

# 基于改进 U-Net 的遥感影像建筑物提取方法

金澍<sup>1</sup>, 关沫<sup>1\*</sup>, 边玉婵<sup>2</sup>, 王舒磊<sup>1</sup>

<sup>1</sup>沈阳工业大学信息科学与工程学院, 辽宁 沈阳 110870;

<sup>2</sup>沈阳工业大学软件学院, 辽宁 沈阳 110870

**摘要** 提取遥感影像中的建筑物对智慧城市建设有着重要意义。针对传统方法提取背景复杂遥感影像时出现的精度低等问题,提出一种基于 U-Net 的遥感影像建筑物提取方法(MA-Unet)。该方法主要由编码器和解码器组成。在编码器中,引入卷积块注意力模块,其中通道注意力模块用来筛选更重要的特征,抑制无效特征,空间注意力模块用来筛选更深层次的语义特征,引入空洞空间金字塔池化模块提取不同尺度的特征。在解码器中,为了融合不同尺度大小的物体特征,将解码器中的特征图上采样后进行串联,这种信息聚合在某种程度上解决了不同尺度物体检测困难的问题。实验结果表明:MA-Unet 方法在 Massachusetts 建筑物数据集上的准确率、精度、交并比(IoU)分别优于 U-Net 网络 1.7 个百分点、2.1 个百分点、1.6 个百分点,在 WHU 建筑物数据集上的准确率、精度、IoU 分别优于 U-Net 网络 1.1 个百分点、1.4 个百分点、2.3 个百分点,是一种更为有效且具有一定实际应用价值的目标提取方法。

**关键词** 遥感影像; 语义分割; 建筑物提取; 注意力机制; 多尺度

中图分类号 TP751.1

文献标志码 A

DOI: 10.3788/LOP213004

## Building Extraction from Remote Sensing Images Based on Improved U-Net

Jin Shu<sup>1</sup>, Guan Mo<sup>1\*</sup>, Bian Yuchan<sup>2</sup>, Wang Shulei<sup>1</sup>

<sup>1</sup>School of Information Science and Engineering, Shenyang University of Technology, Shenyang 110870, Liaoning, China;

<sup>2</sup>School of Software, Shenyang University of Technology, Shenyang 110870, Liaoning, China

**Abstract** Building extraction from remote sensing images is of great significance to the construction of smart cities. Aiming to improve the low accuracy of traditional methods in extracting remote sensing images with a complex background, a remote sensing image building extraction method (MA-Unet) based on U-Net is proposed. This method mainly uses an encoder and a decoder. A convolutional block attention module is introduced into the encoder, in which a channel attention module is used to screen more important features and suppress invalid features, and a spatial attention module is used to screen deeper semantic features. An atrous spatial pyramid pooling module is introduced to extract features with different scales. In the decoder, to fuse object features with different scales, feature maps in the decoder are upsampled and connected in series. This information aggregation solves the difficulty of detecting objects with different scales to some extent. The experimental results show that MA-Unet method is superior to the U-Net method in terms of accuracy, precision, and intersection over union (IoU) by 1.7 percentage points, 2.1 percentage points, and 1.6 percentage points on the Massachusetts building dataset and by 1.1 percentage points, 1.4 percentage points, and 2.3 percentage points on the WHU building dataset, respectively. It is a more effective and practical target extraction method.

**Key words** remote sensing image; semantic segmentation; building extraction; attention mechanism; multi-scale

## 1 引言

近年来,随着深度学习技术的不断发展,对地观测

遥感平台的增多,高分辨率可见光遥感图像的数据量急剧增加,在遥感领域引起了许多关注<sup>[1]</sup>。高分辨率遥感图像具有的地理信息更为丰富,这也就为遥感图

