激光与光电子学进展 先进成像

基于多模块的遥感影像建筑物提取方法

明兴涛，杨德宏*

昆明理工大学国土资源工程学院，云南 昆明 650093

摘要 高分辨率遥感影像建筑物提取是遥感影像解译的一个重要研究方向。针对传统提取方法中小型建筑物容易丢失和大型建筑物边界模糊问题，以Unet为基础，提出一种基于多模块建筑物提取网络(MM-Unet)。首先在网络的编码和解码部分引入多尺度特征组合模块(MFCM)以获取并补充更多的空间信息。之后在解码器末端加入多尺度特征增强模块(MFEF)以增强多尺度特征的提取。在跳跃连接完成后引入双重注意力模块(DAM)，使网络能够自适应地学习通道和空间位置的特征重要性，减小不同深度特征的差异。为了验证所提网络的有效性，分别在空间分辨率为1 m、0.3 m、0.09 m的Massachusetts、WHU以及Vaihingen建筑物数据集上进行实验，MM-Unet的交并比分别达到73.42%、90.11%和85.21%，相比于Unet分别提高2.21个百分点、1.25个百分点和1.55个百分点。结果表明，MM-Unet对于不同尺度的建筑物表现出较高的提取精度和较强的泛化能力。

关键词 遥感影像；建筑物提取；多尺度；注意力机制；特征组合

中图分类号 TP751.1 **文献标志码** A **DOI**：10.3788/LOP231148

Building Extraction from Remote Sensing Image Based on Multi-Module

Ming Xingtao, Yang Dehong*

Faculty of Land and Resources Engineering, Kunming University of Science and Technology, Kunming 650093, *Yunnan, China*

Abstract Building extraction from high-resolution remote sensing imagery is an important research direction for the interpretation of remote sensing imagery. To address the issues of small buildings easily lost and large buildings with blurred boundaries by traditional extraction methods, this paper proposes a multi-module building extraction U-shaped network (MM-Unet) based on Unet. First, Multi-scale feature combination module (MFCM) is introduced in the encoder and decoder sections of the network to obtain and supplement more spatial information. Then, multi-scale feature enhancement module (MFEF) is incorporated at the end of the decoder to enhance the extraction of multi-scale features. After the skip connections, the dual attention module (DAM) is introduced to adaptively learn the feature importance of channel and spatial positions, thereby reducing the differences among features at different depths. In order to validate the effectiveness of the network, experiments are conducted on Massachusetts, WHU, and Vaihingen building datasets with spatial resolutions of 1 m, 0.3 m, and 0.09 m respectively. and the intersection and union ratio of MM-Unet reach 73.42%, 90.11%, and 85.21%, compared to Unet, increased by 2.21 percentage points, 1.25 percentage points , and 1.55 percentage points. These results demonstrate that MM-Unet shows high extraction accuracy and strong generalization ability on buildings of various scales.

Key words remote sensing image; building extraction; multi-scale; attention mechanism; feature combination

1 引 言

遥感影像建筑物提取研究对于城市现代化具有重要的应用价值，通过高效、准确的建筑物提取算法，可以快速获取城市中建筑物的分布、类型等信息，为城市规划、土地利用规划、环境监测等提供数据支持，为城市管理和决策提供科学依据。然而，建筑物在尺度、形态和结构方面的多样性和复杂性等问题使得建筑物提取变得复杂和困难。同时，遥感影像中的建筑物与环境背景也存在颜色、纹理、形状的相似性，这也给建筑物提取带来了挑战[1]。

目前，从高分辨率遥感影像中提取建筑物的方法

收稿日期：2023-04-23；**修回日期**：2023-05-17；**录用日期**：2023-06-01；**网络首发日期**：2023-06-11

通信作者：*1486097650@qq.com

主要有传统方法和基于深度学习的方法。传统方法主要包括特征提取、分类、阈值分割等步骤[2]，例如：谭衢霖等[3]利用多尺度分析技术实现遥感影像的自适应分割；陈行等[4]结合形态学分析和阈值分割技术，实现对建筑物的自动提取；Huang 等[5]提出一种能够同时考虑建筑物和阴影信息的形态学建筑物指数。此外还有决策树[6]、随机森林[7]、支持向量机[8]等机器学习方法也经常被使用。然而传统方法的精度受到特征表示能力以及分类器性能的限制，难以满足场景复杂度高的遥感影像建筑物提取需求[9]。

在大数据时代，基于神经网络的深度学习方法已在计算机视觉领域得到广泛的应用。相比于传统方法，神经网络拥有较好的适用性，并且能够更充分地利用图像信息。2015 年，Long 等[10]提出全卷积神经网络（FCN），将原本的全连接层改为卷积层，使得网络可以处理任意大小的图像，此后 Unet[11]、ResNet[12]、SegNet[13]和 DeepLab[14-15]系列也紧跟着出现。Unet 因其卓越的效果而被广泛应用，很多研究使用 Unet 作为基础架构，并对其进行改进，以更好地应用于建筑物提取领域。文献[16]提出的 Attention Unet 将注意力机制应用于 Unet 以增强特征的重要性，并降低噪声的影响。文献[17]提出的 Unet＋＋将原本的编码器-解码器结构改为多级嵌套，增加了跨层级别的信息传递和融合，使网络具有更好的表达能力和更高的效率。为了解决尺度多样性以及不同层级特征之间融合所存在的语义鸿沟，文献[18]提出的 MultiResUnet 在 Unet 中加入 MultiRes 模块，能够提取不同分辨率下的特征信息，该模型利用残差连接提高了模型的学习能力，并采用了多分辨率的特征提取方式，以更好地处理不同尺度的建筑物目标。文献[19]提出的 Mnet 引入了多分支机制，能够更好地利用图像中的多尺度信息，在提取建筑物全局特征的同时保留局部特征，从而提高建筑物提取的准确性和鲁棒性。

上述网络模型在建筑物提取中精度有明显提升，但仍存在一些不足：模型的全局特征不够充分，容易陷入局部特征中，导致空洞现象；不同深度的特征无法很好地利用，当存在和目标物尺度一致的物体或者目标物尺度过小时，容易出现错分、漏分和边界模糊现象；在连续的下采样后，模型容易丢失多尺度信息，导致提取精度的下降。为此，本文以 Unet 为框架，提出一种基于多尺度特征增强模块、双重注意力模块以及多尺度特征组合模块的高分辨率遥感影像建筑物提取网络（MM-Unet）来实现建筑物精确提取。

