激光与光电子学进展 先进成像

基于自适应多尺度与轮廓梯度的遥感图像分割网络

牛梦佳[1]，张永军[1*]，李智[1]，杨刚[2]，崔忠伟[3]，刘竣文[1]

[1]贵州大学计算机科学与技术学院，贵州 贵阳 550025；

[2]贵阳欧比特宇航科技有限公司，贵州 贵阳 550027；

[3]贵州师范学院大数据科学与智能工程研究院，贵州 贵阳 550018

摘要 遥感图像分割算法易受环境因素干扰，如物体遮挡、光照不均匀等。现有的深度学习遥感图像语义分割方法通常采取端到端的编解码结构，但针对相似度较高物体的结构和轮廓，仍存在分割不准确的问题。为了提高算法鲁棒性、分类准确率，提出一种基于轮廓梯度学习的深度卷积神经网络遥感图像语义分割算法。为了提高预测特征图的质量，首先基于SegNet模型，提出自适应注意力的多通道多尺度特征融合网络(D-MMA Net)，其中D-MA block采用基于注意力的自适应多尺度模块，根据学习到的权重自适应地对不同尺度特征进行提取，以获得更多有效的高级语义特征。为进一步细化提取物体的边界，基于Sobel边缘检测算子原理提出可学习的轮廓提取模块。最后将轮廓信息与多尺度语义特征相结合，以增强对图像空间分辨率的鲁棒性。实验结果表明，所提算法提高分割的准确率，对于不规则物体边界，能有良好的分割效果。

关键词 遥感；遥感图像；多通道特征提取；轮廓梯度；特征融合；语义分割

中图分类号 TP391 **文献标志码** A **DOI**：10.3788/LOP220525

Remote Sensing Image Segmentation Network Based on Adaptive Multiscale and Contour Gradient

Niu Mengjia[1], Zhang Yongjun[1*], Li Zhi[1], Yang Gang[2], Cui Zhongwei[3], Liu Junwen[1]

[1]*College of Computer Science and Technology, Guizhou University, Guiyang* 550025, *Guizhou, China*;

[2]*Guiyang Orbita Aerospace Science & Technology Co., Ltd., Guiyang* 550027, *Guizhou, China*;

[3]*Big Data Science and Intelligent Engineering Research Institute, Guizhou Education University, Guiyang* 550018, *Guizhou, China*

Abstract Remote sensing image segmentation algorithms are susceptible to interference from environmental factors, such as object occlusion and uneven illumination. Existing deep learning remote sensing image semantic segmentation methods usually adopt an end-to-end codec structure. However, they still suffer from inaccurate segmentation for the structure and contours of high similarity objects. Therefore, to improve the algorithm robustness and classification accuracy, a deep convolutional neural network remote sensing image semantic segmentation algorithm based on contour gradient learning is proposed. To improve the quality of the predicted feature maps, the adaptive attention-based multichannel multiscale feature fusion network (D-MMA Net) is proposed based on the SegNet model network. The D-MA block uses an attention-based adaptive multiscale module to adaptively extract different scale features according to the learned weights to obtain more effective high level semantic features. To further refine the extracted object boundaries, the contour extraction module, a learnable contour extraction module, is proposed based on the principle of the Sobel edge detection operator. Finally, the contour information is combined with multi-scale semantic features to enhance the robustness of the spatial resolution of the image. The experimental results show that the proposed method improves the segmentation accuracy and produces good segmentation results for irregular object boundaries.

Key words remote sensing; remote sensing image; multi-channel feature extraction; contour gradient; feature fusion; semantic segmentation

收稿日期：2022-01-12；**修回日期**：2022-02-17；**录用日期**：2022-03-14；**网络首发日期**：2022-03-26

基金项目：国家自然科学基金项目(62062023)、贵州省教育厅创新群体研究项目(黔教合KY字[2021]022)

通信作者：*niumj0130@163.com

1 引 言

遥感图像在如今的人工智能时代有着不可忽视的地位。海量的数据与多样化的数据类型使得遥感图像处理技术被广泛地应用于精细农业[1]、城市规划[2]和防震减灾[3]等领域。因此，性能良好的语义分割模型对于遥感图像的实际运用至关重要。遥感图像与普通光学场景图像相比，由于复杂的地形背景与多样式的物体类别，其同时具有类间方差小和类内方差大的复杂特性[4]，同时环境因素的干扰增加了标准图像采集的难度[5-6]，这些使得遥感数据不能得到充分利用。

随着深度学习在人工智能领域的发展，基于深度卷积神经网络（DCNN）的遥感图像语义分割方法起到重要的作用。语义分割方法能够对图像中的每个像素进行解释分类。语义分割通过深层卷积方法进行像素级的分类，使每个像素获得对应的语义标签。目前图像领域绝大多数先进的深度学习算法均是基于端到端的神经网络，即 full convolutional neural network（FCN）[7]。FCN是一种对图像进行端到端的像素级分割网络，从而开启了语义级别的图像分割时代。基于此，研究者构造端到端的编解码框架，其中具有代表性有 U-Net[8]和 SegNet[9]等，使用堆叠卷积核与池化进行编码，解码器通过上采样或者反卷积得到预测图像。近年来，Chen 等[10]提出 Deeplab 系列，其中 Deeplabv3＋在之前网络的基础上引入深层次膨胀卷积和空间金字塔来对任意尺度的区域进行分类。Hu 等[11]则提出 squeeze-and-excitation module，并在此基础上提出 SENet，SENet 通过学习各通道之间依赖性，自动为通道分配不同权重，构建特征通道之间的相互依赖关系进而提升特征的表达。结合 Deeplab 系列的空洞金字塔池化（ASPP）和 DenseNet[12]思想，Yang 等[13]提出 DenseASPP 结构，该结构具有更大的感受野和更密集的采样点，可改善网络的语义分割结果。Yuan 等[14]提出的 HRNet-OCR 并行连接不同分辨率卷积，同时在并行分支中重复进行多尺度融合来维持高分辨率特征表示，分割后加入每个像素与其他像素的关系权重，以提高场景分类精度。

在边缘提取方面，传统方法由于需要根据不同的光谱、纹理、几何等特征设计特征算子[15-16]，鲁棒性不够强。随着深度学习的发展，Cheng 等[17]提出一种边缘感知卷积网络，该网络在卷积网络的基础上，通过添加边缘感知正则化，进一步利用边缘的输出来细化整个模型。Marmanis 等[18]在 DCNN 模型的末尾添加尾端分类器用来提取物体边缘，从而辅助模型进行更好的分割。Takikawa[19]等提出一种新的用于语义分割的双流 CNN 体系结构，它将形状信息作为一个单独的处理分支显式地连接起来，可以较好地预测物体周边，从而提高分类准确率。