收稿日期: 2021-11-19; 修回日期: 2021-12-15; 录用日期: 2021-12-21; 网络首发日期: 2022-01-04

通信作者: \*42533040@qq.com

像建筑物提取任务提供了有利基础。

深度学习的目标是训练出一个端对端的特征提取器。随着卷积神经网络(CNN)的提出,研究人员在图像分类、目标检测、语义分割等计算机视觉领域取得了重大进展。不断发展的硬件技术使 CNN 在遥感研究领域的影响力不断提高<sup>[2]</sup>。遥感图像语义分割任务对图像中的每一个像素点进行类别划分。目前,基于卷积神经网络的深度学习方法已经在计算机视觉领域取得了广泛的应用,该方法已应用于高分辨率可见光遥感图像的识别以及语义分割中。2015年,Long等<sup>[3]</sup>提出了 full convolution network(FCN),FCN 首次实现了端对端的训练方式,且获得了较为理想的结果。在该领域,具有代表性的网络结构有 U-Net<sup>[4]</sup>、SegNet<sup>[5]</sup>及 Deeplab 系列<sup>[6-9]</sup>等。Badrinarayanan 等<sup>[5]</sup>提出的 SegNet 是由编码器和解码器构成的。其中编码器部分用于特征提取;解码器会存储编码器进行池化后得到的像素索引以提高分割精度,在进行解码操作时释放所存储的索引信息,进行上采样操作后恢复原图尺寸。Chen 等<sup>[6-9]</sup>提出了 Deeplab 系列,其中 Deeplabv3 采用空洞空间金字塔池化(ASPP)模块对目标进行多尺度的特征提取,利用解码器融合不同尺度的特征进行精细分割。2018年,伍广明等<sup>[10]</sup>提出了一种双重约束的改进 U-Net,通过提高参数优化的程度提高了建筑物的分割精度。2019年,刘浩等<sup>[11]</sup>提出了一种特征压缩激活的 U-Net,通过增强网络对特征的识别能力以及对空间信息的恢复能力,进一步提高了对建筑物的分割精度。2022年,杨潇宇等<sup>[12]</sup>提出了一种增加注意力模块以及构建边缘监督网络的 U-Net,该网络通过加强建筑物特征和获取建筑物更精细的边缘特征提高了对建筑物的分割精度。

尽管有许多研究人员在高分辨遥感图像语义分割任务上已经做了许多研究,但是由于高分辨遥感图像存在背景复杂、建筑物的种类丰富以及不同尺度的物体干扰等原因,对遥感图像建筑物的提取精度降低。例如:存在和小型建筑物的尺度一致的物体时,网络可能会出现错分和漏分的情况;当建筑物尺度过小时,网络可能会出现无法完整分割等问题。鉴于此,本文在 U-Net 的基础上,提出了一种针对遥感影像特点的网络,通过引入卷积块注意力模块(CBAM)提升网络对建筑物特征的提取能力,通过融合不同尺度大小的特征图以及引入 ASPP 模块来提升网络对不同尺度建筑物的提取能力。最后,通过实验验证所提方法的性能,并选取其他典型网络进行对比分析。

## 2 研究理论与 MA-Unet 方法

### 2.1 U-Net 结构

U-Net 网络发表于 2015 年,用于对医学图像的语义分割。由于 U-Net 取得了不错的效果,也陆续被应用在语义分割的各种领域。U-Net 是典型的编码器-

解码器结构。在编码器结构中进行了 4 次最大池化操作,在每一次最大池化操作之前会经 2 次卷积操作。在解码器中对特征图进行了 4 次上采样操作,在每一次上采样前会经 2 次卷积操作,上采样后低级特征都会与对应尺寸的高级特征级联。尽管 U-Net 利用跳跃连接保留了低层次的语义信息,但是这些低层次的语义信息对提取尺度不一的建筑物是不足的,并且这些语义信息也存在无效信息。对于小型建筑物的提取,仍然需要更多的尺度信息。

### 2.2 MA-Unet 结构

针对 U-Net 提取建筑物时存在的问题,所提 MA-Unet 结构以 U-Net 为基础对网络进行改进,网络结构如图 1 所示。

MA-Unet 分为两个部分,分别为编码器与解码器。由于 U-Net 在跳跃连接时存在无效的语义信息,故在编码器中每一层的特征图在经过卷积层之后加入 CBAM。在 CBAM 的通道注意力模块中,通过通道注意力机制来加强建筑物特征,抑制其他无效特征,从而提高网络精度;在 CBAM 的空间注意力模块中,空间注意力机制可从已进行上采样的特征图中恢复建筑物的空间信息。针对 U-Net 对建筑物尺度信息提取不足的问题,在编码器最后一层通道数为 1024 的特征图后加入改进的 ASPP 模块。通过 ASPP 模块可以实现对不同尺度特征的提取,在低层阶段获得精细的目标像素分割定位,在高层阶段提取更为有效的建筑物特征,从而提高分割精度。在解码器中,为了解决遥感图像存在的不同尺寸的物体的多尺度问题,对每一个尺度的特征图进行串联,在最后进行提取时,特征图包含了各个尺度的信息特征。通过跨尺度信息聚合的方式可以解决“同物异谱,异物同谱”的问题。

### 2.3 压缩激活模块

Squeeze-and-Excitation 网络(SE-Net)<sup>[13]</sup>是自动驾驶公司 Momenta 在 2017 年发表的图像识别模型。该网络的核心在于注意力模块,根据特定的数学关系可以让网络更加注重信息丰富的建筑物特征,抑制其他背景特征。SE 模块共分为两部分,分别是压缩(squeeze)和激活(excitation)部分。在压缩部分,原始特征图  $U$  的维度大小为  $H \times W \times C$ ,其中  $H$  为高度, $W$  为宽度, $C$  为通道数。 $(i, j)$  表示图片中的像素值。输出的计算公式为

$$Z_c = F_{sq}(U) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W U(i, j), \quad (1)$$

