基于多尺度语义编解码网络的遥感图像语义分割

梁 燕[1,2],易春霞[1,2],王光宇[1,3],胡跃辉[1,2]
(1. 重庆邮电大学通信与信息工程学院,重庆 400065;2. 信号与信息处理重庆市重点实验室,重庆 400065;
3. 移动通信教育部工程研究中心,重庆 400065)

摘 要: 针对遥感图像语义分割中存在的多层次信息提取和多尺度特征图上下文依赖性两个问题,本文分析现有处理方案,提出了一种综合运用多项技术的多尺度语义编解码网络(Multi-scale Semantic Encoder-Decoder Networks,MSEDNet). MSEDNet由编码与解码两部分构成. 编码阶段,首先提出残差协同空间注意(Residuals Coordinate Spatial Attention,RCSA)的MobileNetV3增强型模块,提取语义信息;其次,设计多层增强语义上下文模块(Enhance Semantic Context Module,ESCM),提升多尺度结构特征图的表征能力. 解码阶段,首先提出多核卷积与Focus并行的强化空间细节信息模块(Strengthen Spatial Detail Information Module,SSDIM),增强浅层特征细节和结构信息;其次,设计了三元迭代多尺度特征融合(Triplet Iterative Multi-Scale Feature Fusion,TIMSFF)策略,强化图像深层全局语义信息与浅层局部细节特征的多尺度融合,提升分割精度. 所提模型在ISPRS Vaihingen和Potsdam数据集上验证,总体分割精度(Overall Accuracy,OA)分别达到95.699%、95.534%,平均F_1-score(mean F_1-score,mF_1)分别提高2.661%和2.929%,且平均交并比(mean Intersection over Union,mIoU)分别增长3.973%和4.012%. 所耗参数量Param下降至6.77 M.

关键词: 遥感语义分割;多尺度语义上下文;注意力机制;空间细节;多尺度特征融合
基金项目: 国家自然科学基金(No.61702066);重庆市教委科学技术重点研究项目(No.KJZD-M201900601)
中图分类号: TP183 **文献标识码:** A **文章编号:** 0372-2112(2023)11-3199-16
电子学报URL:http://www.ejournal.org.cn **DOI:**10.12263/DZXB.20220503

Semantic Segmentation of Remote Sensing Image Based on Multi-Scale Semantic Encoder-Decoder Network

LIANG Yan[1,2], YI Chun-xia[1,2], WANG Guang-yu[1,3], HU Yue-hui[1,2]
(1. *School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing* 400065, *China*;
2. *Chongqing Key Laboratory of Signal and Information Processing, Chongqing* 400065, *China*;
3. *Engineering Research Center of Mobile Communications of the Ministry of Education, Chongqing* 400065, *China*)

Abstract: This paper analyzes the existed processing scheme, and proposes a multi-scale semantic encoder-decoder networks (MSEDNet) by comprehensively using multiple technologies for the problems in remote sensing image semantic segmentation both multi-level information extraction and multi-scale feature diagram dependence characteristic. The MSEDNet consists of two parts: encoding part and decoding part. In the encoding part, the enhanced MobileNetV3 with residuals coordinate spatial attention (RCSA) is firstly proposed to extract semantic information, and then a multi-layer enhanced semantic context module (ESCM) is designed to improve representation ability of the multi-scale structure feature map. In the decoding part, a strengthen spatial detail information module (SSDIM) based on Multi-core Convolution and Focus Parallel is proposed to enhance the details and structural information of shallow features. Then triplet iterative multi-scale feature fusion (TIMSFF) strategy is designed to strengthen the multi-scale context fusion both deep global semantic information and shallow local detail features, for improving the segmentation accuracy. The proposed model has been experimentally verified on the ISPRS Vaihingen and Potsdam dataset. The overall segmentation accuracy (OA) reached 95.699% and 95.534% respectively, the mean F_1-score (mF_1) increased by 2.661% and 2.929% respectively, and the mean intersection over union (mIoU) increased by 3.973%and 4.012%, respectively. The number of param dropped to 6.77 M.

Key words: remote sensing semantic segmentation; multi-scale semantic context; attention mechanism; spatial

收稿日期:2022-05-05;修回日期:2022-11-24;责任编辑:朱梅玉

detail; multi-scale feature fusion

Foundation Item(s): National Natural Science Foundation of China (No.61702066); Key Research Project of Science and Technology of Chongqing Education Commission (No.KJZD-M201900601)

1 引言

近年来,随着科学技术的发展,大量搭载高分辨率影像获取设备的卫星被发射并投入使用,由此产生了海量高分辨率遥感图像.在军事领域,遥感图像主要用于敌方目标定位和识别,以提升军事打击精准率.在民用领域,遥感图像主要在土地覆盖制图、交通监测、智能农业等方面发挥作用,特别是在无人驾驶等领域.因此,如何对遥感图像进行语义分割,实现像素级分类和高级语义特征信息提取,是该领域当前研究的热点内容之一.

遥感语义分割将遥感图像像素按照表达语义含义的不同进行分组分割,理解分析图像中蕴含的丰富的理信息.传统图像语义分割方案包括提取纹元森林特征(Semantic Texton Forest,STF)的随机森林分类器[1]方法、基于支持向量机[2](Support Vector Machine,SVM)的二分类方法、基于非监督聚类引导语义分割算法[3]以及基于道路边缘检测方法[4]等.这些方案通常采用整体分割法,从图像的颜色、形状和纹理中提取浅层信息,并在深层空间进行分类[1~4].遥感图像包含丰富光谱信息及纹理信息且地物结构多样性,存在多层次、多尺度、多类别特征提取问题.这些分割方法严重依赖图像特征的质量,不能取得很好的分割效果.

随着深度学习理论和技术的蓬勃发展,深度卷积神经网络因其在特征提取方面的独特优势,在遥感语义分割领域也得到了广泛应用.现有研究主要从三个方面开展,如图1所示.

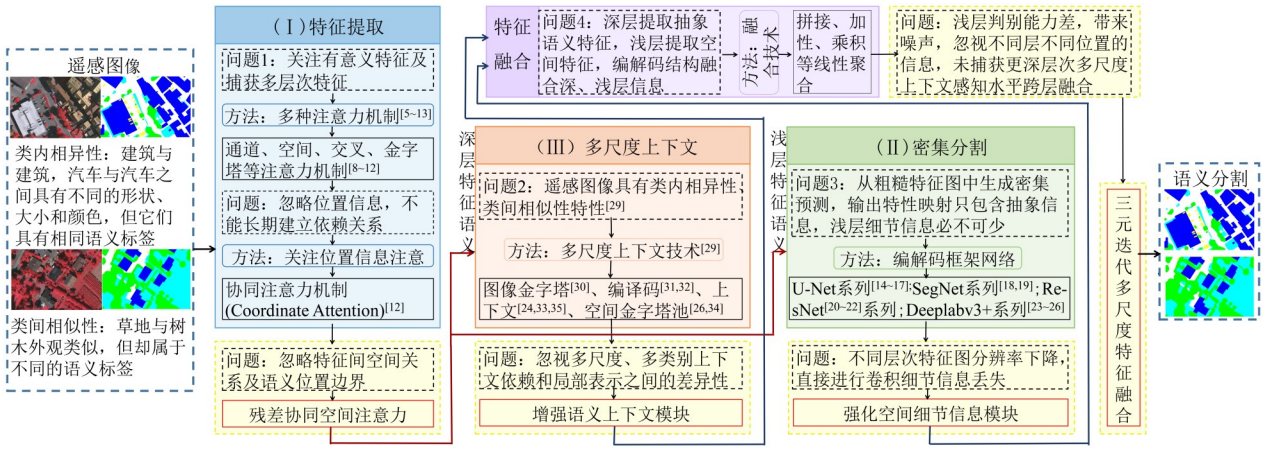


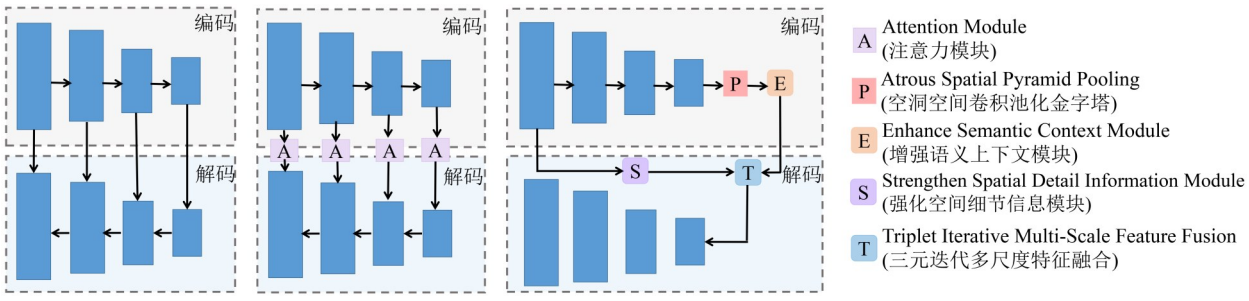
图1 遥感图像语义分割方法

(1)部分研究者在特征提取方面关注有意义特征及捕获多层次特征的问题,采用多种注意力机制.如PANNet(Pixel Aggregation Network)[5]在多尺度模块中引入特征金字塔注意模块,在深层次输出上执行空间金字塔注意结构;MANet(MultiAttention Network)[6]提出核注意机制,采用多尺度策略分层聚合相关上下文特征;CCANet(Class-constraint Coarse-to-fine Attentional deep Network)[7]提出类别约束的粗-精注意深度网络,使类别信息约束的形成获得长范围上下文信息;EAP-Net(Efficient Attention Pyramid Network)[8]提出高效注意金字塔网络,有效地处理多尺度目标分割问题,并利用残差注意融合块处理底层特征等.当前,基于注意力机制的分割方法在位置和形式上存在差异,在实践中表现也各有千秋.在处理多尺度对象分割问题上,更多的注意力机制,如通道注意力[9]、空间注意力[10]、交叉注意力[11]、金字塔注意力[8]、协同注意力机制[12]等应用更加广泛.这些方法提高了网络的特征提取能力,但对特征间的空间关系及语义边界信息处理不足.