尽管上述方法在构建更深更复杂的网络模型进行图像分割时获得较好的效果，但在复杂的遥感图像场景下依然存在诸多挑战。一方面，神经网络中的卷积层较深，对于多尺度的提取往往只是简单叠加，并且卷积核较小且作用于局部，对于较大的目标很难联系上下文获取具有区别的语义信息。同时，现有的基于 DCNN 的网络提取大多通过简单跳跃连接与解码器融合恢复空间定位，然而浅层的特征图包含粗糙的语义信息，这会使得物体分割引入噪声信息[20]。另一方面，深度学习对图像进行下采样容易丢失目标的位置信息，虽然分辨率的降低可能只会影响自然图像中的小物体，但是在正射遥感图像中会出现边界不清晰的问题，导致解码器上采样时目标边界通常很模糊。因此，本文提出一种基于自适应注意力的多尺度多通道特征提取编解码器网络（D-MMA Net），通过分阶段地提取浅层的空间特征，构造深度网络的同时关注网络的宽度，在解码器融合有效的位置信息。此外，引入基于先验知识的轮廓梯度提取模块（CEM），使网络进一步学习有效特征，实现对地物的有效分割。在 Vaihingen 、Potsdam 和 WHU building 数据集上分别进行了验证，所提网络有较好的分割效果。

2 D-MMAE Net模型

本研究结合多尺度多通道特征提取网络和轮廓提取模块对遥感图像进行语义分割，整体网络框架如图 1 所示，主要包括 2 个部分：1 个高效的多尺度特征提取骨干网络和 1 个轮廓学习模块。首先，主干结构采用基于 SegNet 改进的特征提取器，并在编解码器间分阶段融入基于注意力的多尺度特征提取模块（D-MA block），自适应注意力改善多尺度的特征图提取，以多通道的方式将经过学习的不同层级的特征图与解码器特征融合，修改后的特征提取网络具有保留详细多尺度特征的能力，从而达到提高网络的鲁棒性及精度的效果。然后，所设计的轮廓学习模块从梯度信息中学习轮廓，以提取精细的物体边界。

2.1 D-MMA 特征提取网络

遥感图像具有数据信息丰富的固有特性，因此要求卷积神经网络拥有一定的上下文联系。SegNet 是一个经典的编解码器网络，解码器上采样能够保留局部的位置信息，但缺少对于全局的上下文信息的关联，对于遥感图像中不规则物体边界的定位仍有待提高。针对这个问题，基于 SegNet，本研究提出一种基于注意力的多尺度多通道网络。

图 2 为所提 D-MMA Net 框架图。为了更好地恢复物体的形状边缘，减轻因上采样时小卷积核作用于局部导致缺少长距离全局依赖信息的影响，在原有 SegNet 的编码器部分的基础上，通过构造包含膨胀卷积[17]的解码器，使得解码器拥有更大的感受野，高效地建模长距离依赖和位置模式。同时，通过 D-MA block 从浅层的粗语义特征提取更丰富的浅层空间信息，帮

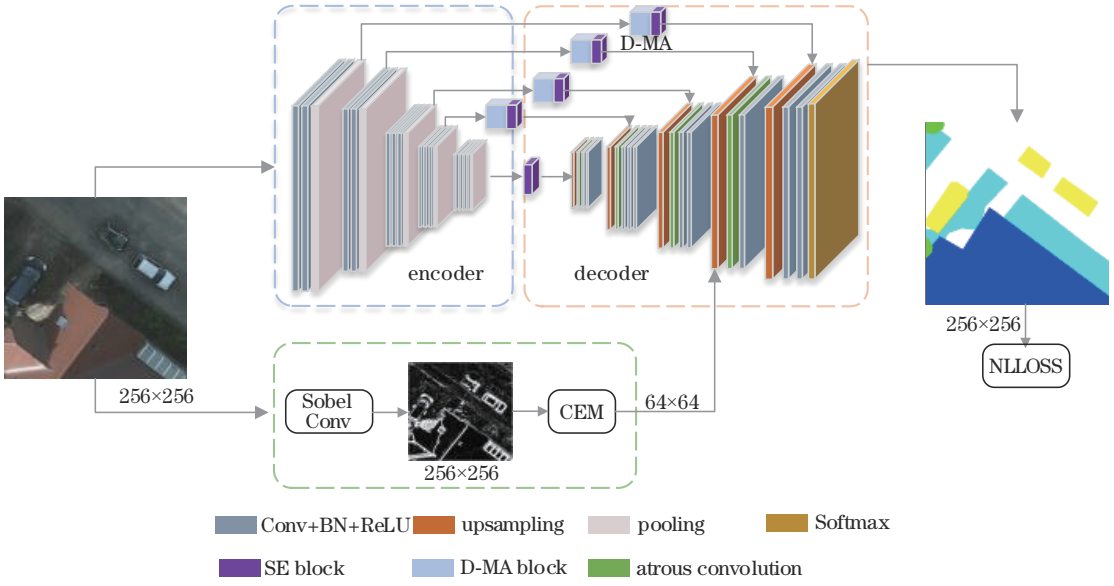

图 1　基于融合轮廓学习的深度卷积神经网络遥感图像语义分割算法框架图

Fig. 1　Framework of remote sensing image semantic segmentation algorithm based on fused contour learning with deep convolutional neural network