式中: $F_{sq}$  为通道注意力模块。输出  $Z_c$  表示一个通道上的全部像素的平均值,维度为  $1 \times 1 \times C$ 。对特征图  $U$  的某一通道进行全局平均池化操作,将二维的特征变为一个实数,该实数一定程度上具备该通道的全局感受野。

第二步操作激活函数  $F_{ex}(\cdot, W)$ ,通过各通道之间

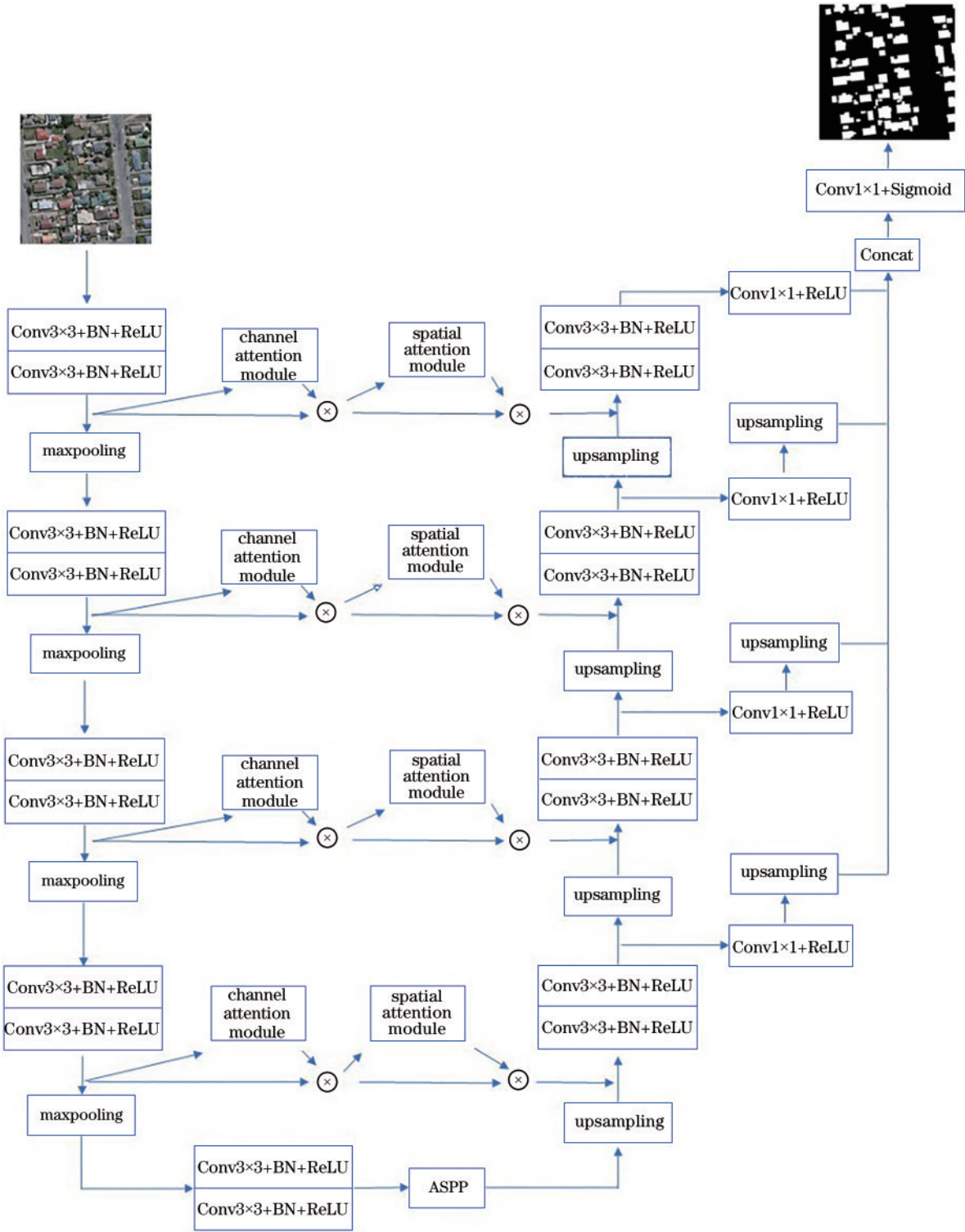


图 1 MA-Unet 结构  
Fig. 1 MA-Unet structure

的相关性来赋予重要通道更多的权重,赋予各通道权重后通过训练进行学习。采用两层全连接层来完成这一操作,在第一个全连接层后接 ReLU,在最后一层全连接层后接 Sigmoid 预测每个通道的重要性,计算方法为

$$S = F_{ex}(z, W) = \sigma[g(z, W)] = \sigma[W_2 \delta(W_1 z)], \quad (2)$$

式中:  $\delta$  指 ReLU 激活函数;  $\sigma$  表示 Sigmoid 激活函数;