2　研究方法

2.1　MM-Unet 结构

所设计的 MM-Unet 高分辨率遥感影像建筑物提取网络如图 1 所示，该网络以 UNet 为基本框架，主要包括 3 个模块：多尺度特征增强模块、双重注意力模块以及多尺度特征组合模块。

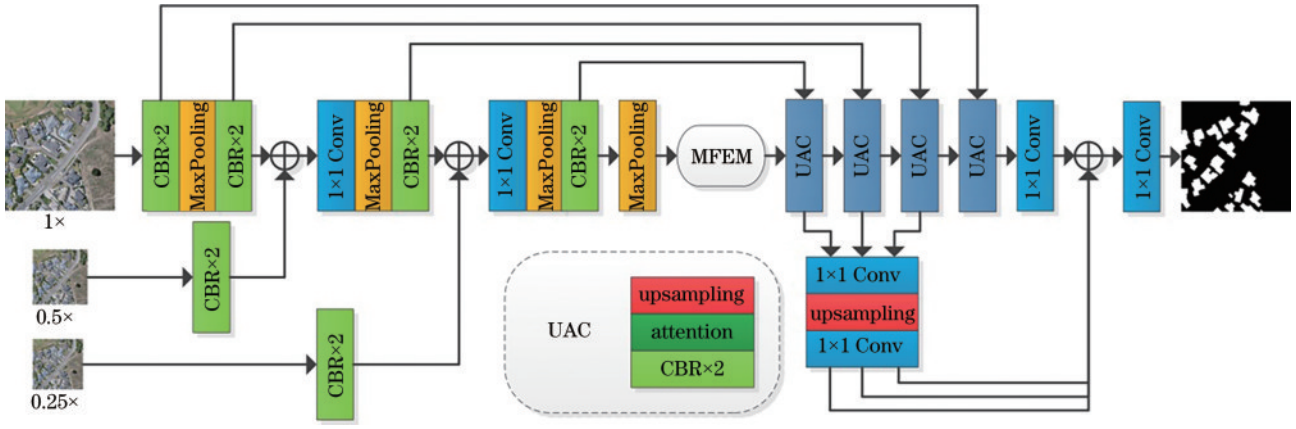

图 1　MM-Unet 结构图

Fig. 1　MM-Unet structure

在模型的编码部分构造多尺度输入，将 3 个不同尺度的输入图像进行卷积处理后，叠加输入网络相对应解码器层中，以增强多尺度信息的获取和实现不同层级感受野的融合。在模型的编码器和解码器之间添加多尺度增强模块，通过在多个不同的空洞卷积中进行特征提取，捕获不同尺度的上下文信息，以获得更加广泛的感受野，提高网络对多尺度特征的感知能力。在模型的跳跃连接完成后，引入双重注意力模块，可以由特征图中不同通道以及不同空间的重要性调整特征的权重，帮助网络更加准确地捕捉建筑物的特征，并抑制背景干扰，从而提高建筑物提取的准确率。在模型的解码部分构造多尺度输出，将每个尺度的特征图通过卷积和双线性插值上采样输出并组合起来，这样就能够有效地补充不同层次特征图的信息，实现对多尺度建筑物的准确提取。

2.2　残差单元

在传统的神经网络中，随着网络层数的增加，反向传播的梯度会不断缩小，导致训练过程变得困难，也就

是深度网络的退化问题，这是梯度消失或梯度爆炸造成的。ResNet[12]的残差单元是通过在网络中引入跨层连接来实现的，即将输入信号直接加到输出信号中，而不是简单地将输入信号作为输出信号的函数输出，这样就能够保留浅层特征的信息，使得深层网络可以更容易训练。残差单元的表达式如式(1)所示，其中，x是输入信号，y是输出信号，F是由权重参数w定义的非线性变换，也就是残差单元的输出结果。

$$y=F(w,x)+x。\tag{1}$$

与普通的神经单元[图 2(a)]相比，使用残差单元的网络在训练过程中能够更加快速地收敛，并且在保持较高准确率的同时，网络更深，从而提升网络的表征能力。所以图 1 中的 CBR×2 卷积块均使用了残差单元，具体结构如图 2(b)所示。

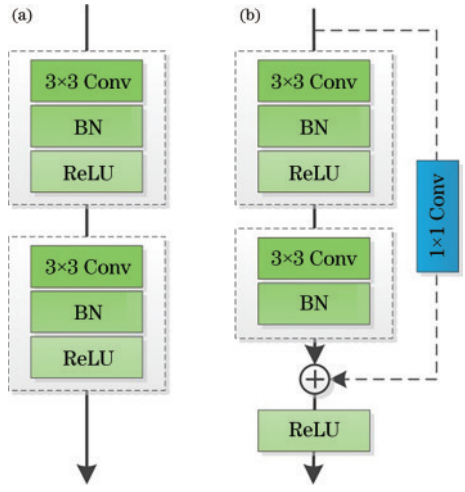

图 2 普通单元与残差单元。(a)普通单元；(b)残差单元

Fig. 2 Plain unit and residual unit. (a) Plain unit; (b) residual unit

2.3 多尺度特征组合模块

为了从遥感图像中准确提取建筑物信息，需要考虑遥感图像中建筑物分布的特点，包括边界是否复杂、尺度差异是否明显和背景是否造成影响等因素。在编码层次对于不同尺度的建筑物的特征信息提取需要不同大小的感受野，所以想要实现多尺度建筑物特征提取，就需要在编码层次对网络中的感受野进行不断调整。而在解码层次对于不同尺度的建筑物提取需要不同深度的特征信息，所以想要实现多尺度建筑物提取，就需要获取解码层次中各个深度的解码特征。传统的 Unet 采用连续的卷积以及下采样操作，容易丢失输入特征图的部分空间信息，之后又采用连续的卷积以及上采样操作，很难得到多个深度的特征信息。

