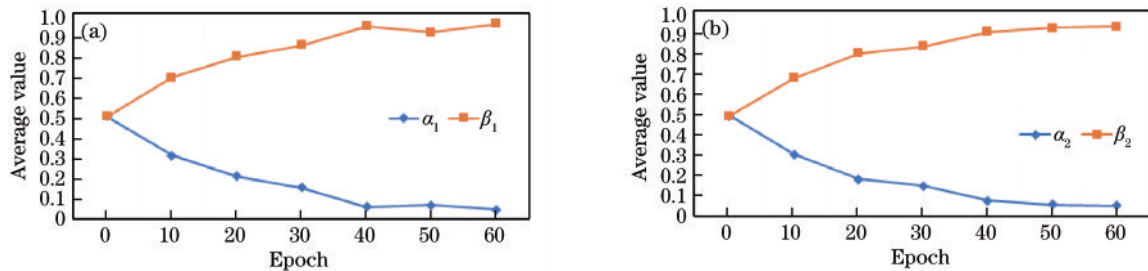


图 8 ADF₁和 ADF₂模块中可训练参数矩阵的参数值变化情况。(a) ADF₁; (b) ADF₂

Fig. 8 Variation of parameter values of the trainable parameter matrix in the ADF₁ and ADF₂ modules. (a) ADF₁; (b) ADF₂

表 4 列出了 Tibet Plateau 数据集的实验结果,不同方法的分割结果如图 9 所示。所有方法中 grass 和

bare ground 类之间都存在着大量的误分,water 类相对而言能准确捕获到。DeepLabv3+ 采用空洞卷积扩大

表 4 Tibet Plateau 数据集的分割结果
Table 4 Segmentation results on the Tibet Plateau dataset

unit: %

Class name	bare ground	construction land	ice/ clouds	water	grass	woodland/ shrubs	A_{OA}	mean S_{F1}	R_{MIoU}
U-Net ^[9]	46.20	16.79	42.79	88.54	78.89	37.97	68.59	51.86	39.07
ResNet34 ^[38]	44.13	15.43	42.19	88.78	78.59	35.59	67.97	50.78	38.26
MAResU-Net ^[37]	42.29	12.88	41.04	89.13	78.50	37.27	68.02	50.19	37.90
DeepLabv3+ ^[22]	46.62	0.00	32.29	85.35	77.96	39.34	67.57	46.93	35.41
BLASeNet-A	44.47	29.10	44.06	90.03	79.07	39.35	68.68	54.31	40.91
BLASeNet-M	49.84	20.62	42.52	88.25	79.36	43.80	69.60	54.06	40.75
BLASeNet-C	49.51	20.50	43.32	88.53	78.79	36.28	68.92	52.82	39.76

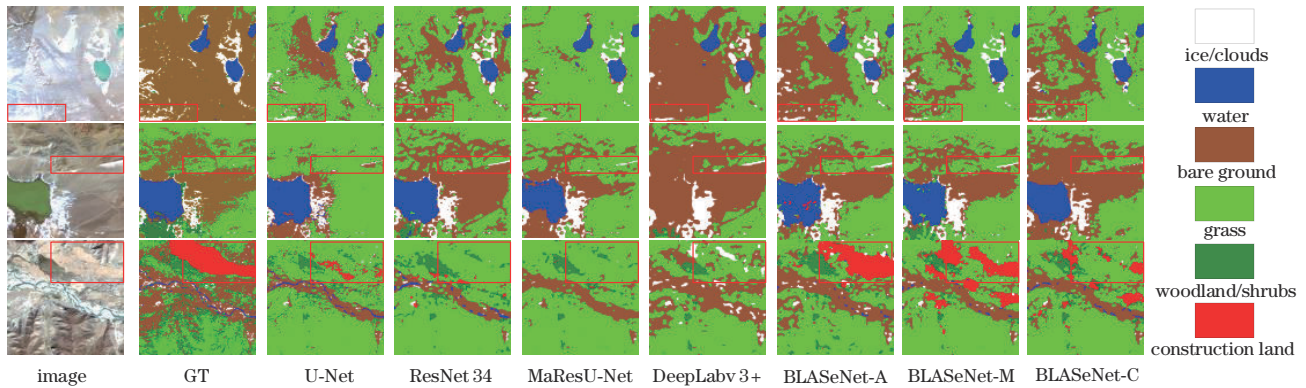


图 9 Tibet Plateau 数据集上不同方法获得的分割结果图

Fig. 9 Segmentation result maps obtained by different methods on the Tibet Plateau dataset