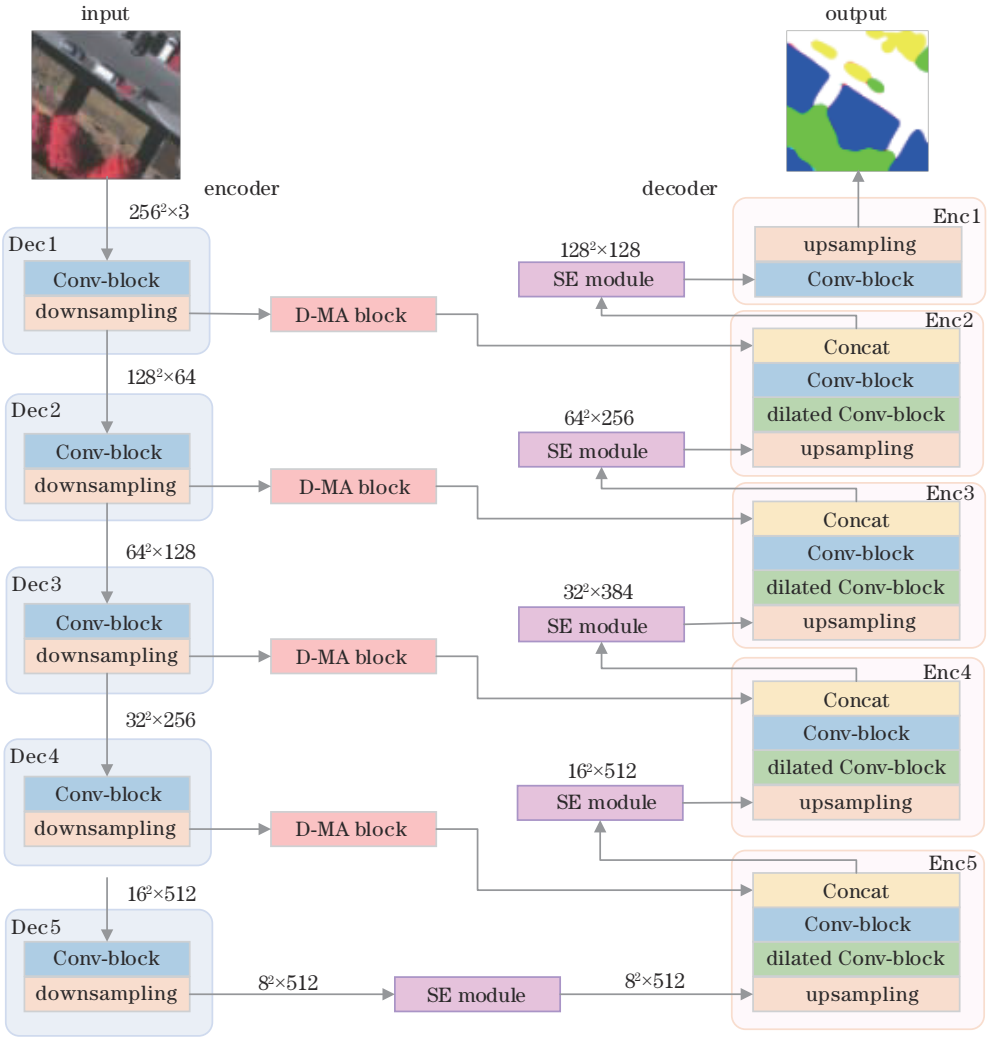

图 2　基于注意力机制的多通道网络框架图

Fig. 2　Multi-channel network framework based on attention mechanism

助模型补充缺失的空间细节信息。随后，将提取到的特征与解码器上采样的部分进行特征融合，经过注意力机制模块（SE module）[11]的作用，重要通道的效用将被放大，有效地将高级语义特征与包含空间信息的浅层语义特征进行特征融合，使得网络拥有良好的分割性能。

在解码器部分，每个解码器块由双线性上采样、膨胀卷积单元与标准卷积单元构成。标准卷积单元（Conv-block）由3×3的卷积核（Conv）、批量归一化层（BN）和ReLU激活函数组成。膨胀卷积单元（dilated Conv-block）由3×3的膨胀卷积核（dilated Conv）、批量归一化层和ReLU激活函数及单个卷积单元组成。在不同阶段的解码器模块中，膨胀卷积单元中的膨胀系数随特征图变大而变大，解码器5～解码器2的膨胀系数依次是2、4、8、8。当扩张率为2时，其感受野为5×5；当扩张率为4时，其感受野为9×9；当扩张率为8时，其感受野为17×17。膨胀卷积可以增加卷积操作的感受野，使更多上下文信息被覆盖，且不增加网络的参数量。

2.2　基于自适应注意力的特征提取模块

多通道特征提取结构由D-MA block在编码器的不同阶段，实现对语义信息由浅到深的特征提取，如图3所示。D-MA block由单个MA block通过密集身份映射连接组成，对原有的卷积进行分解，同时使用大小不同的卷积核对特征图进行特征提取，达到提取更丰富语义信息的效果。MA block引入膨胀卷积代替不同尺度的卷积核，膨胀卷积的膨胀率随着特征图的变小而变小，在不增加参数量的情况下，增大特征图的感受野，使得网络有更高的效率。

在通道注意模块中，通常使用全局平均池化对空间信息进行全局编码后再与自身相乘，但它只将某一尺度的空间信息压缩到通道中，针对不同尺度卷积建立通道依赖关系，可提高对不同感受野信道信息的敏感性。在每个MA block后端使用selective kernel weight module（SK weight module）。SK weight module由分裂、融合和选择等3部分组成。具体而言，首先使用MA block形成3个分支。随后，将各分支进行相加得到融合后的特征F，此时使用全局平均池化获得特征的全局信息S：

$$S_C = F_{gp}(F_C) = \frac{1}{H \times W}\sum_{i=1}^{H}\sum_{j=1}^{W} F_C(i,j), \tag{1}$$