针对以上问题，设计了多尺度特征组合模块，包括多尺度输入和输出，如图 1 所示，多尺度输入设计在编码层次，每张输入的影像通过 1/2 的下采样和 1/4 的下采样得到原始图像、0.5× 图像和 0.25× 图像，经过 CBR×2 残差单元和最大池化的原始图像，再经过 CBR×2 残差单元操作后，与经过 CBR×2 残差单元的 0.5× 图像叠加起来，接下来经过 1×1 的卷积调整通道数得到新的特征图。之后根据同样的步骤使新特征图与 0.25× 图像组合起来。通过以上操作，不但能够获得不同大小的感受野，而且还能保留较多的空间信息。多尺度输出设计在解码层次，将经过前 3 个 UAC 模块后的特征图，用 1×1 的卷积调整通道数与最后一个 UAC 模块特征图的通道数一致，接下来 3 个特征图分别进行 8 倍、4 倍、2 倍上采样使特征图大小一致，然后 4 个特征图都经过 1×1 的卷积调整通道数为 2，并叠加起来再调整通道数为 2，得到最终的预测结果。通过以上的操作，能够在不改变特征感受野的条件下，将不同深度但是信息相似的特征组合起来，以识别不同尺度的建筑物。

2.4 多尺度特征增强模块

在建筑物提取任务中，网络需要同时考虑不同尺度下的建筑物特征，以便能够更好地捕捉到目标信息。但是，传统的卷积神经网络在经过不断的卷积和池化操作后会出现信息瓶颈问题，使得网络只能考虑局部区域的特征，难以捕捉到全局上下文信息，造成在建筑物提取中出现的空洞现象和忽略现象。为了解决这个问题，Chen 等[14]提出空洞空间金字塔池化(ASPP)，如图 1(a)所示，ASPP 将不同空洞率的空洞卷积应用于特征图，得到一系列不同尺度的特征图。然后，这些特征图通过池化操作融合，最后得到来自不同感受野的多尺度特征表示。其中，空洞卷积是一种卷积操作，可以在保持网络参数和卷积核大小不变的情况下，通过扩大卷积核的间隔来改变感受野的大小。

但是 ASPP 使用了较大空洞率的空洞卷积，会导致卷积核感受野扩大过快，无法充分捕捉局部特征信息，进而影响网络性能，并且由于采样的步长增大，卷积操作只能在输入信号的稀疏采样点上进行，使得相距较远的输入点之间的相关性变弱，这可能导致特征提取时的信息损失和退化。所以在 ASPP 和 ResNet 的启发下，设计了多尺度特征增强模块，将其放置在编码器与解码器之间，也就是图 1 中的 MFEM 对应位置处，具体结构如图 3(b)所示，多尺度特征增强模块在 ASPP 的基础上把第 1 个 1×1 卷积改为空洞率为 1 的 3×3 卷积，其他 3 个卷积的空洞率改为 3、5、7。同时为了防止网络训练中多层卷积堆叠引起的梯度消失和梯度爆炸现象影响提取结果，设计了与残差单元相似的一个分支，把输入的特征图通过 1×1 的卷积调整通道数后与叠加输出后的特征图相加，得到最终结果。多尺度特征增强模块使用了较为合适空洞率的空洞卷积，并且利用残差单元结构保证了梯度反向传播，因此多尺度特征增强模块可以增强网络对于建筑物尺度变化的适应性。

2.5 双重注意力模块

在传统网络中由于卷积和池化操作的存在，输入图像的分辨率在网络中逐渐降低，在较深层的特征图

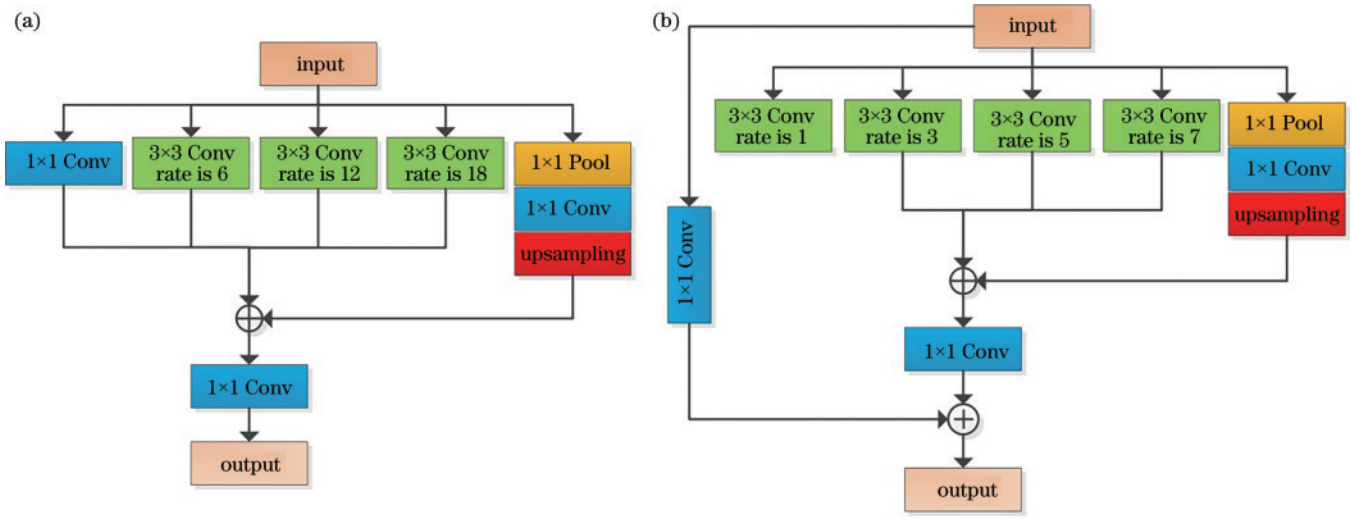

图 3　模块改进。(a)空洞空间金字塔池化模块;(b)多尺度特征增强模块

Fig. 3　Module improvements. (a) Atrous spatial pyramid pooling module; (b) multi-scale feature enhancement module

中可能会丢失一些细节信息。所以 Unet 提出了跳跃连接,在跳跃连接过程中,通过将编码器的浅层特征与解码器的深层特征相结合,能够恢复提取过程中的空间信息,网络可以更好地处理建筑物在不同尺度上的变化。使用不同深度的特征图可以提高提取结果的准确性,但是由于每个深度的特征图经历了不同数量的卷积和非线性激活操作,因此会出现较大的语义差异,在跳跃连接过程中就有可能引入一些冗余无效的建筑物语义信息,从而影响网络对建筑物的识别和分割能力。为了优化网络的特征表达能力,减小不同深度特征图的语义差异,基于卷积注意力[20]设计了双重注意力模块,包括通道注意力和空间注意力机制。通过对特征图进行通道和空间上的自适应特征选择,提高网络对于目标的判别能力并抑制不相关的背景噪声。图 4 显示的是图 1 中 UAC 模块中的过程,深层特征通过上采样后与跳跃连接过来的浅层特征处于同一尺度,把它们叠加起来通过双重注意力模块,得到经过通道方面和空间方面调整后的特征,最后又经过 CBR×2 残差单元得到最终的优化特征。