$W_1 \in \mathbf{R}^{c \times \frac{c}{r}}$  为第一层全连接层的参数,  $r$  为通道的压缩比率,通过压缩来将特征维度降低到原来的  $1/r$ ;  $W_2 \in \mathbf{R}^{c \times \frac{c}{r}}$  为第二层全连接层的参数,第二个全连接层将特征图的维度恢复为  $C$ ,最后经 Sigmoid 激活函数输出权重值。通过两层的全连接层可以增强模型的非线性拟合能力,更加准确地对通道间的相关性进行建模。

最后将通过激活部分得到的权重赋值给相对应的通道,表达式为

$$\tilde{X}_c = F_{\text{scale}}(u_c, s_c) = \mathbf{S}\mathbf{C} \cdot \mathbf{U}\mathbf{C}, \quad (3)$$

式中:  $\tilde{X} = [\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_c]$ ;  $F_{\text{scale}}(u_c, s_c)$  表示给相对应的通道赋予权重。

### 2.4 空洞空间金字塔池化模块

在遥感图像中,建筑物在遥感图像中的目标尺寸大小不一,在进行语义分割的过程中小型建筑物易被忽略,这种空间信息的损失通常是在网络模型中持续的池化操作导致的。在进行池化时,固定的池化核会让提取到的特征具有一定的局限性。通过提取不同尺度的特征,可以提升对小型建筑物的提取效果。针对以上问题,引入 ASPP 模块<sup>[14]</sup>来进行建筑物提取。

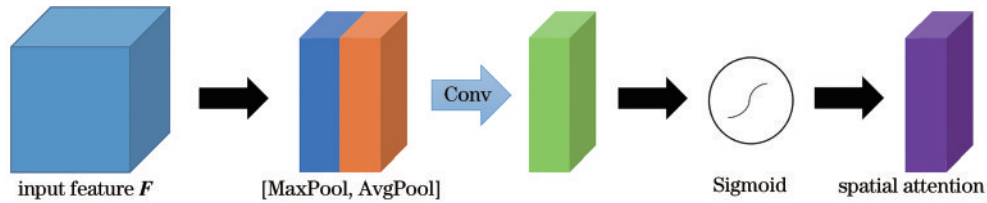


图 2 空间注意力模块

Fig. 2 Spatial attention module

特征图  $F$  的大小为  $H \times W \times C$ , 对特征图  $F$  在通道维度上进行平均池化和最大池化,得到大小为  $H \times W \times 1$  的特征图  $F_{\text{avg}}^S \in \mathbf{R}^{H \times W \times 1}$  和  $F_{\text{max}}^S \in \mathbf{R}^{H \times W \times 1}$ ; 对两个特征图进行融合,融合后对该特征图进行卷积,卷积核大小为  $7 \times 7$ ,经过卷积后的特征图大小为  $H \times W \times C$ ;之后利用 Sigmoid 函数进行激活获得权重,将该权重与原特征图  $F$  逐像素相乘,得到最终的结果  $F'$ 。这一过程的相关表达式为

$$M_s(F) = \sigma \left\{ f^{7 \times 7} \left[ \text{AvgPool}(F); \text{MaxPool}(F) \right] \right\}, \quad (4)$$

$$M_s(F) = \sigma \left\{ f^{7 \times 7} \left[ F_{\text{avg}}^S; F_{\text{max}}^S \right] \right\}, \quad (5)$$

式中:  $M_s(F)$  表示特征图  $F$  输入到空间注意力模块后处理得到的结果;  $f^{7 \times 7}$  为对特征图进行卷积核为  $7 \times 7$

其中也在该模块中使用空洞卷积,它可以在不增加网络参数量以及不降低特征维度的情况下增大感受野,从而提升分割效果。针对遥感图像中存在的不同尺度物体的分割精度问题,本文将改进的 ASPP 模块加入到网络中。ASPP 模块内经过改进的空洞卷积的空洞率为 3、5、7,特征图经不同空洞率的卷积层处理后再进行上采样操作,之后堆叠为特征金字塔,然后对输出结果进行融合。改进的 ASPP 模块通过不同空洞率的并行滤波器提取到了不同尺度的特征,解决了遥感图像分割中建筑物大小各不相同的问题。