式中：C、H、W分别代表特征图的通道、高和宽。使用两个全连接层（FC）将得到的降维特征图转化为一个收缩参数，并对特征图进行维度调整，使得特征图的大

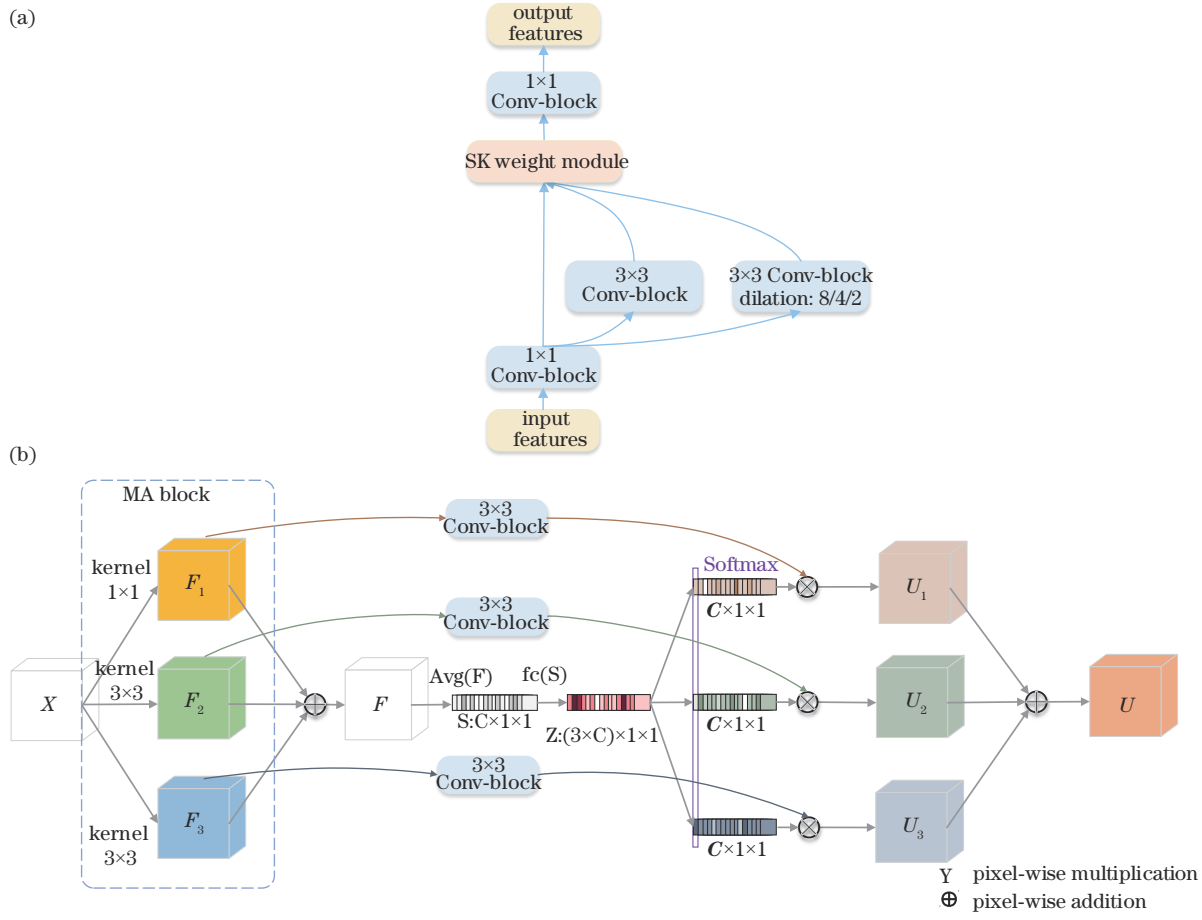

图3　MA block内部结构图。（a）单个MA block的内部结构；（b）SK weight module

Fig. 3　Diagrams of internal structure of MA block. (a) Internal structure of a single MA block; (b) SK weight module

小的维度变为[batch_size, 3×C, 1, 1]。在模块的最后，连接着一个Softmax函数，为了将最终的缩放参数的数值范围框定在(0,1)之间。最终，缩放参数与不同尺度特征图经过Conv-block后的卷积分别相乘，重要的通道被突出表达，获得各尺度分支注意力系数，计算每个尺度对应特征图权重的部分，实现不同尺度自适应的调整，最后连接1×1卷积单元：

$$U_C=\sum_{i=1}^{3}\tilde{F}_i\times F_{\mathrm{Softmax}}\left(F_{\mathrm{fc}}(S_C)\right)=\sum_{i=1}^{3}\tilde{F}_i\times \delta\left\{w_1\left[w_0(S_C)\right]\right\},\tag{2}$$

式中：W_0和W_1分别代表两个全连接层；δ代表Softmax函数；$\tilde{F}$代表不同通道经过标准卷积单元得到的特征图；U_C代表最终经过通道注意力的特征图。

在MA block基础上，进一步得到更丰富的多尺度特征，在DenseNet[12]启发下提出D-MA block。然而过多的层数会引起层数激增，导致特征图出现冗余并增加内存负担，因此每个D-MA unit设置了3层MA block，每一层与之前的所有层拼接，该模块中使用的跳跃连接可以有效地减轻梯度消失和梯度爆炸带来的影响，如图4所示。所提D-MA block包含两个D-MA unit，并在不同阶段连接编码器与解码器。

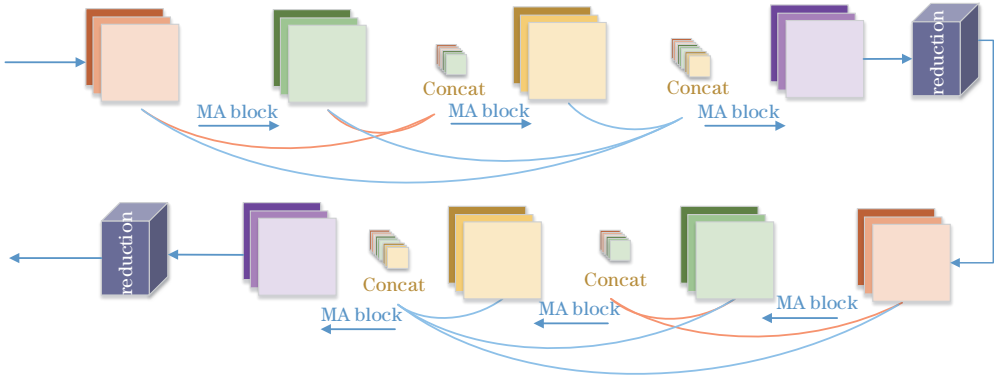

图4　D-MA block内部结构图

Fig. 4　Diagram of internal structure of D-MA block

MA block分别连接在编码器部分中的4个编码块后，形成多通道特征提取D-MMA。编码器的最后一层提取高级的语义信息，因此未采用MA block，而是直接经过通道注意力模块，将提取到的多尺度特征图与解码器对应大小特征图进行级联，补充包含丰富上下文联系的语义信息。通过MA block代替原始网络的跳跃连接，一方面可以弥补简单跳跃连接带来的高层语义与底层语义的语义差异，另一方面使得网络变得更宽，有利于网络提取多样化丰富的语义信息。

2.3　轮廓提取模块

现有的边缘提取方法不能准确地识别边界，特别是对于紧密相邻的物体，物体边缘的信息可能未得到充分利用。另一方面，边界信息的学习对于基于CNN进行地物分割的方法非常重要。本研究将边界检测添加到编解码器体系结构中，构建一个相对简单、高效的模型。

采用引入Sobel算子的卷积层来获取输入图像的梯度信息，该卷积层具有捕捉物体边缘差异的优良特性。将卷积层的卷积核设置为特定的Sobel算子，所使用的Sobel算子为

$$\begin{cases} G_X=\begin{pmatrix}-1 & 0 & 1\\ -2 & 0 & 2\\ -1 & 0 & 1\end{pmatrix}\times F \\ G_Y=\begin{pmatrix}1 & 2 & 1\\ 0 & 0 & 0\\ -1 & -2 & -1\end{pmatrix}\times F \end{cases},\tag{3}$$

式中：G_X和G_Y分别为水平和垂直方向上获取的梯度信息。通过提取输入图像的初始梯度信息，将梯度信息输入由MA block组成的分支网络中获取边缘特征，即所提轮廓提取模块，具体参数如表1所示。将提取的特征作为补充特征连接到分割网络中的解码器部分，通过补充边界信息，可以将初始图像沿水平方向和垂直方向的边缘梯度作用于加强D-MMA网络的特征表达。

表1　轮廓提取模块的配置参数

Table 1　Configuration parameters for profile extraction module

Layer	Output size	Operator	Stride	Size
Branch-1	256×256, 64	MA block	1	1
Branch-2	256×256, 64	MA block	1	1
Down layer-1	128×128, 64	Conv-block	2	1/2
Branch-3	128×128, 64	MA block	1	1/2
Branch-4	128×128, 64	MA block	1	1/2
Down layer-2	64×64, 64	Conv-block	2	1/4
Branch-5	64×64, 128	MA block	1	1/4

3　实验条件与评估方法

3.1　实验数据

使用由国际摄影测量及遥感探测学会(ISPRS)组织发布的Vaihingen数据集与Potsdam数据集[21]。两个数据集均由不透水表明、建筑物、树木、低矮植被、车