(2)部分研究者关注对目标轮廓、形状、面积和空间特征等细节的密集预测,为解决浅层特征图信息缺失问题,提出了许多基于全卷积网络FCN(Fully Convolutional Networks)[13]的编解码网络(图2(a)).如U-Net系列[14~17]、SegNet系列[18,19]、ResNet(Residual Network)系列[20~22]、Deeplabv3+系列[23~26]等,都采用编码、解码结构聚合深层抽象语义特征和浅层空间特征.在文献[15~17]中,U-Net中的普通跳过连接被更微妙和精细地跳过连接取代,减少了编码器和解码器之间的语义差距.如文献[17]提出的MACU-Net(Multiscale connected and Asymmetric-Convolution-based U-Net)设计了一个多尺度跳跃连接和非对称卷积的U-Net网络结构,用于提取

不同层次的特征及捕获细化特征. 在图2(*a*)方法基础上,有研究者提出如图2(*b*)所示的方案,利用各种注意力机制强化编解码结构分割时不同层次特征细节信息和上下文信息. 如MAResU-Net(Multistage Attention ResU-Net)[27]提出的线性注意机制用于重构UNet中的跳跃连接,细化提取的特征图;文献[26]提出SBANet(Semantic Boundary Awareness Network)网络,其在Deeplabv3+基础上引入了边界注意模块,从层次特征聚合中自下而上地获取土地覆盖边界信息;文献[28]提出了一个端到端基于注意的语义分割网络SSAtNet(Semantic Segmentation based on batch-Attention Network),采用金字塔注意池模型,将注意机制引入多尺度模型中进行自适应特征细化. 但图2(*b*)方案中编码器生成的浅级别和细粒度详细特征映射与解码器生成的深级别和粗粒度语义信息融合在一起,没有任何进一步的细化,容易导致特征的利用不足和识别不足[29]. 而且其忽视了语义分割网络中不同层和不同位置的特征信息,未捕获到更深层次的多尺度上下文感知水平跨层融合的语义特性. 因此,不能直接对不同层和不同位置的特征图进行拼接、加性、乘积等线性聚合操作.

(3)遥感图像本身具有类内方差较高、类间方差较低的特点,因此,同一类别对象会出现不同尺度特性(即类内相似性),不同类目标也存在相似特性(即类间相似性)[28]. 当面临多尺度及多类别对象时,多层次特征图只包含相关度较低的上下文依赖关系,融合时不具备多尺度特征聚合的引导能力. 为解决这一问题,需要获取多尺度上下文信息来应对复杂场景. 多尺度技术[28]方法可分为图像金字塔(Image Pyramid,IP)[30]、编译码器框架(Encoder-Decoder Architecture,EDA)[31,32]、上下文模块(Context Module,CM)[33]和空间金字塔池(Spatial Pyramid Pooling,SPP)[34]四大类. IP方法以不同比例图像输出捕获不同范围上下文,此方法消耗大量内存,也会丢失信息. EDA方法包含编、解码两部分:编码捕获多层次语义特征;解码将特征图尽可能恢复和细化. 文献[26]提出空洞空间卷积池化金字塔(Atrous Spatial Pyramid Pooling,ASPP),丰富上下文信息. 由于遥感图像中含有复杂的地物信息,文献[35]提出层次上下文聚合网络HCANet,在提取多语义特征的多尺度上下文信息的同时,也强聚合了上下文信息的路径;文献[24]提出多路径残差网络MP-ResNet(MultiPath Residual Network)结构,通过并行的多尺度分支学习语义上下文,在解码器中采用了多级特征融合设计,有效利用了从不同分支学习到的特征. 虽然上述多种策略捕获上下文信息有利于在不同尺度上表征对象,但忽视了多尺度、多类别对象对上下文的依赖性和局部表示之间的差异性,使整个区域的上下文依赖是同质且非自适应的[29]. 此外,这些多尺度上下文融合策略都是手工设计的,在建模多尺度上下文表示时灵活性有限. 因此,特征图的长期依赖性没有得到充分利用,这对遥感图像语义分割是否能精确感知可能存在至关重要的影响.



(*a*) 编解码　(*b*) 注意力机制与编解码　(*c*) MSEDNet(Ours)

图2　编解码结构对比

针对遥感语义分析中多层次、多尺度、多类别特征提取问题,单一策略效果不足,需要综合运用多种自适应处理方法才可能有效解决. 本文从提取各个层次精细特征和丰富多类别多尺度上下文两个关键点出发,提出了一种多尺度语义编解码网络(Multi-scale Semantic Encoder-Decoder Networks,MSEDNet). 该网络主要分为编码与解码两个阶段,采用如图2(*c*)所示的模型结构. 具体来说,本文贡献与创新点如下(在图1中用浅黄色虚线框表示).

(1)编码阶段. 针对图1所示(Ⅰ)类语义分割特征提取中注意力机制遗留问题,本文基于MobileNetV3[36]主干,提出残差协同空间注意(Residuals Coordinate Spatial Attention,RCSA)机制,关注特征间的空间关系,增强网络捕获图像特征时对浅层详细特征和深层复杂特征增加语义位置边界的特性;针对图1所示(Ⅲ)类遥感图像的多尺度、多类别对象的自适应上下文信息,在ASPP中设计了多层增强语义上下文模块(Enhance Semantic Context Module,ESCM),提升多尺度结构特征图的表征能力,减少特征图生成过程中信息的丢失.

(2)解码阶段. 针对图1所示(Ⅱ)类浅层特征直接

进行卷积特征图空间位置等大部分细节信息丢失的问题,提出特征图多核卷积与Yolov5[37]中Focus并行的强化空间细节信息模块(Strengthen Spatial Detail Information Module,SSDIM),保留浅层图像细节特征和结构信息;针对图1中(Ⅱ)(Ⅲ)类深层与浅层的多尺度上下文感知水平的跨层融合,设计了一种多层次、多尺度的三元迭代多尺度特征融合(Triplet Iterative Multi-Scale Feature Fusion,TIMSFF)策略,将图像深层全局语义信息与浅层局部细节特征跨层融合,提高精细空间分辨率遥感图像语义分割精度.

2 多尺度语义编解码网络MSEDNet

MSEDNet是针对图1中(Ⅰ)(Ⅱ)(Ⅲ)类问题,将深层、浅层特征信息融合的EDA结构.因遥感影像复杂,随着训练的进行,参数量会持续增长,故本文中使用了轻量级网络MobileNetV3-Small版本作为主干网.其特点是参数少、计算量小、推理时间短,更适用于存储空间和功耗受限的场景.

图3所示为本文提出的MSEDNet结构.该结构以增强型MobileNetV3网络为主干,主要包含2个部分:蓝色虚线框为编码阶段,红色虚线框为解码阶段.编码阶段,原始MobileNetV3中引入SE(Squeeze-and-Excitation)注意力结构.SE压缩每个特征图来建立通道间依赖关系,其忽略空间信息,没有充分提取语义信息.因此,我们利用特征间空间关系,提出残差协同空间注意力(Residuals Coordinate Spatial Attention,RCSA)机制,有效细化详细特性且增强特征间的语义位置边界信息.图像经过主干网获得整体抽象表示的相对高级特征,可有效分割大型对象,但不足以处理尺度多样性的小对象,且忽视了多尺度、多类别对象上下文依赖的自适应性.于是设计了增强语义上下文模块(Enhance Semantic Context Module,ESCM)来增强多尺度结构特征表征能力,平衡多尺度感受野.解码阶段,一方面,网络深层高级特征抽象程度高,但空间信息准确性较低;网络浅层不能充分提取抽象信息,但空间信息保持完整,在分割任务中两者同等重要.于是,提出三元迭代多尺度特征融合(Triplet Iterative Multi-Scale Feature Fusion,TIMSFF)实现选择性和动态型融合不同尺度、不同层次的上下文感知水平跨层特征.另一方面,浅层含盖空间信息及大量小对象详细语义信息,直接卷积会使小对象尺度多样性及空间信息被大对象所覆盖[6],导致融合前大部分浅层细节信息已经丢失.因此设计强化空间细节信息(Strengthen Spatial Detail Information Module,SSDIM)模块,在更细粒度水平上有效提取多尺度空间信息,建立多尺度长期通道依赖性.

3 编码阶段

3.1 基于注意力机制RCSA的Enhance MobileNetV3网络

目前,应用最广泛的注意力机制是SE[9].在SE基础上,CBAM(Convolutional Block Attention Module)[10]通过大尺寸核卷积引入空间信息编码.CCNet(Criss-Cross Network)[11]、AFNet(Adaptive Fusion Network)[38]、ACNet(Attention Complementary Network)[39]分别从远程依赖项和中间特征来捕获上下文信息.PANNet[5]引入金字塔注意,结合全局学习更好的特征表示.这些方法提高了多尺度特征提取能力,但它们对多尺度特征中不同通道之间的相互依赖性关注度较少,且忽略了多尺度对象及不同类别对象的上下文依赖性和局部表示之间的差异性.

3.1.1 RCSA

文献[12]提出协作注意(Coordinate Attention,CA),将位置信息嵌入通道注意,沿一个空间方向捕获远程依赖关系,并沿另一个空间方向保留位置信息.但其在对局部详细信息特征提取方面也存在一定的局限性.从空间角度看,通道注意是全局的,空间注意是局部的.空间分布特征包括实体位置、形状及实体间的空间关系、区域空间结构等,不同通道与空间映射之间的相互依赖性可以有效增强特征映射表示特定语义的能力.因此,针对图1所示(Ⅰ)类问题,本文利用特征间空间关系,结合多级和全局上下文特性进行编码,学习具有区分性的特征,提出RCSA机制关注不同层和不同位置的特征信息,正确聚焦目标对象,有效细化详细特性,语义位置边界信息可以分割外界相似性特征.CA与RCSA示意图分别如图4所示.