### 2.5 空间注意力模块

在所提 MA-Unet 网络中增加的空间注意力机制<sup>[15]</sup>通过提取建筑物像素的位置信息来筛选更为有效的建筑物特征。空间注意力模块的结构如图 2 所示。

的卷积操作。

### 2.6 卷积块注意力模块

卷积块注意力模块结合了通道注意力机制与空间注意力机制,相较于只关注通道域的 SE 模块, CBAM 在通道域和空间域两个维度推算注意力图,可筛选出更重要的语义特征,因此可以取得更好的效果,具体结构如图 3 所示。

首先将特征图输入到通道注意力模块中得到对应通道的权重,得到加权结果后将加权后的特征图输入到空间注意力模块中得到对应权重,加权后输出特征图,计算方法表示为

$$A_{cs}(U) = M_s[F_{sq}(U)], \quad (6)$$

式中:  $A_{cs}$  为 CBAM 操作。

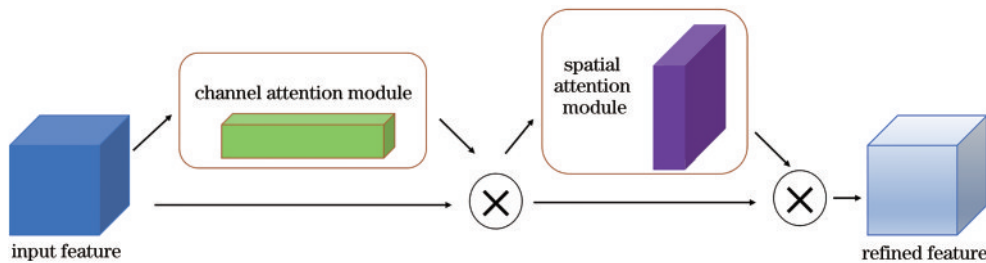


图 3 CBAM 结构

Fig. 3 CBAM structure

## 3 实验结果与分析

### 3.1 实验环境

实验基于 Windows10 操作系统,使用 Pytorch 框

架搭建深度学习模型, GPU 为 NVIDIA GeForce RTX3070, CPU 为 Intel i5-10600KF。

### 3.2 数据集介绍及数据增广

为了防止数据集的独特性对网络的影响,故

采用两个数据集进行对比实验。所用的数据集为 Massachusetts 遥感建筑物数据集<sup>[16]</sup>和 WHU 数据集<sup>[17]</sup>。其中 Massachusetts 建筑物数据集由 151 张分辨率为  $1500 \times 1500$  的遥感图像组成,覆盖马萨诸塞州约  $340 \text{ km}^2$ ,其中 137 张遥感图像作为训练集,4 张遥感图像作为验证集,10 张遥感图像作为测试集。WHU 数据集包括分辨率为  $512 \times 512$  的 4736 张训练集、1036 张验证集以及 2416 张测试集,其标注精度高于 Massachusetts 建筑物数据集。训练网络时需要足够的数据支撑,故使用数据增广来生成新的样本实例。当用于网络训练的数据较少时,数据增广可以增加模型的鲁棒性。对遥感影像进行数据增广<sup>[18]</sup>的方法有很多,考虑到马萨诸塞州数据量较少,对 Massachusetts 数据集中的训练集进行  $512 \text{ pixel} \times 512 \text{ pixel}$  的随机切割,对切割后的图像进行水平与垂直翻转、旋转、增加一定程度的噪声,最终数据集图片量扩充为原本的 6 倍。

### 3.3 评价指标

为了对比所提方法与其他方法的提取效果,采用 3 种指标来评估不同方法的效果,包括准确度 (accuracy,  $A$ )、精确度 (precision,  $P$ )、交并比 (IoU),公式分别为

$$A = \frac{N_{TP} + N_{TN}}{N_{TP} + N_{TN} + N_{FP} + N_{FN}}, \quad (7)$$

$$P = \frac{N_{TP}}{N_{TP} + N_{FP}}, \quad (8)$$

$$R_{IoU} = \frac{N_{TP}}{N_{TP} + N_{FP} + N_{FN}}, \quad (9)$$

式中:准确度为正确分类像素的百分比;精确度为正确分类建筑物像素占建筑物总像素的百分比;交并比评价建筑物与背景分类的效果,其中还考虑到将建筑物像素错分类为背景像素的情况;TP (true positive) 为真正,即正确将建筑物像素分类的像素总数;TN (true negative) 为真负,即正确将背景像素分类的像素总数;FP (false positive) 为假正,即错误将背景像素分类为建筑物像素的像素总数;FN (false negative) 为假负,即错误将建筑物像素分类为背景像素的像素总数。