辆及背景等6个类别组成。Vaihingen数据集包含33幅大小不一的城镇场景遥感图像，对应33幅语义标签图像，其大小约为2500 pixel×2000 pixel，空间分辨率为0.09 m。本研究仅使用正射影像IRRG图像进行实验，它由近红外、红、绿等3个波段组成，随机挑选6张原始图像作为测试集，2张作为验证集，其余为训练集。Potsdam数据集包含38幅城市遥感图像，对应38幅语义标签图像，其大小均为6000 pixel×6000 pixel，空间分辨率为0.05 m。使用RGB图像进行实验，随机挑选7张作为测试集，3张作为验证集，其余为训练集。

WHU building dataset[22]包括航空和卫星子集，具有相应的图像和标签。该数据集包含建筑与非建筑两类。本研究选择在现有工作中广泛使用的航空子集，每幅图像都有3个波段，分别对应R、G和B，大小为512 pixel×512 pixel，空间分辨率为0.075 m。共有8188幅语义标签图像，训练、测试和验证数据集分别为4736、2416和1036。

与自然图像分类数据集相比，遥感图像数据集较小，因此对数据集进行扩充。Vaihingen数据集和Potsdam数据集按照最小尺寸256对图像进行必要的切割，随后进行数据扩充，对图像进行上下翻转、左右翻转及逆时针旋转90°等操作。最终Vaihingen数据集扩充为训练集10102张、测试集2690张、验证集673张，Potsdam建筑数据集扩充为训练集11399张、测试集2850张、验证集950张，WHU building dataset扩充为训练集18944张、测试集9664张、验证集4144张。

3.2 实验参数设置

为了使所提模型达到最优，使用自适应学习率算法Adam训练，学习率设置为0.0005，批量大小为32，动量系数设置为0.9，以此来加速模型收敛。此外，使用惩罚项系数为0.0001的L2正则化以降低过度拟合，达到提高精度的效果。该神经网络模型部署在NVIDIA Tesla V100(32 GB RAM)服务器上，采用CUDA 10.0，以PyTorch 1.8作为开发框架。训练后，选取评价指标最优的模型进行测试。

3.3 评价指标

为了公平客观地定量评估所提网络模型的分割性能，使用公认的语义分割评价指标：平均交并集(mIoU)、总体类别准确度(OA)和F1-score：

$$R_{\mathrm{mIoU}}=\frac{1}{k+1}\sum_{i=0}^{k}\frac{N_{\mathrm{TP}}}{N_{\mathrm{TP}}+N_{\mathrm{FP}}+N_{\mathrm{FN}}},\tag{4}$$

$$A_{\mathrm{OA}}=\frac{N_{\mathrm{TP}}+N_{\mathrm{TN}}}{N_{\mathrm{P}}+N_{\mathrm{N}}},\tag{5}$$

$$S_{\mathrm{F1}}=2\times\frac{P\times R}{P+R},\tag{6}$$

$$P=\frac{N_{\mathrm{TP}}}{N_{\mathrm{TP}}+N_{\mathrm{FP}}},\tag{7}$$

$$R=\frac{N_{\mathrm{TP}}}{N_{\mathrm{TP}}+N_{\mathrm{FN}}},\tag{8}$$

式中：N_{P}表示分类器判定为正样本的数量；N_{N}表示分类器判定为负样本的数量；N_{TP}代表真阳性的数量；N_{FP}代表假阳性的数量；N_{FN}代表假阴性的数量；N_{TN}代表真阴性的数量；$k+1$是总体类别个数(包含背景)。

4 实验结果与分析

对所提模型改进效果进行消融实验，并将其分别与主流语义分割模型U-Net、SegNet、ERFNet[23]和PSPNet[24]在Vaihingen数据集及Potsdam数据集上进行对比。

4.1 D-MMA消融实验

在小节中，对所提模型进行消融实验，具体讨论改进后的语义分割模型(improved SegNet)、自适应多尺度提取特征(with D-MMA)及轮廓提取模块(with CEM)的实验效果。在Vaihingen数据集进行消融实验，并使用F1-score，OA和mIoU这3个指标进行评估，结果如表2所示。

表2 在Vaihingen Vaihingen数据集上的模型消融实验

Table 2 Model ablation experiments on Vaihingen dataset

Model	mIoU / %	OA / %	F1 / %	Params	FLOPs / 10^9
SegNet	76.50	87.96	86.99	20.73	27.56
Improved SegNet	80.69	88.63	89.81	20.91	28.29
with D-MMA	84.28	89.93	90.36	21.82	31.39
with D-MMA +CME	86.56	92.93	92.51	22.69	34.36

表2数据表明，各模型分割效果逐步提升。改进的语义分割模型在mIoU值上提升4.19个百分点，在不增加网络参数量的条件下，获得更多感受野，使得更多上下文语义信息被卷积核覆盖，有助于对不规则物体边缘的有效预测。而使用多通道进行多尺度提取各编码器层特征在mIoU值上提升3.59个百分点，与轮廓提取模块结合的mIoU值为86.56%，最高。可见轮廓提取模块可帮助特征提取网络识别更精确的物体边界，可以帮助小物体更好地补充空间信息，图5中车辆的分割边界优于其他模型，图5为在Vaihingen测试集上的消融实验可视化。此外，OA值与F1值也呈现逐步升高的趋势，这与mIoU值一致。因此，采用该模型进行后续的对比实验。此外，表2列出相关对比数据，还对比了它们的参数量(Params)和浮点运算数(FLOPs)。