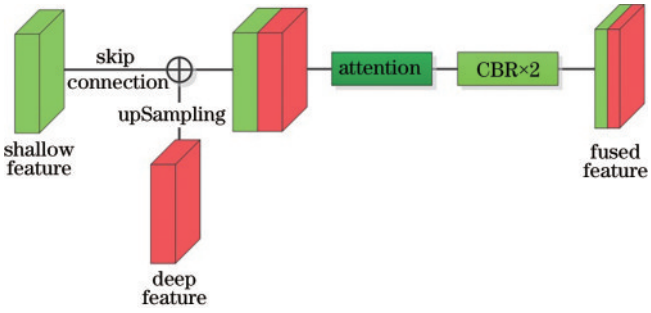

图 4　UAC 模块

Fig. 4　UAC module

图 5 展示的是图 1 中的 attention 模块的具体结构,双重注意力模块将叠加起来的浅层和深层特征 F 作为输入,并通过通道注意力和空间注意力模块对其进行处理。通道注意力模块专注于捕获通道之间的相互依赖性,而空间注意力模块专注于捕获空间位置之间的相互依赖性。这两个模块一起工作,有选择地强调信息的功能和抑制不相关的噪声,从而提高区分能力。此外,为了提高网络的拟合能力,引入了残差连接,允许模型学习原始特征和优化特征之间的差异。这有助于缓解由不同级别的特征图之间的语义间隙引入的信息丢失和噪声的问题,在经过注意力模块和残差连接

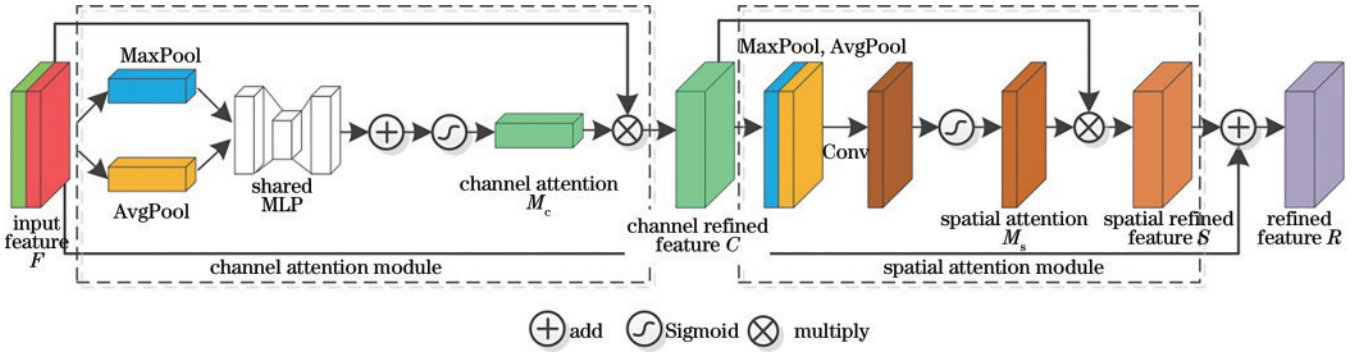

图 5　双重注意力模块

Fig. 5　Dual attention module

之后获得最终特征R。

2.5.1 通道注意力模块

通道注意力主要通过对特征图在通道维度上的加权来对不同通道的特征信息进行选择性整合，以提高特征图的表达能力和减少冗余信息，如图5所示，输入的是大小为$H\times W\times C$的组合特征F，分别在尺度纬度上进行全局平均池化和全局最大池化，得到两个大小为$1\times1\times C$的全局通道特征。然后，将两个池化结果分别送入多层感知器(MLP)中，通过两层权重共享的MLP得到两个通道注意力特征。接着将这两个特征相加，并利用Sigmoid函数进行归一化，得到通道注意力权重M_c。最后，将通道注意力权重M_c与输入特征图F相乘，以得到通道注意力结果特征图C。其计算方法如下：

$$M_c=\sigma\left\{\mathrm{MLP}\left[\mathrm{AvgPool}(F)\right]+\mathrm{MLP}\left[\mathrm{MaxPool}(F)\right]\right\},\tag{2}$$

$$C=F\otimes M_c,\tag{3}$$

式中：F是输入的特征图；σ是Sigmoid函数；AvgPool和MaxPool分别是全局平均池化和全局最大池化操作；MLP是多层感知器；M_c是通道注意力权重；C是通道注意力增强结果。

2.5.2 空间注意力模块

与通道注意力不同的是，空间注意力在空间纬度上通过加权调整特征图中每个像素点的权重，使不同空间位置的特征信息进行选择性整合，从而提高模型对空间位置信息的感知和利用能力，如图5所示，输入的是大小为$H\times W\times C$的通道注意力结果C，分别在通道纬度上进行平均池化和最大池化，得到两个大小为$H\times W\times1$的空间特征，并将两个池化结果叠加起来经过一个7×7的卷积，之后通过Sigmoid函数进行归一化处理，得到空间注意力权重M_s，最后，将空间注意力权重M_s与输入特征图C相乘，得到空间注意力结果特征图S。其计算方法如下：

$$M_s=\sigma\left\{\mathrm{Conv}_{7\times7}\left\{\mathrm{Concat}\left[\mathrm{AvgPool}(C),\mathrm{MaxPool}(C)\right]\right\}\right\},\tag{4}$$

$$S=C\otimes M_s,\tag{5}$$

式中：C是输入的特征图；Concat是叠加；$\mathrm{Conv}_{7\times7}$是7×7的卷积；M_s是空间注意力权重；S是空间注意力增强结果。

在双重注意力模块中，为了进一步提升模型的性能和拟合能力以及防止梯度消失、梯度爆炸现象的出现，将原始的输入特征图F与通过通道注意力和空间注意力加权后得到的优化特征图S相加，得到最终的输出特征图R。