感受野的同时丢失了很多细节信息,导致在小目标 construction land 类上获得了 0 分的 S_{F1} 。基于 BLASeNet 的方法通过建模波段-空间位置权重,很好地保留了细节特征。具体来讲,产生的分割图更加精细,在 grass 与 construction land 类之间的辨别力得到了很大改善,在保留对象边界方面也有着明显改进。特征提取网络较浅的 U-Net 出乎意料地获得了较好的分割结果,这可能是由于数据集本身存在着类别不均衡问题,大类

和小类之间样本数量差异巨大,有着较深特征提取网络的模型反而得到了较差的结果。

4.5 消融实验

在 Potsdam 数据集上进行消融实验,以 BLASeNet-A 为例,探究每个所提模块带来的性能增益。表 5 列出了依次增加不同模块后的实验结果及测试时间。使用 ResNet34 作为基线模型,将 3DRB、BLAS、ADF 和 CAM 模块相继加入模型中。

表 5 Potsdam 数据集上的消融实验

Table 5 Ablation experiments on the Potsdam dataset

Class name	clutter/ background	impervious surfaces	building	low vegetation	tree	car	A_{OA}	mean S_{F1}	R_{MIoU}	Test time / min
Baseline	24.94	82.28	86.90	76.73	74.25	79.63	76.44	70.79	58.07	2.05
+3DRB	35.47	83.07	87.36	77.87	77.56	81.83	78.26	73.86	61.08	2.74
+BLAS	36.72	84.27	88.50	78.58	79.04	82.18	79.29	74.88	62.42	2.75
+ADF	39.32	84.04	88.27	79.29	78.94	82.29	79.49	75.36	62.79	2.75
+CAM	39.77	84.16	88.40	79.33	79.50	81.64	79.64	75.47	62.90	2.76

总的来看,每个模块都在逐步提高 BLASeNet 的模型性能。实验结果表明,对于极难捕获的 clutter/background 类,所提模块都在稳步提高这一类的分割结果。相较于基线模型,clutter/background 类的分割结果提升了 14.83 个百分点,证明 BLASeNet 有能力捕获分布分散的小目标。进一步探究发现,加入 3DRB 后,clutter/background 类的分割精度提升了

10.53 个百分点,说明对于在空间上呈分散分布的小目标,3DRB 能够准确提取到其有鉴别性的光谱-空间特征。加入 BLAS 之后, A_{OA} 、mean S_{F1} 、 R_{MIoU} 都有较大的涨幅,分别提升了 1.03 个百分点、1.02 个百分点、1.34 个百分点,表明 BLAS 能够捕获 MSIs 有效光谱波段的同时建模空间位置权重,使模型更加关注有效区域信息。对于其他类,BLASeNet 在不同程度上都

提高了分割精度,证明了BLASeNet的有效性。

从模型测试时间上来看,增加3DRB后,测试时间相较于基线模型有大幅上涨,增加了0.69 min。由于在光谱域增加了局部卷积,三维卷积带来性能增益的同时,不可避免地增加了模型计算时间。特别地,引入BLAS后,为模型带来性能增益的同时几乎没有增加模型计算时间,证明了BLAS的有效性和轻量化。另外,加入ADF模块之后, A_{OA} 、 $\text{mean } S_{F1}$ 、 R_{MIoU} 分数分别提升了0.2个百分点、0.48个百分点、0.37个百分点,并且没有增加额外的计算时间,表明调整不同级特征的融合比例是必要且有效的。

5 结 论

提出一个新颖的MSIs语义分割模型。所提BLASeNet能够有效应对含有多个光谱波段的MSIs复杂场景分割。在Potsdam、Qinghai和Tibet Plateau数据集上的实验结果证明了BLASeNet的优势:其可以捕捉到MSIs富含信息的光谱波段,有助于增强语义分割结果,并且可以更好地提取到在空间上呈分散分布的小目标特征,通过建模波段-空间位置权重,生成的分割图边界轮廓更加清晰。进一步探究加法、元素乘法和串联等3种特征融合方式发现,加法融合倾向于获得更好的分割性能。