4.2 在Vaihingen数据集上的对比

在Vaihingen数据集上的对比如表3所示，分别列出了各模型的F1-score，OA和mIOU指标及每个类分别的IoU值。从表3可以看出：与另外4种主流的语义

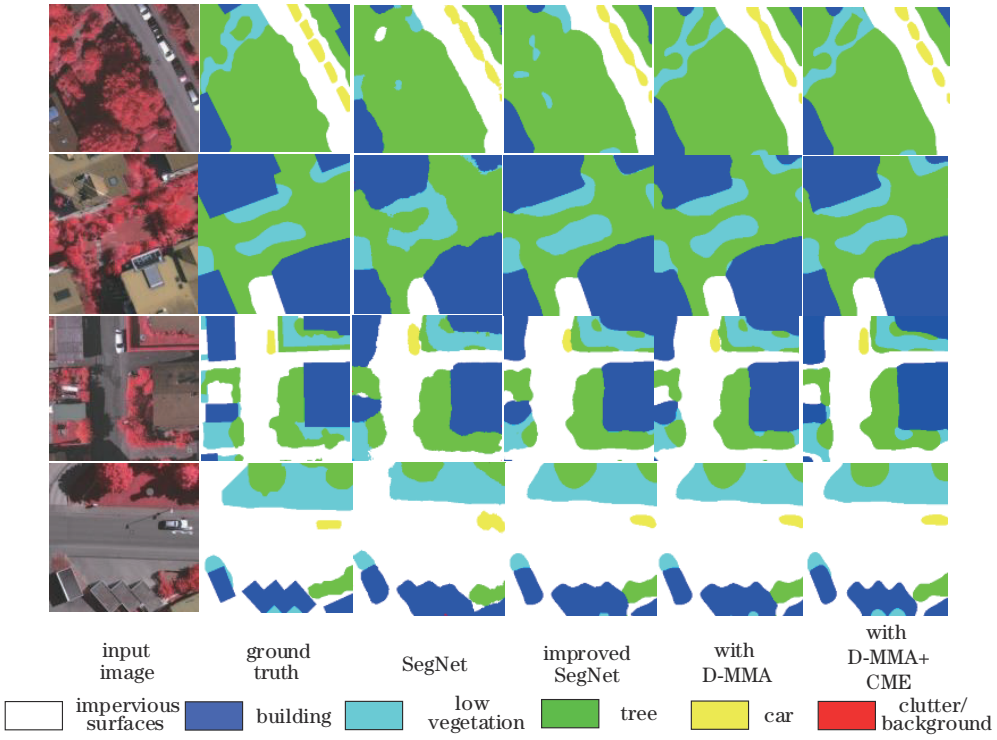

图 5　在 Vaihingen 测试集上消融实验可视化图对比

Fig. 5　Visualization of ablation experiments compared in Vaihingen test set

表 3　在 Vaihingen 数据集上与其他网络对比

Table 3　Comparison with other networks on Vaihingen dataset　　unit: %

Model	IoU					F1	OA	mIoU
	impervious surfaces	building	low vegetation	tree	car			
U-Net	79.45	85.23	64.93	74.77	38.51	86.94	85.43	68.58
SegNet	81.69	86.41	73.43	78.36	42.63	86.99	87.96	76.50
ERFNet	77.51	79.27	62.35	71.27	35.29	83.57	82.09	65.13
PSPNet	87.69	91.94	81.52	84.79	55.79	89.86	90.19	81.16
DSMNet[25]						90.82	91.5	
Fres- MFDNN[26]						92.0	91.0	85.0
Proposed model	91.89	93.61	86.72	88.69	71.93	92.51	92.93	86.56

分割网络相比，所提模型在 mIoU、OA 和 F1-score 上均有所提高；与原 SegNet 相比，所提模型的 mIoU、OA、F1-score 分别提高 5.52 个百分点、4.97 个百分点、10.06%；同时与最近的方法 DSMNet[25] 和 Fres-MFDNN[26] 相比也有较好的分割表现。即基于注意力的多通道分割方法可以提高分割性能。

图 6 为相应的预测效果图。从表 3 和图 6 可以看出，所提模型对不同对象进行语义分割时，分割错误区域较少，更接近真实值，尤其是对于小尺寸物体，比如车辆 IoU 值均高于主流网络。可见基于轮廓学习的多通道分割模型是有效的，有助于补充缺失小物体的空间信息的。此外对于类间相似性较大的问题，如低矮植被和树，所提模型的 IoU 值分别为 86.72% 和 88.69%，也优于其他主流分割模型。模型中的多尺度特征提取丰富了不同类别的语义信息，轮廓学习模块针对物体边缘的特征进行学习，帮助解码器完成有效的特征融合过程，有助于预测类间相似性较高的物体。

4.3　在 Potsdam 数据集上的对比

为了进一步验证所提模型的有效性和泛化能力，在 Potsdam 数据集上继续开展验证实验。所提模型与主流模型的对比结果如表 4 所示。所提模型的 F1、OA 和 mIoU 分别为 91.65%、92.18% 和 82.28%，对比 SegNet 分别提高 6.34 个百分点、4.73 个百分点和 5.24 个百分点。此外，还将所提模型与最新的模型 BAM-Unet-sc[25] 与 ResUNet-a [26] 进行对比，结果显示 OA 有最好的效果，F1 的值略低于 ResUNet-a。

图 7 为相应的预测效果。在对小尺寸物体进行

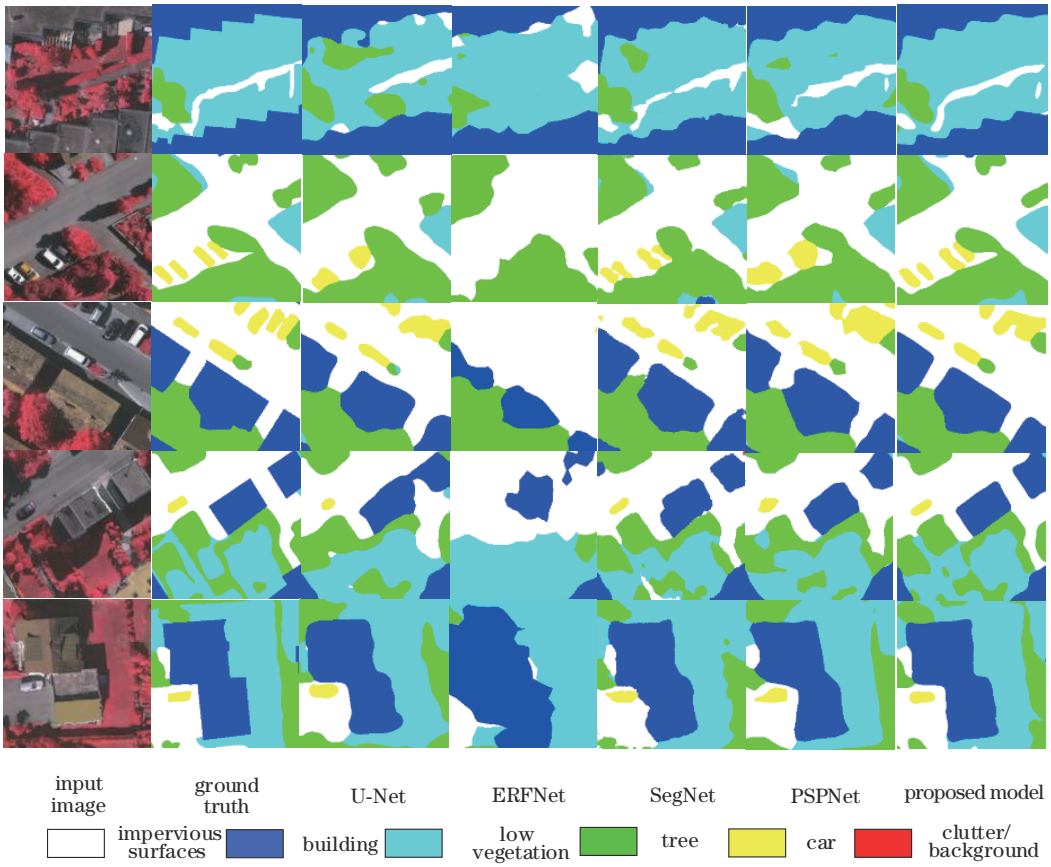

图 6　在 Vaihingen 测试集上与主流模型的预测效果图对比

Fig. 6　Comparison of prediction results with mainstream models on Vaihingen test set