RCSA有两个顺序子模块:CA通道模块和特征间空间关系注意模块.该算法先将输入特征经CA后再进入特征间空间关系注意模块,通过乘积进行特征融合,使有意义信息得以被关注,并抑制无用信息,最后,通过残差拼接得到输出特征.子模块特征注意图的初始化依赖卷积核的初始化,可能存在梯度消失、梯度爆炸等问题.因此,RCSA利用残差加性拼接减少了过拟合现象.如若出现最坏的情况,CA注意模块注意图的值接近于零,但是通过残差连接,就可以正常地训练整个网络,自动更新注意图和其他参数.通道模块本质是对每个通道赋予不同权重,空间关系模块反映地理实体空间分布特征的信息.空间分布特征包括实体的位置、形状及实体间的空间关系、区域空间结构等.空间关系注意关注的是特征图局部信息部分,是对通道注意的补充.因此,RCSA模块能够序列化地在通道和空间维度上产生注意力特征图信息,然后再与之前原输入特征图进行拼接实现自适应特征

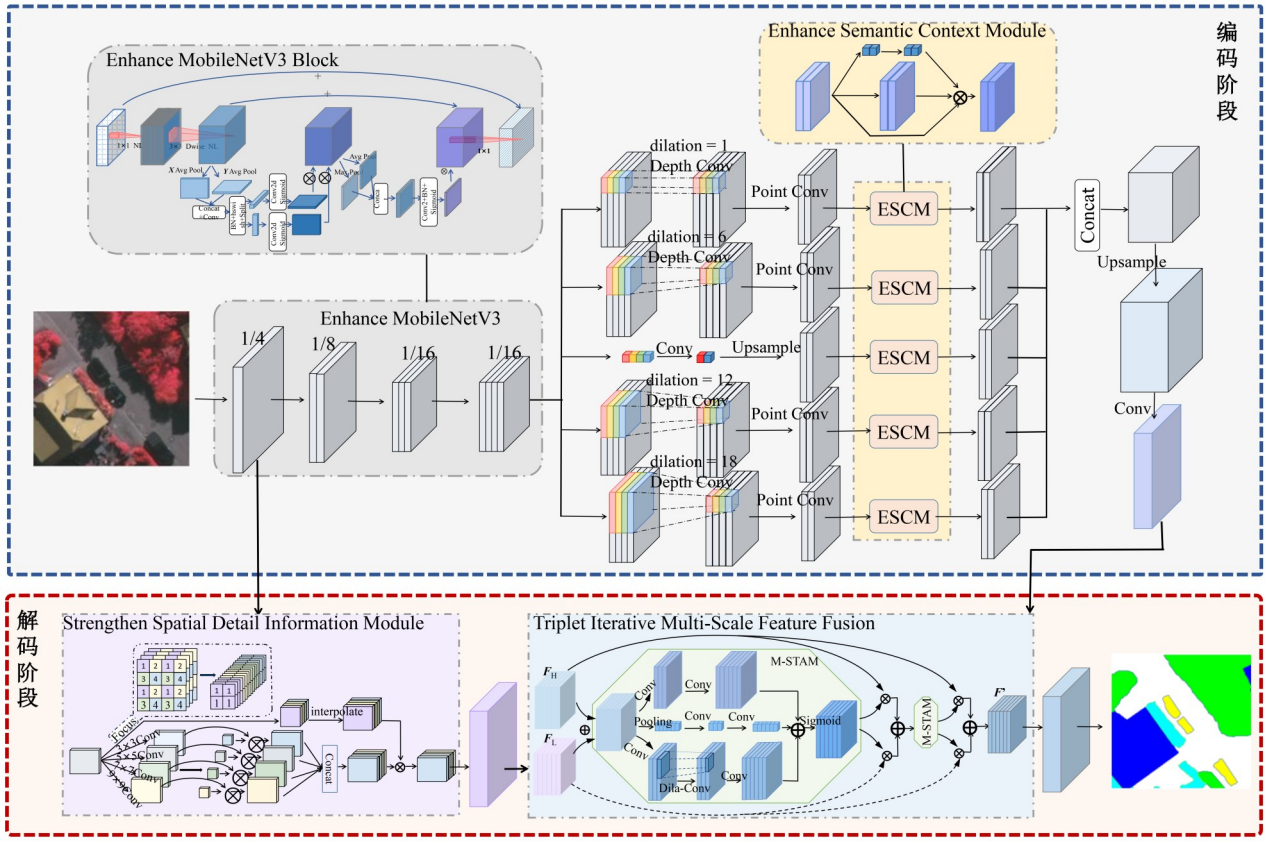
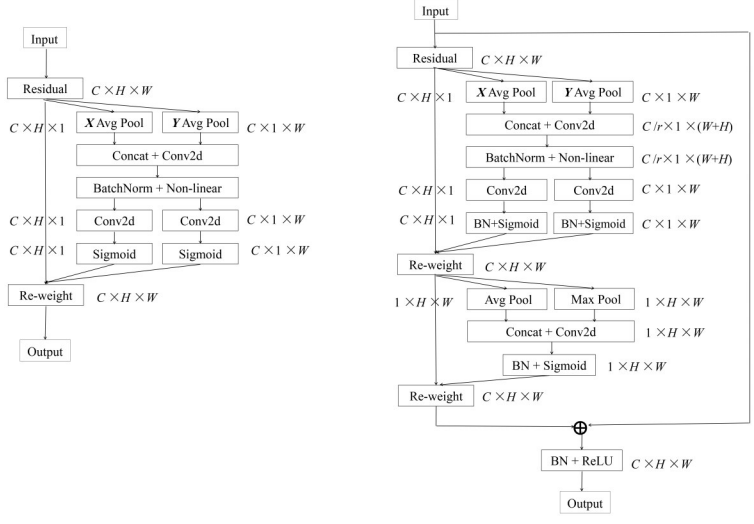


图 3　多尺度语义编解码网络 MSEDNet 结构



(a) CA　　(b) RCSA

图 4　CA 与 RCSA 示意图

修正,产生最后的特征图. RCSA 是一种轻量化的模块,此模块非常灵活和可扩展,可直接嵌入计算机视觉主干网络中以提升性能,这点在消融实验时已得到证实.

3.1.2　Enhance MobileNetV3

本文在轻量级网络 MobileNetV3-Small 版本中引入 RCSA,使主干网在不增加开销情况下,提取图像不同目标更准确的层次特征. 该结构在图 3 中用灰色虚线框标注,其核心模块结构如图 5 所示.

输入特征张量,使用两个空间范围池核$(H,1)$或$(1,W)$分别沿水平和垂直坐标对每个通道编码. 因此,高度h处或宽度w处的第c通道的输出可以分别表示为

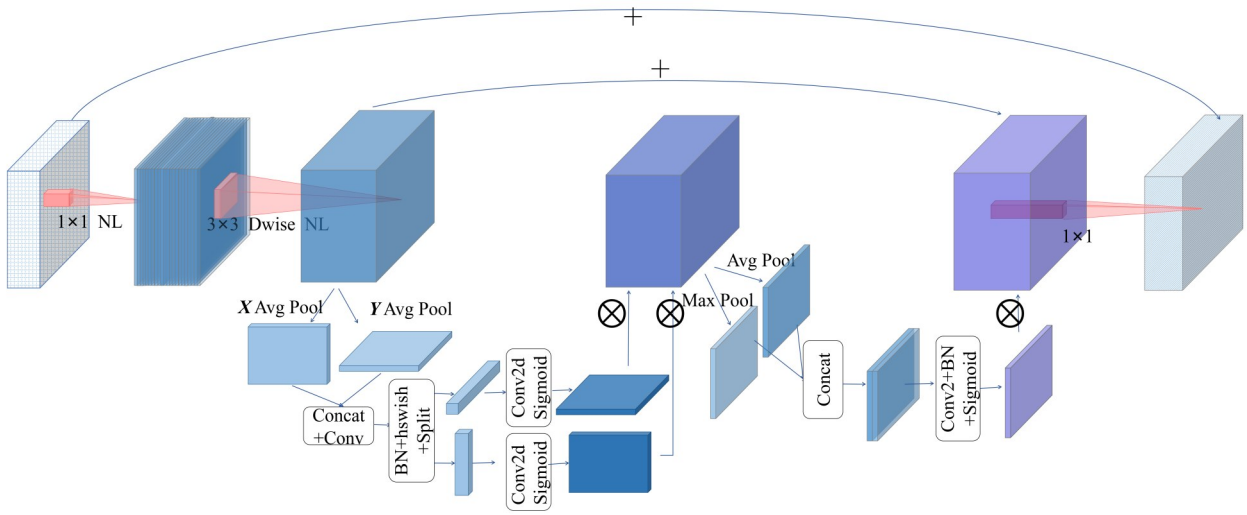


图5 Enhance MobileNetV3 Block

$$z_c^h(h)=\frac{1}{W}\sum_{0\leqslant i<W}\boldsymbol{x}_c(h,i) \tag{1}$$

$$z_c^w(w)=\frac{1}{H}\sum_{0\leqslant j<H}\boldsymbol{x}_c(j,w) \tag{2}$$

其中，z_c是与第c个通道在高h和宽w处相关联的输出；H和W分别为高度与宽度. 对于聚合特征图进行变换函数F_1，得到

$$\boldsymbol{f}=\delta(F_1([z^h,z^w])) \tag{3}$$

其中$[\cdot,\cdot]$表示沿空间维度的连接运算；F_1为卷积运算；δ为非线性激活函数；$\boldsymbol{f}\in\mathbb{R}^{C/r\times(H+W)}$为在两个方向编码的中间特征图；$r$是缩减率. 沿着空间维度将$\boldsymbol{f}$分割为两个单独张量$\boldsymbol{f}^h\in\mathbb{R}^{C/r\times H}$和$\boldsymbol{f}^w\in\mathbb{R}^{C/r\times W}$. 利用另两个变换函数$F_h$和$F_w$分别将$\boldsymbol{f}^h$和$\boldsymbol{f}^w$转换为输入$\boldsymbol{X}$相同通道数的张量，得

$$\boldsymbol{g}^h=\sigma(F_h(\boldsymbol{f}^h)) \tag{4}$$

$$\boldsymbol{g}^w=\sigma(F_w(\boldsymbol{f}^w)) \tag{5}$$

其中，σ是Sigmoid函数. CA输出为

$$\boldsymbol{f}_c(i,j)=\boldsymbol{x}_c(i,j)\times\boldsymbol{g}_c^h(i)\times\boldsymbol{g}_c^w(j) \tag{6}$$

接着对通道维数所有元素计算平均值和最大值，得到不同的空间信息Avg和Max，再融合得到完整空间结构信息特征图后，用卷积conv和激活函数σ对空间特征图进行归一化得到特征图$\boldsymbol{y}_s$，其中$0\leqslant i<W$，$0\leqslant j<H$. 最后，图像输出$\boldsymbol{Y}$张量可以表示成式(10).

$$\mathrm{Avg}=\mathrm{Mean}((\boldsymbol{f}),\mathrm{dim}=1) \tag{7}$$

$$\mathrm{Max}=\mathrm{Max}((\boldsymbol{f}),\mathrm{dim}=1) \tag{8}$$

$$\boldsymbol{y}_s(i,j)=\sigma(\mathrm{conv}(\mathrm{Avg}+\mathrm{Max})) \tag{9}$$

$$\boldsymbol{Y}=\boldsymbol{x}_c(i,j)+(\boldsymbol{f}_c(i,j)\times\boldsymbol{y}_s(i,j)) \tag{10}$$

3.2 多尺度特征与语义上下文信息

众多研究用SPP结构分割挑战性目标. AFNet[38]利用尺度层注意选择性聚焦，将不同区域信息结合. PSPNet[40]通过空间信息并行聚合得多尺度特性. DenseASPP[41]通过级联卷积层，生成更大规模特性. 但是当单一感受野比较大时，有足够的上下文信息，但对象特征会被不相关对象覆盖. 当特征映射有较小感受野时，又缺乏上下文信息. 且使用固定扩张速率或跨距，忽略了目标尺度规模，导致感受野与尺度不匹配.