$$R=S+F。\tag{6}$$

3 实验结果与分析

3.1 数据集

为了验证MM-Unet的适用性和有效性，采用3种不同分辨率的建筑物数据集，分别是Massachusetts Building数据集[21]、ISPRS Vaihingen数据集[22]和WHU Building数据集[23]。Massachusetts Building数据集覆盖马萨诸塞州大约340 km²地区，包含151张大小为1500×1500、分辨率为1 m的遥感图像，训练集有137张，验证集有4张，测试集有10张。ISPRS Vaihingen数据集覆盖Vaihingen市1.38 km²地区，包含33张大小约为2500×2500、分辨率为0.09 m的遥感图像，训练集和验证集共有17张，测试集有16张。WHU Building数据集覆盖新西兰大约450 km²地区，包含8188张大小为512×512、分辨率为0.3 m的遥感图像，训练集有4736张，验证集有1036张，测试集有2416张。在训练过程中，将所有图片裁剪为512×512大小，并进行数据集扩充操作，最终数量如表1所示，并由多个大小相同的RGB图片与其对应的灰度标签图片组合成像素矩阵作为输入特征提供给模型。

表1 实验数据集

Table 1 Experimental dataset

Dataset	Resolution /m	Train number	Validation number	Test number
Massachusetts Building	1.00	7416	288	90
WHU Building	0.30	4736	1036	2416
ISPRS Vaihingen	0.09	2752	736	306

3.2 实验环境设置

本实验基于Ubuntu 18.04.3操作系统，Intel(R) Xeon(R) Gold 6240 CPU @ 2.60 GHz，32 GB内存，通过一台NVIDIA Tesla V100显卡进行训练，显存为32 GB。Python版本为3.7.6，CUDA版本为10.1，使用PyTorch 1.8.1框架。模型训练的迭代次数设置为100，batch size设置为16。采用Adam优化器进行优化，初始学习率设置为0.001，并在训练过程中动态调整学习率。

3.3 精度评价指标

使用5个精度指标评估网络建筑物提取的精度：总体精度(R_{OA})、准确率($R_{\mathrm{precision}}$)、召回率(R_{recall})、F1分数和交并比(R_{IoU})。

$$R_{\mathrm{OA}}=\frac{N_{\mathrm{TP}}+N_{\mathrm{TN}}}{N_{\mathrm{TP}}+N_{\mathrm{TN}}+N_{\mathrm{FP}}+N_{\mathrm{FN}}},\tag{7}$$

$$R_{\mathrm{precision}}=\frac{N_{\mathrm{TP}}}{N_{\mathrm{TP}}+N_{\mathrm{FP}}},\tag{8}$$

$$R_{\mathrm{recall}}=\frac{N_{\mathrm{TP}}}{N_{\mathrm{TP}}+N_{\mathrm{FN}}},\tag{9}$$

$$s_{F1}=\frac{2\times R_{precision}\times R_{recall}}{R_{precision}+R_{recall}},\tag{10}$$

$$R_{IoU}=\frac{N_{TP}}{N_{TP}+N_{FP}+N_{FN}},\tag{11}$$

式中：N_{TP}表示正确分类的建筑物像元数；N_{TN}表示正确分类的背景像元数；N_{FP}表示错误分类为建筑物的像元数；N_{FN}表示错误分类为背景的像元数。R_{OA}表示正确分类的像元和所有像元的比例，$R_{precision}$表示正确分类的建筑物像元数占所有被分类为建筑物像元的比例，R_{recall}表示正确分类的建筑物像元数占所有实际建筑物像元数的比例，F1分数是准确率和召回率的调和平均值，R_{IoU}表示预测建筑物与实际建筑物的交集像元数除以它们的并集像元数。

3.4　实验结果分析

为了验证MM-Unet的性能表现，使用FCN、SegNet、Unet和Unet++等4个经典网络对建筑物提取的性能进行精度对比分析，其中，FCN的基本框架设置为ResNet50，所有网络模型均使用相同的运行环境以及优化参数。最后，为了验证改进模块的有效性和适用性，进行了消融实验。

3.4.1　Massachusetts Building数据集结果分析

图6展示的是Massachusetts Building数据集上多个网络提取建筑物的部分可视化对比结果。从图6可以看出，在中小型的建筑物提取上，MM-Unet不仅改善了错分和漏分的现象，并且边界的完整性要比其他网络高。FCN由于使用了ResNet50作为基本框架，下采样次数不够，局部信息不够精细导致小型建筑物提取效果最差，出现许多漏检现象。SegNet、Unet以及Unet++虽然漏分现象较少，但是边缘问题未能解决，相邻建筑物的边界未能区分，这是因为这3个网络中使用了大量的卷积和池化操作，丢失了部分空间信息，从图6第3行可以看出，g列的MM-Unet能够完整分割相邻建筑物，并未出现其他列的建筑物粘连与缺失现象，与e列的Unet相比，MM-Unet在保持了细节的同时还准确地提取了Unet所漏分的中小型建筑物。