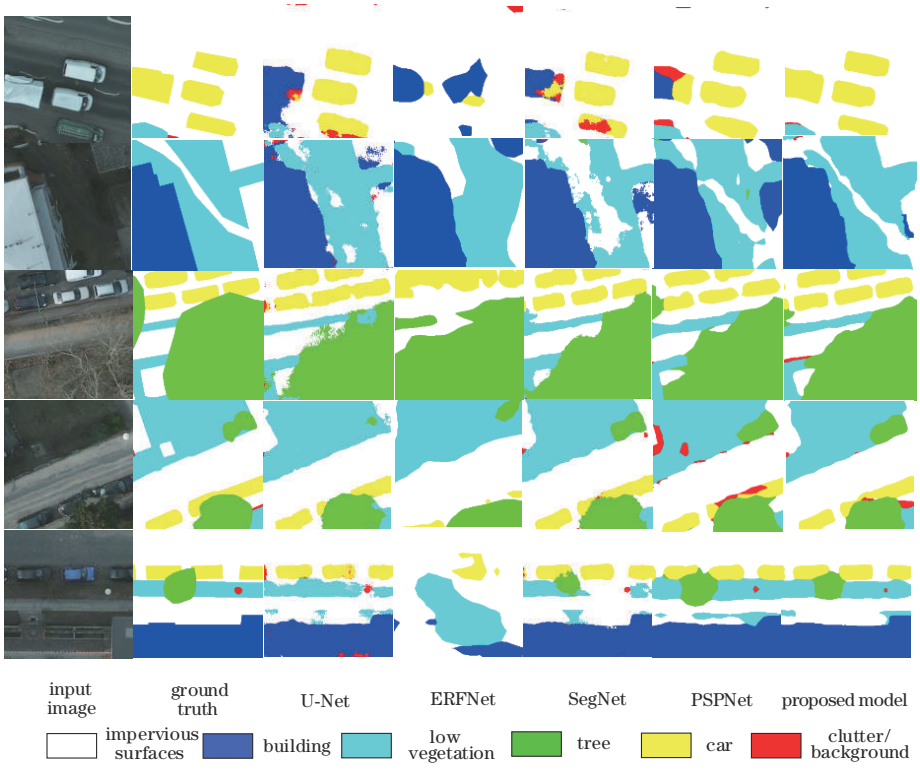

图 7　在 Potsdam 测试集上与主流模型的预测效果图对比

Fig. 7　Comparison of prediction results with mainstream models on Potsdam test set

表 4　在 Potsdam 数据集上与其他网络的对比

Table 4　Comparison with other networks on Potsdam dataset　　unit: %

Model	IoU					F1	OA	mIoU
	impervious surfaces	building	low vegetation	tree	car			
U-Net	76.44	83.22	65.92	58.03	71.87	83.12	77.45	71.10
SegNet	83.63	91.72	74.70	71.67	78.00	85.31	87.45	77.94
ERFNet	61.46	74.78	51.83	45.85	17.07	80.57	72.35	50.20
PSPNet	83.59	92.99	76.28	73.09	77.11	85.78	89.78	80.61
BAM-Unet-sc[27]						88.59	89.13	
ResUNet-a[3]						92.09	91.50	
Proposed model	85.39	93.64	78.82	76.48	81.57	91.65	92.18	83.18

预测时，其余模型均存在较多的误分类情况，即红色区域。同时在类间相似方面，所提模型也达到最优，低矮植被和树的 IoU 值分别为 78.82% 和 76.78%。对比 Vaihingen 数据集，Potsdam 数据集的输入图像颜色区分度较低，这对于模型的学习拟合能力要求更高。

简单地融合从 UNet 和 SegNet 等编码器部分的浅层提取的高分辨率特征图，在解码器部分会同时引入噪声信息。因此，对不规则物体和小尺寸物体的精确定位会造成困难。所提 D-MMA block 采用自适应提取多尺度对不同层进行特征提取，能够提取更多包含细节的语义特征。多通道结合注意力机制将更多浅层的空间信息有效地进行特征融合，同时 CEM 模块针对边缘梯度进行学习，有助于不规则物体边缘和小样本物体边界的预测。

4.4　在 WHU building dataset 上的对比

除了上述两个数据集外，还在 WHU building dataset 上进一步开展了验证实验，结果如表 5 所示。从表 5 可以看出，所提模型与 SegNet 相比，F1-score，OA 和 mIOU 分别提高 3.54 个百分点、1.53 个百分点、2.08 个百分点。同时，从可视化图（图 8）可以看到，所提模型可以更准确地提取建筑物，更好区分背景与建筑，尤其是在建筑物密集的区域，提取的边界较为清晰。DeNet[28] 采用 Inception 思想并用于下采样中，同时使用高密度上采样模块，使网络能够在特征图中编码空间信息；MA-FCN[29] 采用特征金字塔网络（FPN）

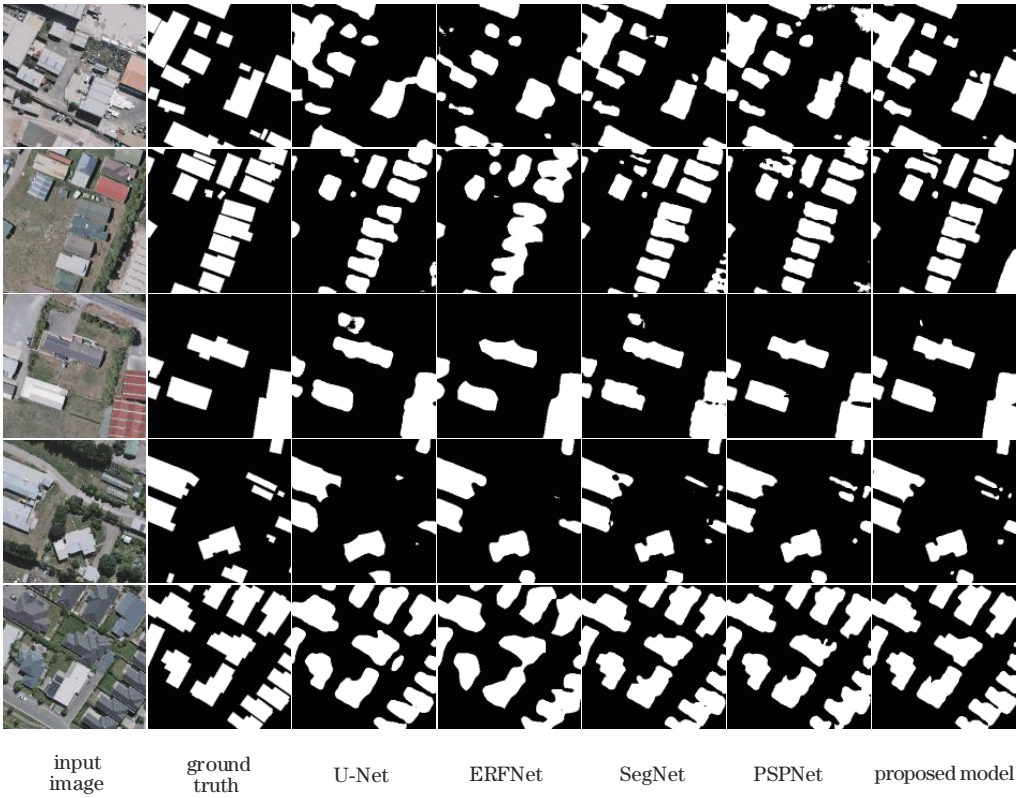

图 8　在 WHU building 测试集上与主流模型的预测效果图对比

Fig. 8　Comparison of prediction results with mainstream models on WHU building test set