针对图1所示(Ⅲ)类问题，本文提出增强语义上下文模块(Enhance Semanic Context Module，ESCM)，在图3中用黄色虚线框标注. 在提取上下文时，ASPP中使用的空洞率是固定的，忽略了具有挑战性对象的尺度规模. 于是将ESCM嵌入ASPP中，并利用由深度卷积(Depthwise Convolution)和点卷积(Pointwise Convolution)构成的深度可分离卷积替代普通卷积保持模型的整体轻量性. 在编码器阶段，Enhance MobileNetV3用作主干网络提取特征，再采用不同空洞率的多个并行空洞卷积层，为每个空洞率提取的特征在单独的分支中进行处理. 该模块通过不同的空洞率构建不同感受野的卷积核，用来获取多尺度物体信息. 每次卷积后，ESCM应用于所有分支的特征图，从而完成第一阶段的特征提取任务. ESCM模块生成全局与局部机制注意特征图，并在融合来自ASPP模块的特征图时，利用注意图对多尺度特征图进行自适应加权处理. 这些特征图具有相同的空间分辨率，但像素的感受野不同. 因此，根据各个不同挑战对象的规模尺度，具有匹配感受野的特征图将得到增强，而其他特征图将受到抑制，增强了具有与对象比例相匹配的感受野的特征地图，确保多尺度语境信息的有效捕获.

ESCM详细结构如图6所示. 特征图通过全局与局部机制来融合上下文信息. 这种尺度感知策略可以使学习特征更具代表性，有助于区分分割对象，从而完成对易混淆对象的分割.

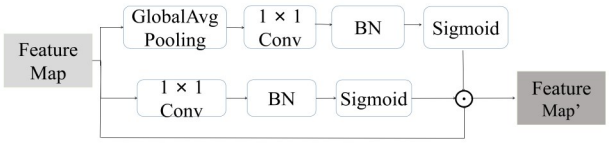


图6　增强语义上下文模块(ESCM)

ESCM输出可以表示为

$$\boldsymbol{F}_m' = \boldsymbol{F}_m \cdot \boldsymbol{F}_0 \cdot \boldsymbol{F}_1 \tag{11}$$

其中,$\boldsymbol{F}_m$为ESCM的输入;$\boldsymbol{F}_0$和$\boldsymbol{F}_1$分别为全局与局部注意图;$\boldsymbol{F}_m'$为输出特征图.$\boldsymbol{F}_0$和$\boldsymbol{F}_1$的生成过程分别表示为

$$\boldsymbol{F}_0 = \mathrm{Sigmoid}\left[\mathrm{BN}\left[k_0 \otimes \left[\mathrm{Pooling}\left(\boldsymbol{F}_m\right)\right]\right]\right] \tag{12}$$

$$\boldsymbol{F}_1 = \mathrm{Sigmoid}\left[\mathrm{BN}\left[k_1 \otimes \boldsymbol{F}_m\right]\right] \tag{13}$$

其中,Pooling是全球平均池化;k_0和k_1为逐点卷积参数;$\otimes$为卷积运算;BN为归一化函数;Sigmoid函数用作激活函数.

4　解码阶段

4.1　特征融合

目前,主要的融合关键点在于上下文感知水平.像拼接、加性及乘积聚合仅提供固定特征映射的线性聚合,完全与上下文无关.EDA在融合深、浅层信息时,浅层判别能力较差,也带来噪声等特征.而且现有结构仅将最后一层编码产生的高级信息进行一次融合,忽略了解码特征聚合过程的引导能力.

4.1.1　多尺度注意力机制融合

ResNet及其后续优化结构中,映射特征和残差学习通过短跳跃连接,而U-Net通过长跳跃连接.最近,SKNet[42]和ResNeSt[43]都基于通道注意对同一层的多个核或组的特征进行动态加权平均.这些方法为融合提供了非线性策略,但其只关注同一层的特征选择,而没有解决跳过连接的跨层融合问题.

本文中针对图1所示(Ⅱ)(Ⅲ)类深层、浅层融合问题,提出三元迭代多尺度特征融合(Triplet Iterative Multi-Scale Feature Fusion,TIMSFF)模块,图3中用浅蓝色虚线框标注.该模块解决了深、浅层多尺度特征上下文聚合和初始集成的问题,且实现了捕获更深层次多尺度上下文感知水平的跨层融合.

4.1.2　TIMSFF

TIMSFF以所设计的多分支跨维交互连接聚合邻域信息的多尺度三元注意力模块(Multi-Scale Triplet Attention Module,M-STAM)为核心单元,着重于多层次融合.模块中双分支方式与单分支方式为

$$\begin{aligned}\boldsymbol{F}_{\mathrm{dc}}(\boldsymbol{X}) = {} & B\left(\mathrm{Conv}_2\left(R\left(B\left(\mathrm{Conv}_1(\boldsymbol{X})\right)\right)\right)\right) \\ & + B\left(\mathrm{Conv}_2\left(R\left(B\left(\mathrm{Conv}_1\left(\mathrm{Gap}(X)\right)\right)\right)\right)\right)\end{aligned} \tag{14}$$

$$\boldsymbol{F}_{\mathrm{ss}}(\boldsymbol{X}) = B\left(\mathrm{Conv}_2\left(B\left(\mathrm{DilConv}_3\left(B\left(\mathrm{Conv}_1(\boldsymbol{X})\right)\right)\right)\right)\right) \tag{15}$$

其中,Conv_1和Conv_2为1×1卷积;$\mathrm{DilConv}_3$为3×3空洞卷积,B代表BN,R代表Relu,Gap代表全局平均池化函数.$\boldsymbol{F}_{\mathrm{dc}}(\boldsymbol{X})$和$\boldsymbol{F}_{\mathrm{ss}}(\boldsymbol{X})$与输入特征图具有相同形状,可保留和突出原始特征细节.通过M-STAM可得到多尺度精细特征注意权重$\boldsymbol{F}' \in \mathbb{R}^{C\times H\times W}$,可表示为式(16),其中,$\sigma$是Sigmoid激活函数.

$$\boldsymbol{F}'(\boldsymbol{X}) = \sigma(\boldsymbol{F}_{\mathrm{dc}}(\boldsymbol{X}) + \boldsymbol{F}_{\mathrm{ss}}(\boldsymbol{X})) \tag{16}$$

TIMSFF模块结构如图7所示,其设计目的是解决接收特征初始集成的融合瓶颈[26].TIMSFF模块中[·,·,·]参数分别表示[channel,height,width].其中,channel代表维度,height与width分别表示特征图的高度和宽度.如图7所示,TIMSFF模块输入张量为64维,经过M-STAM模块中双分支与单分支方式进行多尺度细化特征时,先经过卷积降维到16,再升维到64的方式来捕获特征信息,最后经过线性聚合得到与输入维度相同的输出特征图.TIMSFF可表示为

$$\begin{aligned}\boldsymbol{F}_{f1} = {} & \boldsymbol{F}_{\mathrm{H}} \otimes (\boldsymbol{F}'(\boldsymbol{F}_{\mathrm{H}} \oplus \boldsymbol{F}_{\mathrm{L}})) \\ & + \boldsymbol{F}_{\mathrm{L}} \otimes (1 - (\boldsymbol{F}'(\boldsymbol{F}_{\mathrm{H}} \oplus \boldsymbol{F}_{\mathrm{L}})))\end{aligned} \tag{17}$$

$$\boldsymbol{F}_{f2} = \boldsymbol{F}_{\mathrm{H}} \otimes \boldsymbol{F}'(\boldsymbol{F}_{f1}) + \boldsymbol{F}_{\mathrm{L}} \otimes (1 - \boldsymbol{F}'(\boldsymbol{F}_{f1})) \tag{18}$$

其中,$\boldsymbol{F}_{\mathrm{H}}$为高阶特征图;$\boldsymbol{F}_{\mathrm{L}}$为低阶特征图;$\boldsymbol{F}'$为多尺度三元注意力特征权重;$\otimes$表示按元素乘积运算;$\oplus$表示加性拼接运算;$\boldsymbol{F}_{f1}$表示为迭代M-STAM之前的特征图;$\boldsymbol{F}_{f2}$表示经过迭代之后的特征图.TIMSFF框架关注多尺度的多分支通道和空间注意力,基于多注意的特征融合从同层场景推广到跨层场景,包括长短连接.

4.2　强化多尺度特征图空间细节

浅层特征含盖大量小对象详细语义信息,直接进行卷积特征图空间位置等大部分细节信息会丢失.于是多尺度变化层面上增强空间细节信息并且建立长期依赖关系尤其重要.因此,针对图1所示(Ⅱ)类问题,提出了一种基于Focus强化空间细节信息(Strengthen Spatial Detail Information Module,SSDIM)的方法(图3中浅紫色虚线框),能在更细粒度水平上有效提取浅层语义和多尺度上下文细节信息.

SSDIM详细结构如图8所示,它由提取空间信息的多分支并行卷积层和将特征图宽、高信息整合到空间中的Focus分支层构成,避免浅层空间细节信息丢失.Focus分支层首次将Yolov5中的Focus与双线性插值结合用于语义分割.Focus对图像进行切片操作,将低阶特征图的宽度W和高度H信息整合到空间中,得到没有信息丢失情况的二倍下采样特征图.

SSDIM利用多卷积核自适应地学习特征图不同尺度信息,使上下文特征的相邻尺度可以更精确地合并.