另外,MSIs不同地物对不同光谱波段的敏感性不同,BLASeNet首先依据所有地物信息自适应地为每个波段学习一个权重,之后建模空间位置权重,从而隐式地根据不同地物类型为每个波段不同空间位置学习不同权重。这样做可能会导致建模的波段-空间位置权重不能充分反映所有地物类型的真实分布。因此,如何根据不同地物类型直接为每个波段不同空间位置学习到不同权重是下一步的研究方向。

参 考 文 献

- [1] Wang Q, He X, Li X L. Locality and structure regularized low rank representation for hyperspectral image classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57(2): 911-923.
- [2] Tobias O J, Seara R. Image segmentation by histogram thresholding using fuzzy sets[J]. IEEE Transactions on Image Processing, 2002, 11(12): 1457-1465.
- [3] Dhanachandra N, Mangleam K, Chanu Y J. Image segmentation using K-means clustering algorithm and subtractive clustering algorithm[J]. Procedia Computer Science, 2015, 54: 764-771.
- [4] 刘金香, 班伟, 陈宇, 等. 融合多维度CNN的高光谱遥感图像分类算法[J]. 中国激光, 2021, 48(16): 1610003. Liu J X, Ban W, Chen Y, et al. Multi-dimensional CNN fused algorithm for hyperspectral remote sensing image classification[J]. Chinese Journal of Lasers, 2021, 48(16): 1610003.
- [5] 张世宽, 吴清潇, 林智远. 焊缝图像中结构光条纹的检

测与分割[J]. 光学学报, 2021, 41(5): 0515002.

Zhang S K, Wu Q X, Lin Z Y. Detection and segmentation of structured light stripe in weld image[J]. Acta Optica Sinica, 2021, 41(5): 0515002.

- [6] 袁媛, 陈明惠, 柯舒婷, 等. 基于集成卷积神经网络和Vit的眼底图像分类研究[J]. 中国激光, 2022, 49(20): 2007205. Yuan Y, Chen M H, Ke S T, et al. Fundus image classification research based on ensemble convolutional neural network and vision transformer[J]. Chinese Journal of Lasers, 2022, 49(20): 2007205.
- [7] 罗禹杰, 张剑, 陈亮, 等. 基于自适应空间特征融合的轻量化目标检测算法[J]. 激光与光电子学进展, 2022, 59(4): 0415004. Luo Y J, Zhang J, Chen L, et al. Lightweight target detection algorithm based on adaptive spatial feature fusion[J]. Laser & Optoelectronics Progress, 2022, 59(4): 0415004.
- [8] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA, USA. New York: IEEE, 2015: 3431-3440.
- [9] Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation [M]//Navab N, Hornegger J, Wells W M, et al. Medical image computing and computer-assisted intervention-MICCAI 2015. Lecture notes in computer science. Cham: Springer, 2015, 9351: 234-241.
- [10] Badrinarayanan V, Kendall A, Cipolla R. SegNet: a deep convolutional encoder-decoder architecture for image segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(12): 2481-2495.
- [11] Lin G S, Milan A, Shen C H, et al. RefineNet: multi-path refinement networks for high-resolution semantic segmentation[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 5168-5177.
- [12] Noh H, Hong S, Han B. Learning deconvolution network for semantic segmentation[C]//2015 IEEE International Conference on Computer Vision, December 7-13, 2015, Santiago, Chile. New York: IEEE Press, 2015: 1520-1528.
- [13] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]//2017 IEEE International Conference on Computer Vision, October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 2999-3007.
- [14] Sudre C H, Li W Q, Vercauteren T, et al. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations[M]//Cardoso M J, Arbel T, Carneiro G, et al. Deep learning in medical image analysis and multimodal learning for clinical decision support. Lecture notes in computer science. Cham: Springer, 2017, 10553: 240-248.
- [15] Hao S J, Zhou Y, Guo Y R. A brief survey on semantic