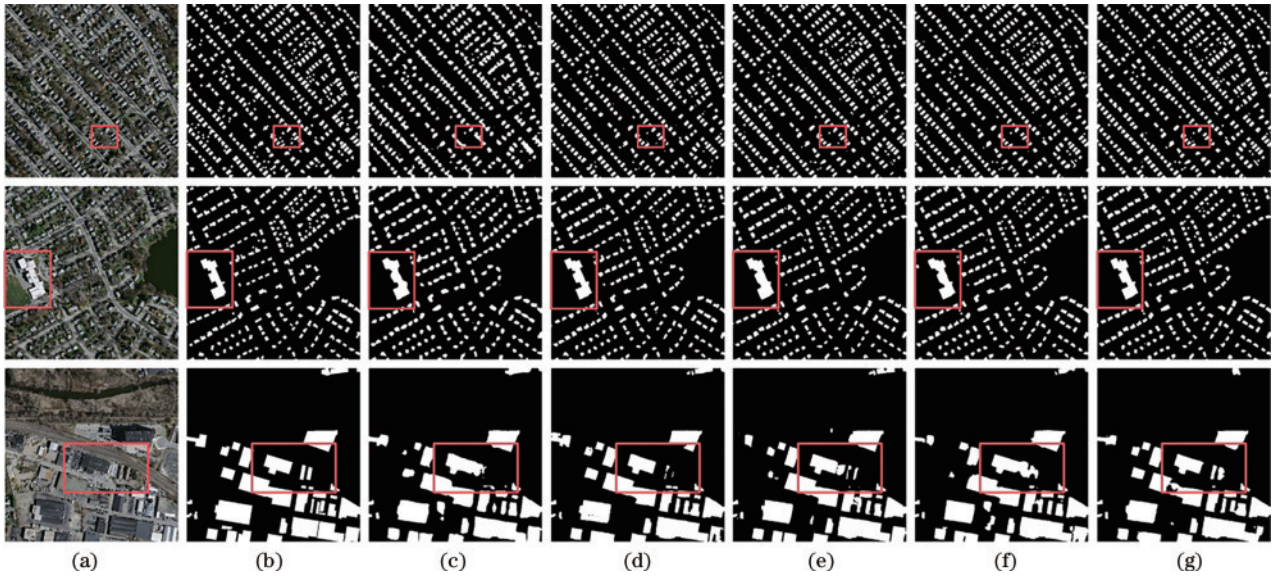

图6　Massachusetts Building数据集对比结果。(a)图像；(b)标签；(c)FCN；(d)SegNet；(e)Unet；(f)Unet++；(g)MM-Unet

Fig. 6　Comparison results of Massachusetts Building dataset. (a) Images; (b) labels; (c) FCN; (d) SegNet; (e) Unet; (f) Unet++; (g) MM-Unet

表2展示的是Massachusetts Building数据集上多个网络提取建筑物的精度对比。从表2可以看出，MM-Unet在总体精度、准确率、召回率、F1分数和交并比中均取得了最高精度，分别达到了94.46%、86.39%、83.12%、84.72%和73.42%。与基础网络Unet相比，分别提高了0.70个百分点、2.92个百分点、1.22个百分点、2.04个百分点和2.21个百分点。

表2　Massachusetts Building数据集精度对比

Table 2　Accuracy comparison of Massachusetts Building dataset

unit: %

Model	R_{OA}	$R_{precision}$	R_{recall}	s_{F1}	R_{IoU}
FCN	92.78	80.53	80.70	80.61	67.52
SegNet	93.00	81.83	81.68	81.75	69.87
Unet	93.76	83.47	81.90	82.68	71.21
Unet++	94.14	85.52	82.47	83.97	72.37
MM-Unet	**94.46**	**86.39**	**83.12**	**84.72**	**73.42**

3.4.2　WHU Building数据集结果分析

图7展示的是WHU Building数据集上多个网络提取建筑物的部分可视化对比结果。从图7可以看出，在大中型的建筑物提取上，MM-Unet不仅改善了空洞现象，并且边界的区分程度要比其他网络高。对于中型建筑物，4个经典网络提取的建筑物边界比较平滑，无法展示建筑物的边角。对于大型建筑物，FCN、SegNet以及Unet都出现了空洞现象，虽然Unet++没有出现

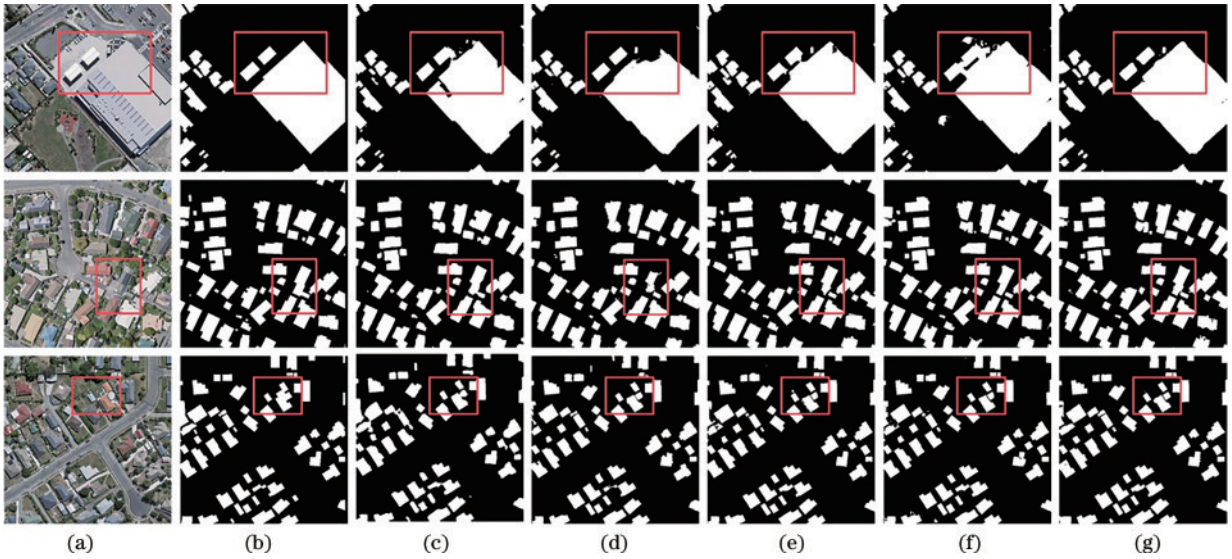

图 7　WHU Building 数据集对比结果。(a)图像;(b)标签;(c)FCN;(d)SegNet;(e)Unet;(f)Unet++;(g)MM-Unet

Fig. 7　Comparison results of WHU Building dataset. (a) Images; (b) labels; (c) FCN; (d) SegNet; (e) Unet; (f) Unet++; (g) MM-Unet

空洞现象,但是错分现象较为严重,这是由于Unet++特殊的跳跃连接能够获得较多的全局特征,弥补空间信息损失,但过多的跳跃连接可能引入一些冗余无效的信息,使错分现象发生。从图 7 第 1 行可以看出,g 列的 MM-Unet 能够完整提取大型建筑物,并未出现空洞和错分现象,与 e 列的 Unet 相比,MM-Unet 的建筑物边缘更加平滑与完整,更加接近实际的边界。

表 3 展示的是 WHU Building 数据集上多个网络提取建筑物的精度对比。从表 3 可以看出,MM-Unet 在总体精度、准确率、召回率、F1 分数和交并比中均取得了最高精度,分别达到了 98.85%、95.35%、94.25%、94.80% 和 90.11%,与基础网络 Unet 相比,分别提高了 0.17 个百分点、1.38 个百分点、0.02 个百分点、0.70 个百分点和 1.25 个百分点。