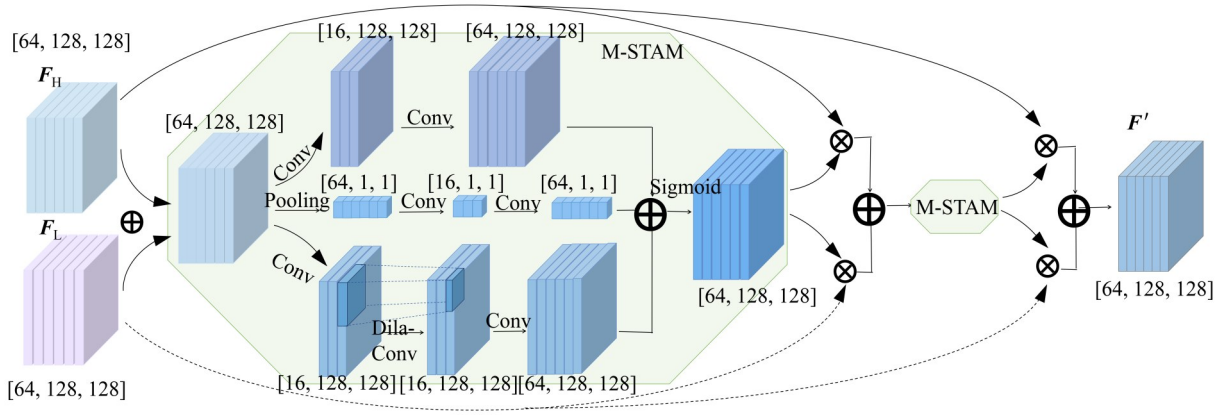


图7 三元迭代多尺度特征融合(TIMSFF)

接着,与Focus分支层特征进行融合.Focus的具体操作如图8虚线框所示,在一张图片中每隔一个像素拿到一个值,形成四张有信息丢失的互补图片,但输入通道数扩充四倍,特征图W和H变成了原来的1/2.再与多核卷积自适应获得的特征图进行元素乘法运算,最终得到的特征图具有丰富的多尺度上下文信息.这在一定程度上缓解了卷积使通道数增加而造成的空间细节信息丢失.

5 实验与分析

5.1 实验环境与对比模型

本文基于Pytorch搭建框架,所有实验都基于NVIDIA-SMI 460.67 GPU(11 GB)和Python3.7实现.学习率初始值设为0.001,batch size设置为8,优化器采用Adam,在类别像素上采用Negative Log Liklihood Loss(NLLLoss)损失函数,Epoch设置为600轮.为了便于比较,同一个数据集所有实验都在相同训练和测试设置下进行.

为证明本文所提网络有效性,与以下网络模型进行比较.

(1)与经典模型U-Net[14],SegNet[18],ResNet[20],ICNet[44]等进行对比实验,证实EDA有效性及实用性.

(2)与现有模型MANet[6],MAResU-Net[27],MACU-Net[17],SCAttNet[45]等网络进行对比实验,证明所提出的多层次、多尺度、多类别策略的上下文语义信息及空间细节信息的重要性.

(3)为证明提出网络轻量性,也与目前较新轻量性模型MobileVit[46]进行了性能及Param对比.

(4)为评估网络中每个改善模块的作用,与MobileNetV3基准模型进行对比,完成消融实验.

5.2 数据集

本文利用国际摄影测量与遥感协会提供的两个公开数据集ISPRS Vaihingen和Potsdam进行实验.数据集中都包含最常见的土地覆盖类别,分别是道路、建筑、草地、树木、车辆以及背景.Vaihingen数据集是一个相对较小的村庄,有着独立的建筑物和小的多层建筑物,其语义标注数据集由33幅平均大小为2 500×2 000像素、空间分辨率为9 cm的遥感图像.数据集提供了近红外、红色、绿色3个波段以及数字表面模型(Digital Surface Model,DSM).我们利用ID:1,3,5,7,11,13,15,17,21,23,26,28,32,34,37进行训练,ID:30进行验证,剩下17幅用于测试.在实验中只使用了RGB三通道的正射影像.为了训练,将图像裁剪成512×512像素,通过随机轴旋转、水平轴翻转、垂直轴翻转、随机缩放、添加随机高斯噪声等方式对其进行数据集增强扩充.

Potsdam数据集是一个典型的历史城市,有大型建筑群、狭窄的道路和密集的聚落结构,其包含38幅6 000×6 000像素、空间分辨率为5 cm的精细分辨率图像,提供了红外、红色、绿色和蓝色通道,以及数字地表模型(DSM)和标准化数字地表模型(Normalized Digital Surface Model,NDSM).我们利用ID:2_11,2_12,3_10,3_11,3_12,4_10,4_11,4_12,5_10,5_11,5_12,6_7,6_8,6_9,6_10,6_11,6_12,7_7,7_8,7_9,7_11和7_12进行训练,ID:2_10进行验证,剩余的15幅图像进行测试.在实验中只使用了RGB三通道的正射影像.为了训练,将图像剪裁成600×600像素,然后利用与数据集Vaihingen相同的方式进行数据集增强操作.数据集Potsdam的训练过程与数据集Vaihingen训练过程一致.

5.3 评价指标

为了评价多尺度语义编解码网络MSEDNet在数据集Vaihingen和Potsdam上的精细分割性能,与多数遥感语义分割方法采用的指标一致,在实验中使用了4个常用语义分割评价指标,分别是总体准确率(Overall Accuracy,OA)、平均交并比(mean Intersection Over Union,mIoU)、平均F_1-score(mean F_1-score,mF_1)和参数量(Param).

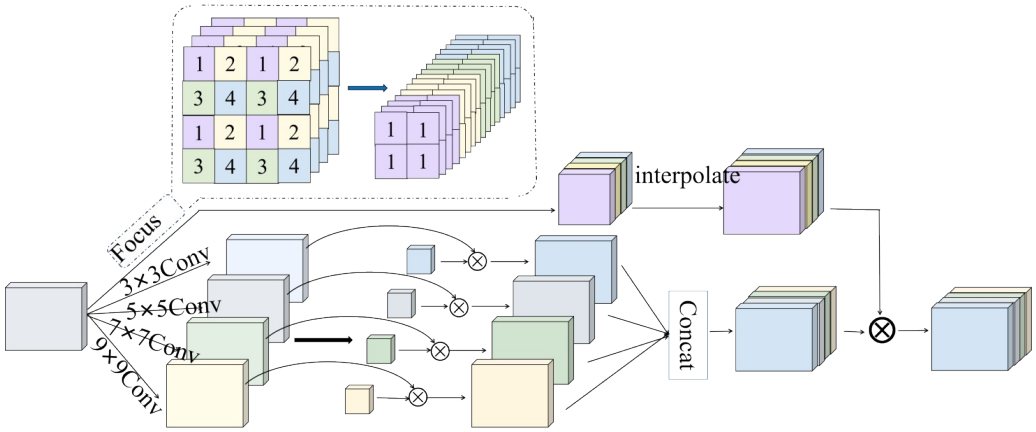


图8　强化空间细节信息模块(SSDIM)

$$\mathrm{OA}=\frac{\sum_{k=1}^{N}\mathrm{TP}_k}{\sum_{k=1}^{N}\mathrm{TP}_k+\mathrm{FP}_k+\mathrm{TN}_k+\mathrm{FN}_k} \tag{19}$$

$$\mathrm{mIoU}=\frac{1}{N}\sum_{k=1}^{N}\frac{\mathrm{TP}_k}{\mathrm{TP}_k+\mathrm{FP}_k+\mathrm{FN}_k} \tag{20}$$

$$F_1=2\times\frac{\mathrm{precision}\times\mathrm{recall}}{\mathrm{precision}+\mathrm{recall}} \tag{21}$$

$$\mathrm{precision}=\frac{\sum_{k=1}^{N}\mathrm{TP}_k}{\sum_{k=1}^{N}\mathrm{TP}_k+\mathrm{FP}_k} \tag{22}$$

$$\mathrm{recall}=\frac{\sum_{k=1}^{N}\mathrm{TP}_k}{\sum_{k=1}^{N}\mathrm{TP}_k+\mathrm{FN}_k} \tag{23}$$

其中,TP_k代表真正类的数量;FP_k代表假正类的数量;TN_k代表真负类的数量;FN_k代表假负类的数量;k表示索引为k类的对象;OA对包括背景在内的所有类别进行计算;precision和recall是准确率和召回率.

5.4　实验结果

表1和表2分别展示了在Vaihingen数据集和Potsdam数据集上的对比实验结果. 表中,加粗标注表示在不同网络模型中各项指标最好的数据. ↑与↓分别表示与最好数据相比较所提出网络的验证数据是增长还是下降.

表1显示了在Vaihingen数据集上不同方法的实验结果. 所提网络模型MSEDNet比现有网络的性能更好,参数量更少,特别是对于小对象,如车辆. 即使使用单一的RCSA模块改进后的Enhance MobileNetv3增强特征图语义信息,OA至少可以获得94.656%的改善,mF_1为89.568%,mIoU为71.833%. 此外,浅层解码改进模块比深层编码模块的贡献更大,因为后者包含丰富的上下文空间细节信息. 当所有模块均附加时,OA显著增加到95.699%,mF_1增加到91.949%,mIoU显著增加到75.948%. 这些结果表明,本文提出的网络模型可从不同角度挖掘特征图全局与局部的上下文信息,对不同尺度的对象都有较好分割效果,且网络参数量显著下降到6.77 M.

表2显示了在Potsdam数据集上各种方法的实验结果. 实验数据证明,MSEDNet网络也适应Potsdam数据集. 仅仅使用Enhance MobileNetv3网络结构进行实验,OA, mF_1, mIoU分别可以达到93.904%, 89.955%, 71.659%. 在MSEDNet整体网络架构中,OA增长到95.534%,mF_1增长到92.038%,mIoU增长到75.683%,与其他现有网络相比,OA,mF_1,mIoU分别增加了1.299%, 2.929%, 4.012%. 这些结果表明,本文提出的ESCM, SSDIM,TIMSFF模块,不仅提升了多尺度结构特征图的表征能力,且强化了特征图空间细节信息,提深高了分辨率遥感图像语义分割精度.

为了检验模型训练参数设置的合理性及收敛性,由Vaihingen训练集、Potsdam训练集的前200个epoch绘制得到图9所示的两个数据集的训练准确度(Overall Accuracy, OA)曲线和训练损失(Loss)曲线. 其中,红色、紫色、黑色表示训练OA曲线,蓝色、绿色、青色表示训练Loss曲线. 相比基线(MobileNet3+ASPP),可以明显看出编码阶段MSEDNet(RCSA+ESCM)和解码阶段MSEDNet(RCSA+ESCM+TIMSFF+SSDIM)的训练曲线(黑色)都带来了OA方面的改进. 其中,解码阶段的OA(黑色)曲线及Loss(青色)曲线比编码阶段对应曲线更快达到稳态,这意味着浅层语义信息与空间细节信息加速了网络的融合. 从图9也可以看出,MSEDNet网络的附加模块也适应不同的数据集,具有一定的鲁棒性.