### 3.4 网络训练优化器

网络训练优化器采用 AdamOptimizer 算法<sup>[19]</sup>,学习率为 0.0001。Epoch 设置 100 个,训练完每个 epoch 后对测试集进行评价,若超过 15 个 epoch 后评价指标训练不再提高,则提前终止模型训练<sup>[20]</sup>。

### 3.5 对比实验分析

#### 3.5.1 Massachusetts 建筑物数据集提取结果

使用在马萨诸塞州训练好的模型对测试集进行预测,输入图像的分辨率为  $512 \times 512$ ,输出结果为建筑物提取图。对所提方法 MA-Unet 与相关文献方法进行对比,其中 SegNet 和 PSPnet 均为典型的语义分割网络。实验中所对比的网络模型具有相同的运行环境和同样的网络训练优化参数。部分提取图如图 4 所示。

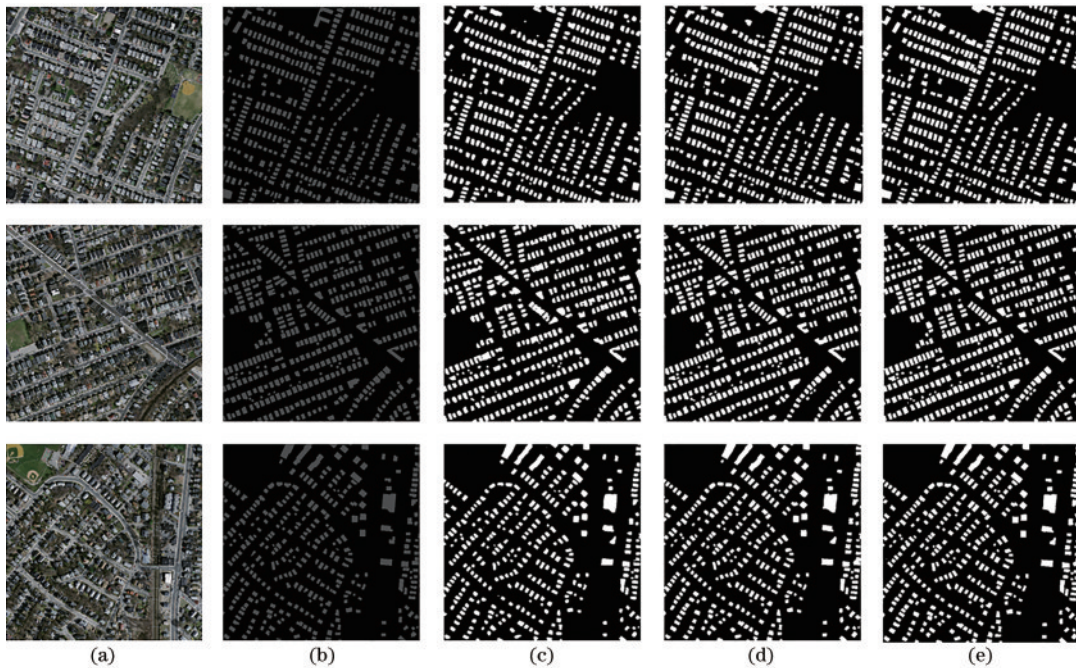


图 4 Massachusetts 数据集上的对比效果。(a)原始图像;(b)标签图;(c) SegNet; (d) U-Net; (e) MA-Unet

Fig. 4 Contrast effects on the Massachusetts dataset. (a) Original images; (b) label images; (c) SegNet; (d) U-Net; (e) MA-Unet

所提 MA-Unet 模型在基线模型 U-Net 的基础上,在编码器中通过加入 CBAM 让网络从通道域和空间域提升识别特征的能力,在解码器中通过融合不同尺

度的特征让网络具备多尺度信息。通过对比提取图可以发现,MA-Unet 模型在 Massachusetts 建筑物测试集上的效果优于其他对比模型,在某些车辆密集的情况

下,对小型建筑物预测得不够准确和将密集车辆错误预测为建筑的情况得到改善。MA-Unet 可以提取出更高质量的建筑物分割图,具体的模型的泛化能力实验结果如表 1 所示。可以看出:所提网络的评价指标均优于 U-Net、SegNet、PSPnet;所提网络的准确率、精确度、交并比较 U-Net 提高 1.7 个百分点、2.1 个百分点、1.6 个百分点。