表 3　WHU Building 数据集精度对比

Table 3　Accuracy comparison of WHU Building dataset unit: %

Model	R_{OA}	$R_{precision}$	R_{recall}	s_{F1}	R_{IoU}
FCN	98.64	93.65	94.20	93.92	88.54
SegNet	98.62	93.90	94.00	93.95	88.59
Unet	98.68	93.97	94.23	94.10	88.86
Unet++	98.79	94.90	93.91	94.40	89.39
MM-Unet	**98.85**	**95.35**	**94.25**	**94.80**	**90.11**

3.4.3　ISPRS Vaihingen 数据集结果分析

图 8 展示的是 ISPRS Vaihingen 数据集上多个网络提取建筑物的部分可视化对比结果。从图 8 可以看出,在大型的建筑物提取上,MM-Unet 不仅有效防止了错分、漏分和空洞现象,并且建筑物的边界要比其他

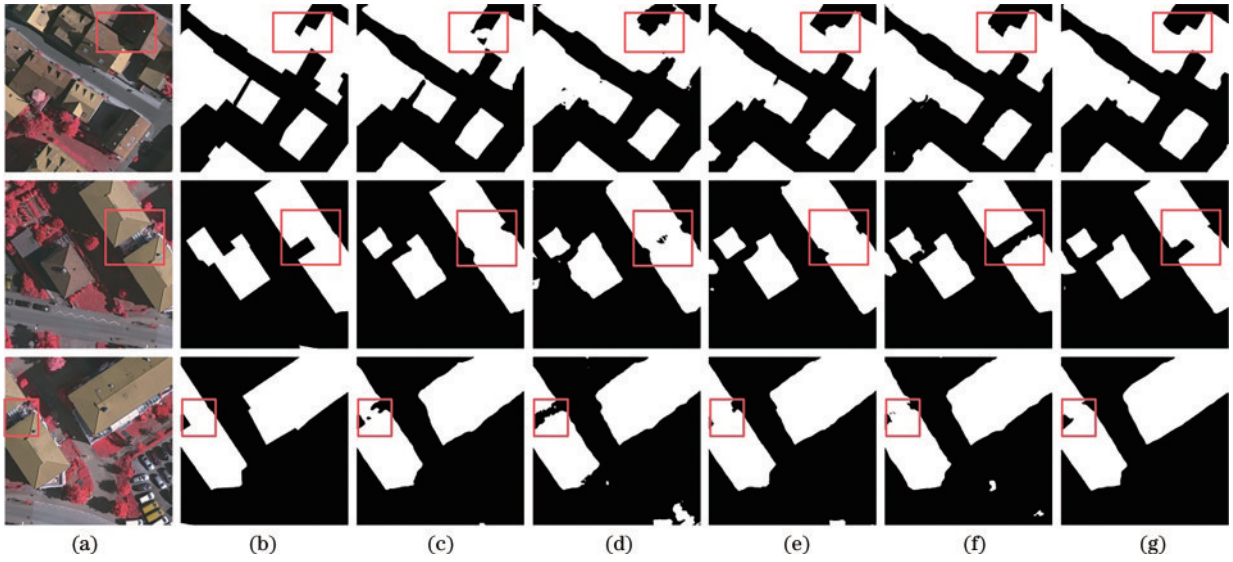

图 8　ISPRS Vaihingen 数据集对比结果。(a)图像;(b)标签;(c)FCN;(d)SegNet;(e)Unet;(f)Unet++;(g)MM-Unet

Fig. 8　Comparison results of ISPRS Vaihingen dataset. (a) Images; (b) labels; (c) FCN; (d) SegNet; (e) Unet; (f) Unet++; (g) MM-Unet

网络精确。FCN、SegNet、Unet以及Unet++均出现了错分和空洞现象，但是FCN的空洞现象不明显而且错分现象较少，这是由于FCN的基础框架使用ResNet50，下采样完成后尺寸为原始图像的1/8，而其他网络在下采样完成后尺寸为原始图像的1/16，所获取的全局特征不够，导致分割结果出现空洞现象。从图8第3行可以看出，g列的MM-Unet能够较为准确地区分颜色和纹理相近的物体与建筑物，与e列的Unet相比，MM-Unet不但没有把颜色相近的物体识别为建筑物，还能够完整保持不同颜色建筑物的边界。

表4展示的是ISPRS Vaihingen数据集上多个网络提取建筑物的精度对比。从表4中可以看出，MM-Unet在总体精度、准确率、召回率、F1分数和交并比中均取得了最高精度，分别达到了95.95%、92.95%、91.68%、92.31%和85.21%，与基础网络Unet相比，分别提高了0.38个百分点、1.08个百分点、1.42个百分点、1.26个百分点和1.55个百分点。

表4 ISPRS Vaihingen数据集精度对比

Table 4 Accuracy comparison of ISPRS Vaihingen dataset

unit: %

Model	R_{OA}	$R_{precision}$	R_{recall}	s_{F1}	R_{IoU}
FCN	95.80	92.56	90.33	91.43	84.57
SegNet	95.46	91.98	90.09	91.02	83.53
Unet	95.57	91.87	90.26	91.05	83.66
Unet++	95.63	91.18	91.64	91.40	84.17
MM-Unet	**95.95**	**92.95**	**91.68**	**92.31**	**85.21**

3.4.4 消融实验

为了验证所提多个模块的有效性和适用性，在Massachusetts Building数据集、ISPRS Vaihingen数据集和WHU Building数据集上，以Unet为基础模型，分别对Unet+DAM、Unet+DAM+MFCM、Unet+DAM+MFCM+MFEM（MM-Unet）进行消融实验和精度评估以研究各个模块的效果。在Massachusetts Building数据集上的结果如表5所示。

表5 Massachusetts Building数据集消融实验

Table 5 Ablation experiments of Massachusetts Building dataset

unit: %

Model	R_{OA}	$R_{precision}$	R_{recall}	s_{F1}	R_{IoU}
Unet	93.76	83.47	81.90	82.68	71.21
Unet+DAM	94.08	85.46	82.05	83.72	72.16
Unet+DAM+MFCM	94.23	86.15	82.97	84.53	73.15
MM-Unet	**94.46**	**86.39**	**83.12**	**84.72**	**73.42**

在Massachusetts Building数据集上与Unet相比，加入各个模块的模型的s_{F1}和R_{IoU}分别提高了1.04个百分点、1.85个百分点、2.04个百分点和0.95个百分点、1.94个百分点、2.21个百分点。各个模块在MM-Unet中对建筑物提取精度都具有提升效果。

