武汉大学学报(信息科学版)
Geomatics and Information Science of Wuhan University
ISSN 1671-8860,CN 42-1676/TN

《武汉大学学报(信息科学版)》网络首发论文

题目：基于多特征融合与对象边界联合约束网络的建筑物提取
作者：高贤君，冉树浩，张广斌，杨元维
DOI：10.13203/j.whugis20210520
收稿日期：2022-06-20
网络首发日期：2022-07-18
引用格式：高贤君，冉树浩，张广斌，杨元维．基于多特征融合与对象边界联合约束网络的建筑物提取[J/OL]．武汉大学学报(信息科学版).
https://doi.org/10.13203/j.whugis20210520

DOI:10.13203/j.whugis20210520
引用格式：
高贤君，冉树浩，张广斌，等. 基于多特征融合与对象边界联合约束网络的建筑物提取[J]. 武汉大学学报 •信息科学版，2022，DOI：10.13203/j.whugis20210520（GAO Xianjun, RAN Shuhao, ZHANG Guangbin, et al. Building Extraction Based on Multi-feature Fusion and Object-boundary Joint Constraint Network[J]. Geomatics and Information Science of Wuhan University, 2022, DOI: 10.13203/j.whugis20210520）

基于多特征融合与对象边界联合约束网络的建筑物提取

高贤君[1] 冉树浩[1] 张广斌[1] 杨元维[1,2,3]

1 长江大学 地球科学学院，湖北 武汉，430100
2 城市空间信息工程北京市重点实验室，北京，100045
3 湖南科技大学测绘遥感信息工程湖南省重点实验室，湖南 湘潭，411201

摘要：针对现有全卷积神经网络因光谱混杂，造成建筑物漏检和误检，以及边界缺失的问题，设计了一种基于多特征融合与对象边界联合约束网络的高分辨率遥感影像建筑物提取方法。该方法基于编解码结构，并在编码阶段末端融入连续空洞空间金字塔模块，以在不损失过多有效信息的前提下进行多尺度特征提取和融合；在解码阶段，通过实现基于对象和边界的多输出融合约束结构，为网络融入更多准确的建筑物特征并细化边界；在编码与解码阶段间的横向跳级连接中引入卷积块注意力机制，以增强有效特征。此外，解码阶段的多层级输出结果还被用于构建分段多尺度加权损失函数，实现对网络参数的精细化更新。在 WHU 和 Massachusetts 建筑物数据集上进行对比试验分析，其中 IoU 和 F_1 分数分别达到了 90.44%、94.98%和 72.57%、84.10%，且模型的复杂度与效率均优于 MFCNN 与 BRRNet。

关键词：建筑物提取；全卷积神经网络；多尺度特征；注意力机制；联合约束
中图分类号：TP751.1　　**文献标志码：**A

1. 引言

从高分辨率遥感影像中进行建筑物的精确自动提取研究在城市规划、地图数据更新、应急响应等方面都具有极为重要的意义[1]。根据特征提取方式的不同，现有从高分辨率遥感中进行建筑物提取方法主要包括两大类。一类是传统人工设计提取特征，并结合图像处理与分析方法来提取建筑物，包括基于对象分割方法[2]、基于建筑物特征的方法[3]、以及基于辅助

收稿日期：2022-06-20
第一作者：高贤君，博士，副教授，主要从事高分辨率图像目标自动识别的理论和方法研究。电子邮件：junxgao@yangtzeu.edu.cn
通讯作者：冉树浩，硕士。电子邮件：201500880@yangtzeu.edu.cn
基金资助：城市空间信息工程北京重点实验室开放基金（20210205）；城市轨道交通数字化建设与评价技术国家工程实验室开放基金（2021ZH02）；湖南科技大学测绘与遥感地理信息工程湖南省重点实验室开放基金（E22133）；海南省地球观测重点实验室开放研究基金（2020LDE001）；国家自然科学基金（41872129）。

信息的方法[4]等。然而，传统方法的性能受到特征表示能力的极大限制，易受季节、光照、传感器质量、建筑物风格和环境等的影响。此外，特征设计与选取过程过度依赖先验知识和可变参数调节，难以做到建筑物特征的全面、多层次化描述，无法真正做到自动化和通用化。另一类是通过神经网络自动提取影像高低维度特征并进行像素级分类的提取方法，相较于传统方法，其特征提取能力更强，自动化程度更高，适用范围更广，极大的推动了基于深度学习技术的建筑物提取任务的发展。在神经网络发展的早期，主要以“滑动窗口+分类网络”的策略进行建筑物提取，但其提取结果通常连续性较差且计算效率偏低。后期，随着全卷积神经网络（fully convolutional networks, FCN）[5]的提出，端到端的像素级图像分割成为现实，从遥感影像进行建筑物提取的研究重点也逐步从卷积神经网络（convolutional neural networks, CNN）转向 FCN。