5.5　可视化分析

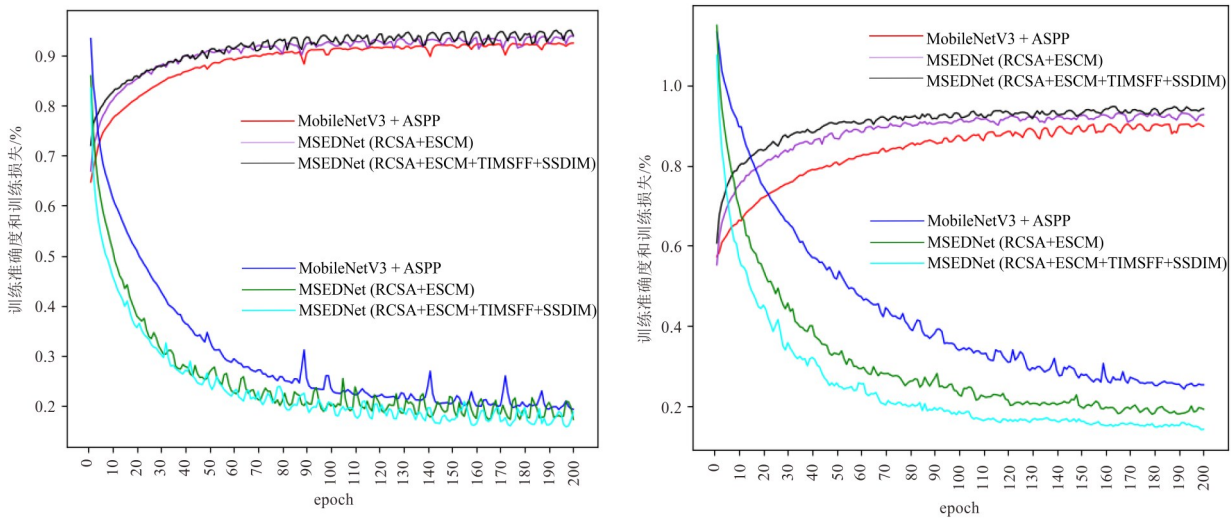
图10所示为Vaihingen测试集(前三行)和Potsdam测试集(后两行)的标签可视化结果. 其中,有明显改善的区域用红色虚线框突出显示. 如图10所示,MACU-Net

表1 在Vaihingen数据集上的语义分割结果比较

网络	#Param	道路/%	建筑/%	草地/%	树木/%	车辆/%	mIoU/%	mF_1/%	OA/%
U-Net	34.96 M	87.836	94.254	84.109	89.366	70.383	65.405	85.189	92.561
SegNet	**8.19 M**	85.991	91.529	82.786	87.917	69.738	63.644	83.592	91.566
ICNet	109.71 M	88.965	91.368	83.494	86.750	71.394	67.841	84.394	90.725
ResNet18	58.40 M	91.032	93.489	88.057	90.643	77.620	70.973	88.168	92.954
ResNet152	237.57 M	**91.476**	94.360	89.792	91.951	**78.025**	71.593	89.121	**94.592**
MANet	136.80 M	90.264	**94.955**	86.831	91.820	74.792	69.273	87.732	93.363
MAResU-Net	100.24 M	89.468	94.825	87.783	92.440	73.426	69.058	87.588	93.492
MACU-Net	19.65 M	82.114	90.299	79.026	84.682	67.785	70.884	80.781	89.659
SCAttNet	105.45 M	90.735	94.402	88.979	91.663	76.201	**71.975**	88.396	94.020
ResNet34+Deeplabv3+	85.28 M	90.713	94.565	**90.295**	**92.485**	77.484	71.212	89.108	94.524
MobileVit-XS	8.35 M	90.956	94.861	85.614	92.121	74.935	68.566	87.697	93.677
MobileVit-S	19.67 M	91.372	94.914	89.925	92.432	77.794	70.981	**89.288**	94.328
Enhance MobileNetV3	**6.35 M** ↓	**91.845** ↑	**94.927** ↓	**90.422** ↑	**92.696** ↑	**77.951** ↓	**71.833** ↓	**89.568** ↑	**94.656** ↑
MSEDNet(Ours)	**6.77 M** ↓	**93.211** ↑	**96.146** ↑	**92.310** ↑	**94.765** ↑	**83.312** ↑	**75.948** ↑	**91.949** ↑	**95.699** ↑

表2 在Potsdam数据集上的语义分割结果比较

网络	#Param	道路/%	建筑/%	草地/%	树木/%	车辆/%	mIoU/%	mF_1/%	OA/%
ResNet18	58.40 M	91.348	88.287	89.506	90.643	70.256	70.237	86.015	91.837
ResNet152	237.57 M	92.281	89.804	90.925	**91.006**	81.525	71.218	**89.109**	94.010
MANet	136.80 M	90.138	89.680	89.539	89.850	76.203	70.847	87.082	92.405
MAResU-Net	100.24 M	91.574	**90.895**	90.328	90.217	80.965	71.004	88.796	93.646
MACU-Net	**19.65 M**	89.246	86.742	88.845	87.325	78.655	70.540	86.163	91.764
SCAttNet	105.45 M	90.856	88.036	89.887	87.541	**81.598**	70.952	87.584	92.194
ResNet34+Deeplabv3+	85.28 M	**92.489**	88.314	**91.472**	89.652	80.254	71.065	88.436	**94.235**
MobileVit-S	19.67 M	91.801	89.597	90.463	90.511	75.089	**71.671**	87.492	92.821
Enhance MobileNetV3	**6.35 M** ↓	**93.334** ↑	**90.736** ↓	**92.911** ↑	**90.657** ↓	**82.135** ↑	**71.659** ↓	**89.955** ↑	**93.904** ↓
MSEDNet(Ours)	**6.77 M** ↓	**95.490** ↑	**92.207** ↑	**94.531** ↑	**93.129** ↑	**84.835** ↑	**75.683** ↑	**92.038** ↑	**95.534** ↑



(*a*) Vaihingen (*b*) Potsdam

图9 训练准确度和训练损失曲线图

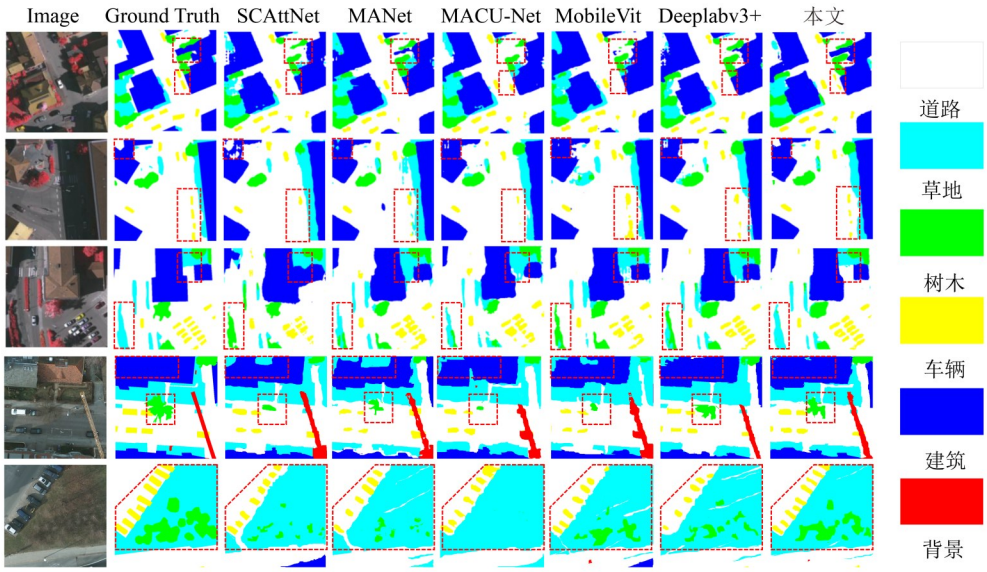


图 10　标签可视化结果图

分割效果最差,因其空间信息丢失,无法提取多尺度上下文信息. SCAttNet, DeepLabv3+, MANet 和 MobileVit 会产生部分错误分割. 如第一行红色虚线框区域所示,建筑和车辆进行了错误分割及漏分割;第二、四行区域的道路和建筑存在更多错误分割和边界混淆;第三、四、五行区域的树木和草地两者出现混淆分割. 可以看出,所提 MSEDNet 可捕获多级自适应比例特征,实现较好的分割. 这可归功于多层注意块为每种类别对象提取适当上下文信息,不仅减少了不完整和不规则语义对象,而且更好地保留了几何细节和复杂轮廓. 同时,由于增强了浅层细节特性,小对象分割更加准确,空间位置信息得以保留. 这表明,该方法对关键类别分割有了显著改善.

图 11 所示为 Vaihingen 测试集(前三行)和 Potsdam 测试集(后两行)的热力图可视化结果. 热力图以特殊高亮的形式显示测试集中每类对象所在的地理区域的位置分布,其中有明显改善的区域用黑色虚线框突出显示. 如图 11 所示,相比所提 MSEDNet 网络具有关注多尺度上下文信息和浅层细节特性的特性, MACU-Net, SCAttNet, DeepLabv3+, MANet 和 MobileVit 会产生部分建筑阴影及小对象的错误分割. 如第一行黑色虚线框区域所示,建筑阴影的颜色高亮与真实建筑没有完全进行区分且出现漏分割真实建筑现象;第二、四区域的热力图展示出建筑阴影下的车辆和真实建筑存在更多错误分割和边界混淆现象;第三、五行区域的树木和草地的高亮颜色深浅大多数是一类的,导致两者出现混淆分割. 在 MSEDNet 网络下,建筑、草地、树木、车辆等部分都能得到突出显示. 由热力图可视化看出,所提 MSEDNet 可捕获多级自适应比例特征,实现较好的多类、多尺度语义分割.

5.6　消融实验

在所提 MSEDNet 中,各个模块都针对性地被用来改善图 1 所示(Ⅰ)(Ⅱ)(Ⅲ)类存在的问题,是为了增强高分辨率遥感图像特征提取和表征能力. 为了评估附加每个模块的性能及更直观地体现每个单独模块的贡献,本文分别使用表 3 和表 4 中列出的不同设置,在数据集 Vaihingen 和 Potsdam 上完成消融实验. 其中,×表示未加入对应模块,√表示加入对应模块. 表 3 和表 4 中加粗数字分别表示加入各模块改善后的结果.

表 3 展示出在编码阶段 MobileNeV3 主干中引入了 RCSA 模块,利用特征间空间关系及语义位置边界信息获取更准确的特征,以及利用 ESCM 增强特征表征能力的金字塔结构,相比基线,在 Vaihingen 和 Potsdam 数据集上 OA, mIoU, mF_1 各指标增长 2% 左右. 在解码阶段引入结合局部和全局思想的 TIMSFF 以解决深、层多尺度特征上下文感知水平跨层聚合问题,此外, SSDIM 的引入更证实了特征图多尺度空间细节信息的重要性. 相比编码阶段,解码阶段改善模块贡献更大,在 Vaihingen 上 OA, mIoU, mF_1 各指标分别增长 0.996%, 3.497%, 1.993%,且在 Potsdam 上 OA, mIoU、mF_1 各指标分别提高 1.062%, 2.336%, 1.237%. 因此,实验数据表明,所提 MSEDNet 对高分辨率遥感图像语义分割有一定的成效.