在WHU Building数据集上的结果如表6所示，在WHU Building数据集上与Unet相比，加入各个模块的模型的s_{F1}和R_{IoU}分别提高了0.22个百分点、0.47个百分点、0.70个百分点和0.39个百分点、0.84个百分点、1.25个百分点。各个模块在MM-Unet中对建筑物提取精度都具有提升效果。

表6 WHU Building数据集消融实验

Table 6 Ablation experiments of WHU Building dataset unit: %

Model	R_{OA}	$R_{precision}$	R_{recall}	s_{F1}	R_{IoU}
Unet	98.68	93.97	94.23	94.10	88.86
Unet+DAM	98.73	94.55	94.09	94.32	89.25
Unet+DAM+MFCM	98.80	95.09	94.05	94.57	89.70
MM-Unet	**98.85**	**95.35**	**94.25**	**94.80**	**90.11**

在ISPRS Vaihingen数据集上的结果如表7所示，在ISPRS Vaihingen数据集上与Unet相比，加入各个模块的模型的s_{F1}和R_{IoU}分别提高了0.60个百分点、0.77个百分点、1.26个百分点和0.93个百分点、1.22个百分点、1.55个百分点。各个模块在MM-Unet中对建筑物提取精度都具有提升效果。

表7 ISPRS Vaihingen数据集消融实验

Table 7 Ablation experiments of ISPRS Vaihingen dataset

unit: %

Model	R_{OA}	$R_{precision}$	R_{recall}	s_{F1}	R_{IoU}
Unet	95.57	91.87	90.26	91.05	83.66
Unet+DAM	95.84	92.06	91.25	91.65	84.59
Unet+DAM+MFCM	95.90	92.74	90.93	91.82	84.88
MM-Unet	**95.95**	**92.95**	**91.68**	**92.31**	**85.21**

4 结 论

针对高分辨率的遥感影像在建筑物提取过程中常出现的边界模糊、空洞、误分和漏分现象，提出一种基于多模块的建筑物提取网络（MM-Unet）。使用多尺度特征组合模块以减少空间信息的丢失和加强多深度特征的利用；在跳跃连接完成后引入双重注意力模块，通过加强通道和空间上的自适应特征选择抑制不相关的背景噪声；在网络中加入多尺度特征增强模块，通过使用空洞卷积扩大感受野，加强网络全局特征和多尺度信息的提取。在Massachusetts、WHU以及Vaihingen建筑物数据集上的实验结果表明，MM-Unet对于不同分辨率的遥感影像建筑物提取任务表现良好，能够有效地解决中小型建筑物容易丢失和大型建筑物边界模糊问题，与改进网络Unet相比，各个提取精度指标均有提升。消融实验结果表明，多尺度特征组合模块、双重注意力模块和多尺度特征增强模块均

能够提高模型提取的精度，并具有较好的适用性。后续将进一步结合轻量化模块进行研究，以减少模型参数和训练时间。

参 考 文 献

[1] Shao P, Yi Y Q, Liu Z W, et al. Novel multiscale decision fusion approach to unsupervised change detection for high-resolution images[J]. IEEE Geoscience and Remote Sensing Letters, 2022, 19: 2503105.

[2] Shaloni, Dixit M, Agarwal S, et al. Building extraction from remote sensing images: a survey[C]//2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), December 18-19, 2020, Greater Noida, India. New York: IEEE Press, 2021: 966-971.

[3] 谭衢霖, 刘正军, 沈伟. 一种面向对象的遥感影像多尺度分割方法[J]. 北京交通大学学报, 2007, 31(4): 111-114, 119.
Tan Q L, Liu Z J, Shen W. An algorithm for object-oriented multi-scale remote sensing image segmentation [J]. Journal of Beijing Jiaotong University, 2007, 31(4): 111-114, 119.

[4] 陈行, 卓莉, 陶海燕. 基于MMBI的高分辨率影像建筑物提取研究[J]. 遥感技术与应用, 2016, 31(5): 930-938.
Chen H, Zhuo L, Tao H Y. Study on building extraction from high spatial resolution images using MMBI[J]. Remote Sensing Technology and Application, 2016, 31 (5): 930-938.

[5] Huang X, Zhang L P. Morphological building/shadow index for building extraction from high-resolution imagery over urban areas[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2012, 5(1): 161-172.

[6] Blockeel H, Struyf J. Efficient algorithms for decision tree cross-validation[J]. Journal of Machine Learning Research, 2003, 3(4/5): 621-650.

[7] Zhang H X, Li Q Z, Liu J G, et al. Image classification using RapidEye data: integration of spectral and textual features in a random forest classifier[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2017, 10(12): 5334-5349.

[8] Melgani F, Bruzzone L. Classification of hyperspectral remote sensing images with support vector machines[J]. IEEE Transactions on Geoscience and Remote Sensing, 2004, 42(8): 1778-1790.

[9] Shi Y L, Li Q Y, Zhu X X. Building segmentation through a gated graph convolutional neural network with deep structured feature embedding[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2020, 159: 184-197.

[10] 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 Press, 2015: 3431-3440.

[11] 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.

[12] 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), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 770-778.

[13] 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.

[14] 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.

[15] 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.

[16] Schlemper J, Oktay O, Schaap M, et al. Attention gated networks: learning to leverage salient regions in medical images[J]. Medical Image Analysis, 2019, 53: 197-207.

[17] Zhou Z W, Siddiquee M M R, Tajbakhsh N, et al. UNet: redesigning skip connections to exploit multiscale features in image segmentation[J]. IEEE Transactions on Medical Imaging, 2020, 39(6): 1856-1867.

[18] Ibtehaz N, Rahman M S. MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation[J]. Neural Networks, 2020, 121: 74-87.

[19] Mehta R, Sivaswamy J. M-net: a Convolutional Neural Network for deep brain structure segmentation[C]//2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), April 18-21, 2017, Melbourne, VIC, Australia. New York: IEEE Press, 2017: 437-440.

[20] 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.

[21] Mnih V. Machine learning for aerial image labeling[D]. Toronto: University of Toronto, 2013.

[22] Rottensteiner F, Sohn G, Gerke M, et al. ISPRS semantic labeling contest[R]. Leopoldshöhe: ISPRS, 2014.

[23] 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.