- segmentation with deep learning[J]. *Neurocomputing*, 2020, 406: 302-321.
- [16] Li R, Zheng S Y, Zhang C, et al. Multiattention network for semantic segmentation of fine-resolution remote sensing images[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, 60: 5607713.
- [17] Zhao H S, Shi J P, Qi X J, et al. Pyramid scene parsing network[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 6230-6239.
- [18] Yu B, Yang L, Chen F. Semantic segmentation for high spatial resolution remote sensing images based on convolution neural network and pyramid pooling module [J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2018, 11(9): 3252-3261.
- [19] Ding L, Zhang J, Bruzzone L. Semantic segmentation of large-size VHR remote sensing images using a two-stage multiscale training architecture[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2020, 58(8): 5367-5376.
- [20] Mou L C, Hua Y S, Zhu X X. Relation matters: relational context-aware fully convolutional network for semantic segmentation of high-resolution aerial images[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2020, 58(11): 7557-7569.
- [21] Chen L C, Papandreou G, Kokkinos I, et al. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, 40(4): 834-848.
- [22] Chen L C, Zhu Y K, Papandreou G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[M]//Ferrari V, Hebert M, Sminchisescu C, et al. *Computer vision-ECCV 2018. Lecture notes in computer science*. Cham: Springer, 2018, 11211: 833-851.
- [23] Li G, Li L L, Zhu H, et al. Adaptive multiscale deep fusion residual network for remote sensing image classification[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2019, 57(11): 8506-8521.
- [24] Mi L, Chen Z Z. Superpixel-enhanced deep neural forest for remote sensing image semantic segmentation[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2020, 159: 140-152.
- [25] Roy S K, Krishna G, Dubey S R, et al. HybridSN: exploring 3-D-2-D CNN feature hierarchy for hyperspectral image classification[J]. *IEEE Geoscience and Remote Sensing Letters*, 2020, 17(2): 277-281.
- [26] Chen Y S, Jiang H L, Li C Y, et al. Deep feature extraction and classification of hyperspectral images based on convolutional neural networks[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2016, 54(10): 6232-6251.
- [27] He M Y, Li B, Chen H H. Multi-scale 3D deep convolutional neural network for hyperspectral image classification[C]//2017 IEEE International Conference on Image Processing, September 17-20, 2017, Beijing, China. New York: IEEE Press, 2017: 3904-3908.
- [28] Liu R, Mi L, Chen Z Z. AFNet: adaptive fusion network for remote sensing image semantic segmentation[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2021, 59(9): 7871-7886.
- [29] Peng C L, Zhang K N, Ma Y, et al. Cross fusion net: a fast semantic segmentation network for small-scale semantic information capturing in aerial scenes[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, 60: 5601313.
- [30] Guo M H, Xu T X, Liu J J, et al. Attention mechanisms in computer vision: a survey[J]. *Computational Visual Media*, 2022, 8(3): 331-368.
- [31] Hu J, Shen L, Sun G. Squeeze-and-excitation networks [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 7132-7141.
- [32] Wang Q L, Wu B G, Zhu P F, et al. ECA-net: efficient channel attention for deep convolutional neural networks [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 11531-11539.
- [33] Qin Z Q, Zhang P Y, Wu F, et al. FcaNet: frequency channel attention networks[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV), October 10-17, 2021, Montreal, QC, Canada. New York: IEEE Press, 2021: 763-772.
- [34] Woo S, Park J, Lee J Y, et al. CBAM: convolutional block attention module[M]//Ferrari V, Hebert M, Sminchisescu C, et al. *Computer vision-ECCV 2018. Lecture notes in computer science*. Cham: Springer, 2018, 11211: 3-19.
- [35] Li H F, Qiu K J, Chen L, et al. SCAAttNet: semantic segmentation network with spatial and channel attention mechanism for high-resolution remote sensing images[J]. *IEEE Geoscience and Remote Sensing Letters*, 2021, 18(5): 905-909.
- [36] Zhao Q, Liu J H, Li Y W, et al. Semantic segmentation with attention mechanism for remote sensing images[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2021, 60: 1-13.
- [37] Li R, Zheng S Y, Duan C X, et al. Multistage attention ResU-Net for semantic segmentation of fine-resolution remote sensing images[J]. *IEEE Geoscience and Remote Sensing Letters*, 2022, 19: 1-5.
- [38] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 770-778.