### 3.5.2 WHU 建筑物数据集提取结果

使用在 WHU 建筑物数据集训练好的模型对测试集进行预测,输入图像的分辨率为  $512 \times 512$ ,输出结

表 1 Massachusetts 数据集上的对比结果

Table 1 Comparison results on Massachusetts dataset

Model	Accuracy/%	Precision/%	IoU/%
U-Net	89.80	65.28	67.61
PSPnet	89.40	63.91	66.36
SegNet	87.76	61.84	65.24
MA-Unet	91.50	67.35	69.24

果为建筑物提取图。实验中所对比的网络模型具备相同的运行环境和同样的网络训练优化参数,部分建筑物提取图如图 5 所示。

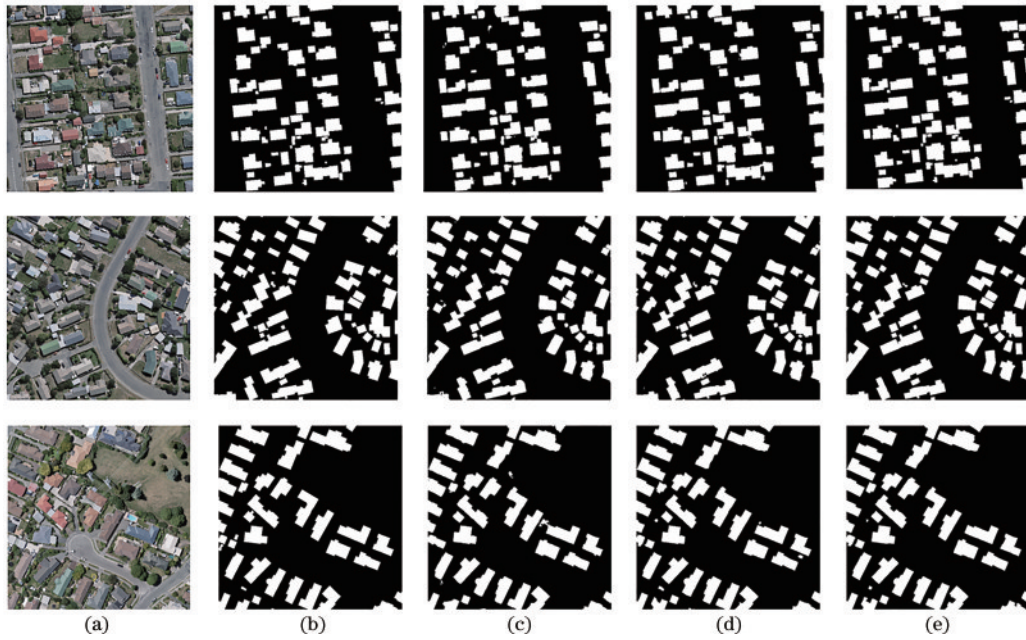


图 5 WHU 数据集上的对比效果。(a)原始图像;(b)标签图;(c) SegNet; (d) U-Net; (e) MA-Unet

Fig. 5 Contrast effects on the WHU dataset. (a) Original images; (b) label images; (c) SegNet; (d) U-Net; (e) MA-Unet

所提模型在 WHU 测试集上的预测效果优于其他模型,并且最终提取的建筑物的完整程度也更高,对部分小型建筑物漏分和错分的情况有所改善,在小型建筑物与大型建筑物相连的部分也精细提取到特征,部分地表物体颜色形状与建筑物相似时提取的结果也较为准确,建筑物的边缘也较为平滑。根据各类评价指标,对模型的泛化能力实验结果如表 2 所示。可以看出:所提网络的评价指标均优于 U-Net、SegNet、PSPnet;所提网络的准确率、精确度、交并比较 U-Net 提高 1.1 个百分点、1.4 个百分点、2.3 个百分点。

表 2 WHU 数据集上的对比结果

Table 2 Comparison results on WHU dataset

Model	Accuracy/%	Precision/%	IoU/%
U-Net	96.86	88.97	86.34
PSPnet	96.63	88.32	86.56
SegNet	97.12	86.27	85.74
MA-Unet	97.92	90.41	88.65

## 4 结 论

对遥感影像建筑物目标提取进行研究,针对 U-Net 模型的不足,提出了一种多尺度信息聚合的深度学习网络 MA-Unet。由于 U-Net 在跳跃连接时存在无效的语义信息,通过在网络中引入卷积块注意力模块增强网络识别重要特征的能力。其中通过通道注意力机制来加强建筑物特征,抑制其他无效的背景特征;通过空间注意力机制从已进行上采样的特征图中恢复建筑物的空间信息。由于遥感影像存在不同尺度大小的目标特征,故对网络中各尺度特征图进行串联来提高网络提取精度;在网络中加入空洞空间金字塔池化模块来丰富网络的尺度信息。实验结果表明,MA-Unet 在 Massachusetts 建筑物数据集和 WHU 建筑物数据集上所预测的结果与真实标签更为接近。通过对比实验可以看出:MA-Unet 的评价指标均得到提高,一定程度上改善了基线网络 U-Net 的错分和漏分情况,建筑物提取图也较为完整。融合卷积块注意力