表 4 中,前面四组数组主要对 RCSA 模块的灵活性和可扩展性进行了证实,表明 RCSA 可直接嵌入计算机视觉主干网络中以提升性能. 可以看出,这四组数据不管是在 Vaihingen 数据集,还是在 Potsdam 数据集上,都表现出较好的提升效果. 表中最后加粗的 4 行分别是针对基于基线模型(MobileNet3+ASPP)分别加入

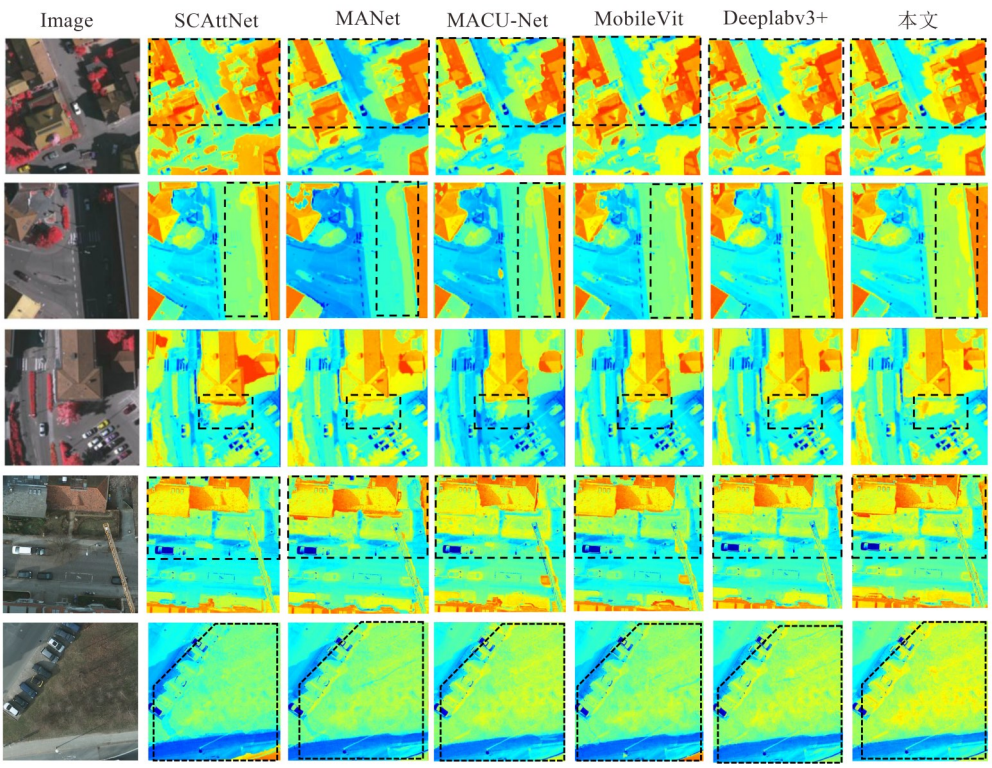


图 11 热力图可视化结果图

表 3 基于多尺度语义编解码网络 MSEDNet 各模块消融实验对比

数据集	模块				Vaihingen			Potsdam		
网络	RCSA	ESCM	TIMSFF	SSDIM	mIoU/%	mF_1/%	OA/%	mIoU/%	mF_1/%	OA/%
ResNet34 + DeepLabv3+	×	×	×	×	71.212	89.108	94.524	71.065	88.436	94.235
MobileNetv3 + ASPP	×	×	×	×	70.368	87.985	91.382	71.258	89.602	92.549
Enhance MobileNetv3	√	×	×	×	71.833	89.568	94.656	71.659	89.955	93.904
MSEDNet(本文)	√	√	×	×	**72.451**	**89.956**	**94.703**	**73.347**	**90.787**	**94.472**
MSEDNet(本文)	√	√	√	×	73.377	90.785	94.986	75.070	91.419	94.848
MSEDNet(本文)	√	√	√	√	**75.948**	**91.949**	**95.699**	**75.683**	**92.024**	**95.534**

RCSA,ESCM,TIMSFF,SSDIM 这四个文中提出的模块,以便更能体现出每个模块单独的功能. 与基线模型进行对比,从数据指标上可以看出在分别加入这几个模块时两个数据集的 mIou,mF_1,OA 都能得到一定的提升,也可以证明所提各个模块对高分辨率遥感图像语义分割有一定的针对性改善. 为了更直观地体现出每个单独模块的贡献,通过对表 4 中最后四行结果进行可视化,如图 12 所示. 图 12 以 Vaihingen 测试集来呈现各个单模块可视化,因为 Vaihingen 是一个村庄数据集,而 Potsdam 是一个城镇数据集,相比 Potsdam 城镇数据集,Vaihingen 村庄数据集更具复杂性、杂乱性和代表性.

图 12 以基线(Base)模型(MobileNet3+ASPP)为前提,分别加入文中所提 RCSA,ESCM,TIMSFF,SSDIM 四个模块所得到的单独模块测试的可视化结果图,其中有明显改善的区域用红色虚线标记显示. 在第一组可视化(Base+RCSA)结果中,可明显看出相比 Base,建筑、道路、草地与树木的边界都得到明显改善,且各类的空间范围更加精确. RCSA 模块可正确聚焦目标对象,有效细化特征且语义位置边界信息可以有效分割外界相似特征,针对性地缓解了图 1 中的(Ⅰ)类问题. 在第二组(Base+ESCM)结果中,明显得到改善的是草地与树木的区别分割,证明 ESCM 提升了具有类内相似性与类间相似性对象的多尺度特征图的表征能力,有效地改善了图 1 中的(Ⅲ)类问题. 在第三组(Base+

表 4　基于多尺度语义编解码网络 MSEDNet 单独模块实验对比

数据集	模块				Vaihingen			Potsdam		
网络	RCSA	ESCM	TIMSFF	SSDIM	mIoU/%	mF_1/%	OA/%	mIoU/%	mF_1/%	OA/%
ResNet34 + Deep-Labv3+	×	×	×	×	71.212	89.108	94.524	71.065	88.436	94.235
ResNet34 + Deep-Labv3+	√	×	×	×	71.689	89.395	94.582	71.316	88.981	94.372
ResNet18	×	×	×	×	70.973	88.168	92.954	70.237	86.015	91.837
ResNet18	√	×	×	×	71.157	88.415	93.182	70.613	86.476	92.305
MACU-Net	×	×	×	×	70.884	80.781	89.659	70.540	86.163	91.764
MACU-Net	√	×	×	×	71.351	81.062	90.195	70.775	86.602	91.925
MobileNetv3 + ASPP	×	×	×	×	70.368	87.985	91.382	71.258	89.602	92.549
Enhance MobileNetv3	√	×	×	×	**71.833**	**89.568**	**94.656**	**71.659**	**89.955**	**93.904**
MSEDNet(本文)	×	√	×	×	**70.945**	**89.106**	**92.750**	**71.537**	**90.132**	**92.961**
MSEDNet(本文)	×	×	√	×	**71.125**	**89.732**	**93.007**	**72.054**	**90.315**	**94.011**
MSEDNet(本文)	×	×	×	√	**73.021**	**90.435**	**94.982**	**73.514**	**91.153**	**94.704**

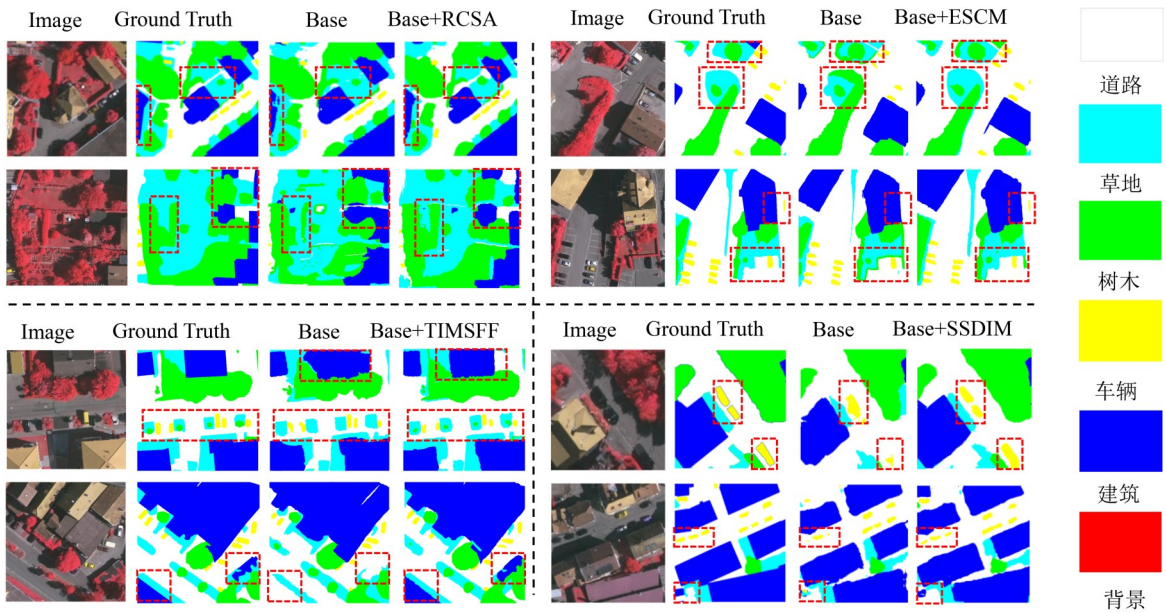


图 12　相比基线模型的各单模块测试可视化结果图

TIMSFF)测试可视化结果中,基于 Base 模型出现了部分漏分割或错误分割,建筑、草地、树木是明显得到改善的,TIMSFF 模块可以捕获更深的、浅层次多尺度上下文感知,实现跨层融合增强特征聚合过程的引导能力,针对性改善了图 1 中(Ⅱ)(Ⅲ)类问题. 在第四组(Base+SSDIM)结果中,主要是对小对象的改善,可明显看出车辆及小建筑物的改善能力,证明了 SSDIM 保留住了小对象的空间细节详细信息及结构信息,针对性地改善了图 1 中存在的(Ⅱ)类问题. 图 12 所示也证实了文中所提各个模块对遥感语义分割效果有明显改善.

5.7　实验结果分析

综合分析数据集结果、可视化结果和消融实验结果,本文得到以下结论.