表 5 在 WHU building dataset 上与其他网络各类指标对比

Table 5 Comparison with other network indicators on WHU building dataset unit: %

Model	F1	OA	mIoU
U-Net	89.33	87.59	83.01
SegNet	93.73	92.53	88.76
ERFNe	89.33	87.59	78.72
PSPNet	93.28	92.15	87.26
DeNet[28]	94.80		90.12
MA-FCN[29]	95.15		90.70
Proposed model	95.81	94.06	92.30

进行多尺度特征提取，采用多边形正则化策略进行边界优化。所提模型的 mIoU 要高于 DeNet 的 90.12% 和 MA-FCN 的 90.70%，这意味着以边界优化作为补充信息可以有效地提高分割精度。

5 结 论

针对遥感图像由于地面信息复杂而导致小尺寸物体和相似物体分类错误的问题，设计一种基于轮廓学习的多通道遥感图像语义分割网络 D-MMA Net。该网络采用编解码器的结构，设计了含有膨胀卷积的解码器的特征提取网络，同时设计了自适应多尺度特征提取模块，分别在编码器网络的不同阶段提取空间信息，此外以轮廓学习模块为分支作为补充信息，在解码器阶段使用通道注意力机制对特征进行有效融合。先后在 Vaihingen 数据集、Potsdam 数据集及 WHU building dataset 上验证了该方法，实验结果表明，所提模型具有较好的分割性能。对比其他主流模型，该模型可以满足现阶段遥感图像实际应用的需求。下一步将在分割精细边缘上进一步研究，可以从多尺度并行卷积增强特征图分辨率入手。

参 考 文 献

[1] Sishodia R P, Ray R L, Singh S K. Applications of remote sensing in precision agriculture: a review[J]. Remote Sensing, 2020, 12(19): 3136.

[2] Diakogiannis F I, Waldner F, Caccetta P, et al. ResUNet-a: a deep learning framework for semantic segmentation of remotely sensed data[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2020, 162: 94-114.

[3] 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.

[4] Liu F S, Wang Q. A sparse tensor-based classification method of hyperspectral image[J]. Signal Processing, 2020, 168: 107361.

[5] 龚希, 陈占龙, 吴亮, 等. 用于高分辨遥感影像场景分类的迁移学习混合专家分类模型[J]. 光学学报, 2021, 41(23): 2301003.
Gong X, Chen Z L, Wu L, et al. Transfer learning based mixture of experts classification model for high-resolution remote sensing scene classification[J]. Acta Optica Sinica, 2021, 41(23): 2301003.

[6] 朱淑鑫, 周子俊, 顾兴健, 等. 基于RCF网络的遥感图像场景分类研究[J]. 激光与光电子学进展, 2021, 58(14): 1401001.
Zhu S X, Zhou Z J, Gu X J, et al. Scene classification of remote sensing images based on RCF network[J]. Laser & Optoelectronics Progress, 2021, 58(14): 1401001.

[7] Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(4): 640-651.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] Huang G, Liu Z, van der Maaten L, et al. Densely connected convolutional networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 2261-2269.

[13] Yang M K, Yu K, Zhang C, et al. DenseASPP for semantic segmentation in street scenes[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: 3684-3692.

[14] Yuan Y H, Chen X L, Wang J D. Object-contextual representations for semantic segmentation[M]//Vedaldi A, Bischof H, Brox T, et al. Computer vision-ECCV 2020. Lecture notes in computer science. Cham: Springer, 2020, 12351: 173-190.

[15] Li Z B, Shi W Z, Wang Q M, et al. Extracting man-made objects from high spatial resolution remote sensing images via fast level set evolutions[J]. IEEE Transactions

on Geoscience and Remote Sensing, 2015, 53(2): 883-899.

[16] Liasis G, Stavrou S. Building extraction in satellite images using active contours and colour features[J]. International Journal of Remote Sensing, 2016, 37(5): 1127-1153.

[17] Cheng D C, Meng G F, Xiang S M, et al. FusionNet: edge aware deep convolutional networks for semantic segmentation of remote sensing harbor images[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2017, 10(12): 5769-5783.

[18] Marmanis D, Schindler K, Wegner J D, et al. Classification with an edge: improving semantic image segmentation with boundary detection[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2018, 135: 158-172.

[19] 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), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 5228-5237.

[20] Zhu Q, Liao C, Hu H, et al. MAP-net: multiple attending path neural network for building footprint extraction from remote sensed imagery[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 59(7): 6169-6181.

[21] International Society for Photogrammetry and Remote Sensing. 2D semantic labeling contest[EB/OL]. [2021-05-06]. https://www. isprs. org/education/benchmarks/UrbanSemLab/semantic-labeling.aspx.

[22] Ji S P, Wei S Q, Lu M. Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery data set[J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57(1): 574-586.

[23] Romera E, Álvarez J M, Bergasa L M, et al. ERFNet: efficient residual factorized ConvNet for real-time semantic segmentation[J]. IEEE Transactions on Intelligent Transportation Systems, 2018, 19(1): 263-272.

[24] 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.

[25] Cao Z Y, Fu K, Lu X D, et al. End-to-end DSM fusion networks for semantic segmentation in high-resolution aerial images[J]. IEEE Geoscience and Remote Sensing Letters, 2019, 16(11): 1766-1770.

[26] 张小娟, 汪西莉. 完全残差连接与多尺度特征融合遥感图像分割[J]. 遥感学报, 2020, 24(9): 1120-1133.
Zhang X J, Wang X L. Image segmentation models of remote sensing using full residual connection and multiscale feature fusion[J]. Journal of Remote Sensing, 2020, 24(9): 1120-1133.

[27] Nong Z X, Su X, Liu Y, et al. Boundary-aware dual-stream network for VHR remote sensing images semantic segmentation[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2021, 14: 5260-5268.

[28] Liu H, Luo J C, Huang B, et al. DE-net: deep encoding network for building extraction from high-resolution remote sensing imagery[J]. Remote Sensing, 2019, 11(20): 2380.

[29] Wei S Q, Ji S P, Lu M. Toward automatic building footprint delineation from aerial images using CNN and regularization[J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 58(3): 2178-2189.