模块筛选后的建筑物特征图,配合空洞空间金字塔池化模块提取的不同尺度的特征,提升了模型在复杂背景下对小型建筑物的提取能力。但是仍需进一步改进 MA-Unet,提升其在更为复杂的背景信息中提取建筑物的能力,下一步可结合更为完善的其他深度学习方法进行研究。

### 参 考 文 献

- [1] 何小飞, 邹峥嵘, 陶超, 等. 联合显著性和多层卷积神经网络的高分影像场景分类[J]. 测绘学报, 2016, 45(9): 1073-1080.  
He X F, Zou Z R, Tao C, et al. Combined saliency with multi-convolutional neural network for high resolution remote sensing scene classification[J]. Acta Geodaetica et Cartographica Sinica, 2016, 45(9): 1073-1080.
- [2] 田莹, 王亮, 丁琪. 基于深度学习的图像语义分割方法综述[J]. 软件学报, 2019, 30(2): 440-468.  
Tian X, Wang L, Ding Q. Review of image semantic segmentation based on deep learning[J]. Journal of Software, 2019, 30(2): 440-468.
- [3] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition, June 7-12, 2015, Boston, MA, USA. New York: IEEE Press, 2015: 3431-3440.
- [4] 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.
- [5] 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.
- [6] Chen L C, Papandreou G, Kokkinos I, et al. Semantic image segmentation with deep convolutional nets and fully connected CRFs[EB/OL]. (2014-12-22) [2021-05-08]. <https://arxiv.org/abs/1412.7062>.
- [7] 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.
- [8] 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.
- [9] Chen L C, Papandreou G, Schroff F, et al. Rethinking atrous convolution for semantic image segmentation[EB/OL]. (2017-06-17) [2021-05-04]. <https://arxiv.org/abs/1706.05587>.
- [10] 伍广明, 陈奇, Shibasaki R, 等. 基于 U 型卷积神经网络的航空影像建筑物检测[J]. 测绘学报, 2018, 47(6): 864-872.  
Wu G M, Chen Q, Shibasaki R, et al. High precision building detection from aerial imagery using a U-net like convolutional architecture[J]. Acta Geodaetica et Cartographica Sinica, 2018, 47(6): 864-872.
- [11] 刘浩, 骆剑承, 黄波, 等. 基于特征压缩激活 Unet 网络的建筑物提取[J]. 地球信息科学学报, 2019, 21(11): 1779-1789.  
Liu H, Luo J C, Huang B, et al. Building extraction based on SE-Unet[J]. Journal of Geo-Information Science, 2019, 21(11): 1779-1789.
- [12] 杨潇宇, 汪西莉. 结合多尺度注意力和边缘监督的遥感图像建筑物分割模型[J]. 激光与光电子学进展, 2022, 59(22): 2228004.  
Yang X Y, Wang X L. Building segmentation model of remote sensing image combining multi-scale attention and edge supervision[J]. Laser & Optoelectronics Progress, 2022, 59(22): 2228004.
- [13] 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.
- [14] He K M, Zhang X Y, Ren S Q, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904-1916.
- [15] 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.
- [16] Mnih V. Machine learning for aerial image labeling[D]. Toronto: University of Toronto, 2013.
- [17] 季顺平, 魏世清. 遥感影像建筑物提取的卷积神经网络与开源数据集方法[J]. 测绘学报, 2019, 48(4): 448-459.  
Ji S P, Wei S Q. Building extraction via convolutional neural networks from an open remote sensing building dataset[J]. Acta Geodaetica et Cartographica Sinica, 2019, 48(4): 448-459.
- [18] Lashgari E, Liang D H, Maoz U. Data augmentation for deep-learning-based electroencephalography[J]. Journal of Neuroscience Methods, 2020, 346: 108885.
- [19] Kingma D P, Ba J. Adam: a method for stochastic optimization[EB/OL]. (2014-12-22) [2021-04-05]. <https://arxiv.org/abs/1412.6980>.
- [20] 向阳, 赵银娣, 董霁红. 基于改进 UNet 孪生网络的遥感影像矿区变化检测[J]. 煤炭学报, 2019, 44(12): 3773-3780.  
Xiang Y, Zhao Y D, Dong J H. Remote sensing image mining area change detection based on improved UNet siamese network[J]. Journal of China Coal Society, 2019, 44(12): 3773-3780.