(1)由表 1~3 和表 4 数据结果,仅使用单一的 RCSA 模块改进的 Enhance MobileNetv3 提取语义信息后,各类评价指标值都得到了提升;由图 10、图 11 和图 12 可视化结果可知,各类边界性能得到改善. 因此,RCSA 关注特征间的空间关系,增强网络捕获图像语义位置边界信息特征,针对性地缓解了图 1 中的(Ⅰ)类问题.

(2)数据集中草地和树木存在类内与类间的相似特性,大型建筑群、狭窄的道路和密集的聚落存在多尺度状态. 由表 3 和表 4 数据结果可知,在使用 ESCM 模块后各类指标得到提升;由图 10、图 11 和图 12 可视化结果可知,各类别得到了明显区别分割. 因此,ESCM

提升了多尺度、多类别结构特征图的表征能力,缓解了图1中的(Ⅲ)类问题.

(3)通过对比表1、表2实验数据和图10~12可视化结果可知,小对象(车辆)的相对F_1提升比例较大. 浅层特征涵盖大量小对象详细语义信息,提出的SSDIM模块能在更细粒度水平上有效提取浅层语义和多尺度上下文细节信息,缓解了图1所示(Ⅱ)类问题.

(4)综合分析图10~12、表3和表4,通过TIMSFF实现了图像深层全局语义信息与浅层局部细节特征跨层融合,提升了高分辨率遥感图像语义分割精度,有效且可靠地缓解了图1中的(Ⅰ)(Ⅱ)(Ⅲ)类问题.

因此,本文所提基于多尺度语义编解码网络MSEDNet综合运用多项技术,针对性地缓解了高分辨率遥感图像语义分割存在的多层次信息提取和多尺度特征图上下文依赖性两个问题.

6 结束语

为了提高遥感图像语义分割精度,本文提出了基于多尺度语义编解码网络MSEDNet. 该网络在编、解码阶段,采用了RCSA,ESCM,SSDIM,TIMSFF等针对性的设计,加强不同特征图中多层次、多类别的深层、浅层语义信息的提取和特征图多尺度上下文感知水平的依赖性. 提出的网络模型在Vaihingen和Potsdam数据集上取得了较好的效果,所耗参数量显著下降至6.77 M. 在Vaihingen数据集上,总体分割精确度(OA)达到95.699%,平均F_1-score(mF_1)增加到91.949%,且平均交并比(mIoU)显著增加到75.948%. 在Potsdam数据集上,OA,mF_1,mIoU分别增长了1.299%,2.929%,4.012%. 通过可视化及各种消融实验分析验证,所提方案实现了更精确语义边界分割以及多类别各尺度对象的语义分割. 在未来的研究中,我们将把MSEDNet中的各个模块推广到更多的数据集,并通过调整或提出新模块来适应多类提取、道路检测和土地覆盖分类等任务,提高遥感图像语义分割性能.

参考文献

[1] SHOTTON J, JOHNSON M, CIPOLLA R. Semantic texton forests for image categorization and segmentation[C]// 2008 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2008: 1-8.

[2] ARBELÁEZ P, HARIHARAN B, GU C H, et al. Semantic segmentation using regions and parts[C]//2012 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2012: 3378-3385.

[3] HUANG Q, XIA C Y, LI S Y, et al. Unsupervised clustering guided semantic segmentation[C]//2018 IEEE Winter Conference on Applications of Computer Vision (WACV). Piscataway: IEEE, 2018: 1489-1498.

[4] LYU H R, FU H Y, HU X J, et al. Esnet: Edge-based segmentation network for real-time semantic segmentation in traffic scenes[C]//2019 IEEE International Conference on Image Processing (ICIP). Piscataway: IEEE, 2019: 1855-1859.

[5] WANG W H, XIE E Z, SONG X G, et al. Efficient and accurate arbitrary-shaped text detection with pixel aggregation network[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2020: 8439-8448.

[6] DAI Y M, GIESEKE F, OEHMCKE S, et al. Attentional feature fusion[C]//2021 IEEE Winter Conference on Applications of Computer Vision (WACV). Piscataway: IEEE, 2021: 3559-3568.

[7] DENG G H, WU Z C, WANG C J, et al. CCANet: Class-constraint coarse-to-fine attentional deep network for sub-decimeter aerial image semantic segmentation[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 1-20.

[8] YANG Q R, KU T, HU K Y. Efficient attention pyramid network for semantic segmentation[J]. IEEE Access, 2021, 9: 18867-18875.

[9] HU J, SHEN L, ALBANIE S, et al. Squeeze-and-excitation networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(8): 2011-2023.

[10] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional block attention module[C]//European Conference on Computer Vision. Cham: Springer, 2018: 3-19.

[11] HUANG Z L, WANG X G, HUANG L C, et al. CCNet: Criss-cross attention for semantic segmentation[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2020: 603-612.

[12] HOU Q B, ZHOU D Q, FENG J S. Coordinate attention for efficient mobile network design[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2021: 13708-13717.

[13] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C]/2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2015: 3431-3440.

[14] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation [C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer,

2015: 234-241.

[15] ZHOU Z W, SIDDIQUEE M M R, TAJBAKHSH N, et al. UNet++: A nested U-net architecture for medical image segmentation[EB/OL]. (2018-07-18) [2022-05-05]. https://arxiv.org/abs/1807.10165.

[16] HUANG H M, LIN L F, TONG R F, et al. UNet 3: A full-scale connected UNet for medical image segmentation [C]//ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway: IEEE, 2020: 1055-1059.

[17] LI R, DUAN C X, ZHENG S Y, et al. MACU-net for semantic segmentation of fine-resolution remotely sensed images[J]. IEEE Geoscience and Remote Sensing Letters, 2022, 19: 1-5.

[18] 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.

[19] ISOBE S, ARAI S. Deep convolutional encoder-decoder network with model uncertainty for semantic segmentation[C]//2017 IEEE International Conference on INnovations in Intelligent SysTems and Applications (INISTA). Piscataway: IEEE, 2017: 365-370.

[20] 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 (CVPR). Piscataway: IEEE, 2016: 770-778.

[21] IBTEHAZ N, RAHMAN M S. MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation[J]. Neural Networks: the Official Journal of the International Neural Network Society, 2020, 121: 74-87.

[22] GUO C L, LI C Y, GUO J C, et al. Hierarchical features driven residual learning for depth map super-resolution [J]. IEEE Transactions on Image Processing, 2019, 28(5): 2545-2557.

[23] LIU J Q, WANG Z L, CHENG K X. An improved algorithm for semantic segmentation of remote sensing images based on DeepLabv3+[C]//Proceedings of the 5th International Conference on Communication and Information Processing. New York: ACM, 2019: 124-128.

[24] 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.

[25] CHEN L C, ZHU Y K, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]//European Conference on Computer Vision. Cham: Springer, 2018: 833-851.

[26] LI A J, JIAO L C, ZHU H, et al. Multitask semantic boundary awareness network for remote sensing image segmentation[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 1-14.

[27] 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.

[28] 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.

[29] 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: 1-13.

[30] WU Z F, SHEN C H, VAN DEN HENGEL A. Wider or deeper: Revisiting the ResNet model for visual recognition[J]. Pattern Recognition, 2019, 90: 119-133.

[31] PENG C, ZHANG X Y, YU G, et al. Large kernel matters —Improve semantic segmentation by global convolutional network[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2017: 1743-1751.

[32] TAKIKAWA T, ACUNA D, JAMPANI V, et al. Gated-SCNN: Gated shape CNNs for semantic segmentation [C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2020: 5228-5237.

[33] 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.

[34] 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 (CVPR). Piscataway: IEEE, 2017: 6230-6239.

[35] BAI H W, CHENG J, HUANG X, et al. HCANet: A hierarchical context aggregation network for semantic segmentation of high-resolution remote sensing images[J]. IEEE Geoscience and Remote Sensing Letters, 2022, 19:

6002105.
[36] HOWARD A, SANDLER M, CHEN B, et al. Searching for MobileNetV3[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2020: 1314-1324.
[37] WANG J F, CHEN Y, GAO M Y, et al. Improved YOLOv5 network for real-time multi-scale traffic sign detection[EB/OL]. (2021-12-16)[2022-05-05]. https://arxiv.org/abs/2112.08782.
[38] 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.
[39] HU X X, YANG K L, FEI L, et al. ACNET: Attention based network to exploit complementary features for RGBD semantic segmentation[C]//2019 IEEE International Conference on Image Processing (ICIP). Piscataway: IEEE, 2019: 1440-1444.
[40] NOGUEIRA K, DALLA MURA M, CHANUSSOT J, et al. Dynamic multicontext segmentation of remote sensing images based on convolutional networks[J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57(10): 7503-7520.
[41] VOLPI M, TUIA D. Dense semantic labeling of subdecimeter resolution images with convolutional neural networks[J]. IEEE Transactions on Geoscience and Remote Sensing, 2017, 55(2): 881-893.
[42] LI X, WANG W H, HU X L, et al. Selective kernel networks[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 510-519.
[43] LI D Q, HU X Q, WANG S Q, et al. Hyperspectral images ground object recognition based on split attention[C]//2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE). Piscataway: IEEE, 2021: 324-330.
[44] ZHAO H S, QI X J, SHEN X Y, et al. ICNet for real-time semantic segmentation on high-resolution images[C]//European Conference on Computer Vision. Cham: Springer, 2018: 418-434.
[45] LI H F, QIU K J, CHEN L, et al. SCAttNet: 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.
[46] MEHTA S, RASTEGARI M. MobileViT: Light-weight, general-purpose, and mobile-friendly vision transformer [EB/OL]. (2021-10-05) [2022-05-05]. https://arxiv. org/abs/2110.02178.

作者简介

梁 燕 女,1977年生,重庆人. 于重庆邮电大学获硕士学位. 现任重庆邮电大学通信与信息工程学院副教授、硕士生导师. 主要研究方向为移动通信、物联网AI、图像处理.
E-mail: liangyan@cqupt.edu.cn

易春霞 女,1996年生,重庆人. 现为重庆邮电大学通信与信息工程学院硕士研究生. 主要研究方向为计算机视觉、AI图像处理.
E-mail: 1638362782@qq.com

王光宇 男,1964年生,贵州人. 于德国基尔大学获博士学位. 现任重庆邮电大学海外特聘教授,就职于德国英飞凌半导体公司. 主要研究方向为5G/6G移动通信、AI人工智能.
E-mail: wangguangyu@cqupt.edu.cn

胡跃辉 男,2002年生,重庆人. 现为重庆邮电大学通信与信息工程学院本科生,参与重庆邮电大学本科生科研训练计划. 主要研究方向为AI图像处理.
E-mail: 2372230575@qq.com