为实现对原始输入影像高层级特征的提取与学习，FCN 通常会在编码阶段采用池化操作，以缩小特征图并减少计算量。尽管 FCN 能够在解码阶段实现对特征图大小的还原，但由于池化操作而丢失的原始信息却难以恢复，导致细节信息丢失，建筑物提取精度降低。针对上述问题，现阶段对 FCN 的改进工作主要集中在如下两个方面：一是改善特征提取能力，二是优化解码过程。

在改善模型特征提取能力方面，主要是通过提高模型的多尺度特征提取融合能力来提高网络性能。Szegedy 等[6]提出了 Inception 结构，通过并行不同卷积核大小的卷积运算，实现对不同尺度物体的检测。但由于多分支卷积为普通卷积，这导致了大量的冗余计算，降低了提取效率。Zhao 等[7]在金字塔场景解析网络（pyramid scene parsing network, PSPNet）提出了一种金字塔池化模块（pyramid pooling module, PPM）,通过采用多尺度池化方式以实现多尺度特征的提取和融合。Chen 等基于空洞卷积的思想，在 DeepLabv2[8]和 DeepLabv3[9]中提出并改进了空洞空间金字塔池化(atrous spatial pyramid pooling, ASPP)模块，通过并行多个不同空洞率的空洞卷积来获取不同尺度的特征信息。然而，上述的 PPM 和 ASPP 模块常采用较大的池化窗口和空洞率以获取更大范围的图像信息，即更大的感受野，这通常会造成提取过程中大量有效信息的损失，削弱长距离信息之间的关联性，并降低模型对具有可变光谱特征建筑物提取的完整度。此外，特征图中每个通道都可以看作是网络模型对特定语义信息的响应[10],而现有的多尺度特征提取模块未能实现对特征图通道间关联性的建立，影响了模型对特征图通道注意力的分配。

优化解码过程的策略主要是为解码阶段提供更丰富、更准确的浅层特征信息，使其能够恢复部分原始信息，以此提高分割精度。SegNet[11]利用池化索引结构记录 MaxPooling 操作

中最大值的位置信息，并在解码阶段进行恢复，提高了边界的划分精度。Ronneberger 等人通过融合编码阶段的浅层位置信息和解码阶段的高级语义信息，提出了一种具有 U 形结构的 U-Net 模型[12]。此后，诸如 ResUNet-a[13]、BRRNet[14]、PRCUnet[15]等多种建筑物提取模型均基于 U-Net 拓展而来。然而，上述模型只在网络末端进行预测输出，且训练过程中参数更新动量会逐步衰减，这就导致了远离输出端网络参数优化程度不足，进而影响了网络对浅层特征的学习能力。而后，MA-FCN[16]、MFCNN[17]、U 型 CNN[18]等模型通过对解码阶段单层级预测输出进行改进，构建了多重约束网络模型，并在建筑物提取任务中取得了良好效果。然而，多重约束结构却未能实现对解码阶段其他层级中有效特征信息的进一步利用，导致模型难以对具有“异物同谱”现象建筑物进行准确提取。此外，编码阶段的浅层特征虽能提供一定的边缘特征信息，但随着网络层级的加深，这种特征信息会逐渐减弱或产生误差，因此需要进一步强化解码阶段边缘特征信息的表达，以实现对建筑物边缘的准确恢复。

针对现有建筑物提取全卷积神经网络在多尺度特征提取融合过程中存在大量有效信息损失，且在解码阶段对于建筑物有效特征利用不足以及对边界细节信息感知能力较弱的问题，本文提出了一种多特征融合与对象边界联合约束网络，提升了光谱混杂区域建筑物提取的正确率以及完整度，并保持良好的边界。

2. 基本原理

2.1. 整体框架

为更好的缓解因光谱混淆造成利用神经网络从高分辨率遥感影像所提建筑物存在较多漏检、误检以及边界粘连缺失的问题，本文设计了一种基于多特征融合与对象边界联合约束网络的建筑物提取方法，其技术流程图如图 1 所示。首先利用滑动窗口对经过预处理后的高分辨率遥感影像进行分割处理，然后按一定比例将数据划分为训练集、验证集、测试集三部分。将训练集和验证集数据输入本文所提网络进行训练，并采用随机梯度下降法实现权重参数的学习更新。最后再利用训练阶段所得到的网络模型对测试集影像进行预测，得到建筑物提取结果。

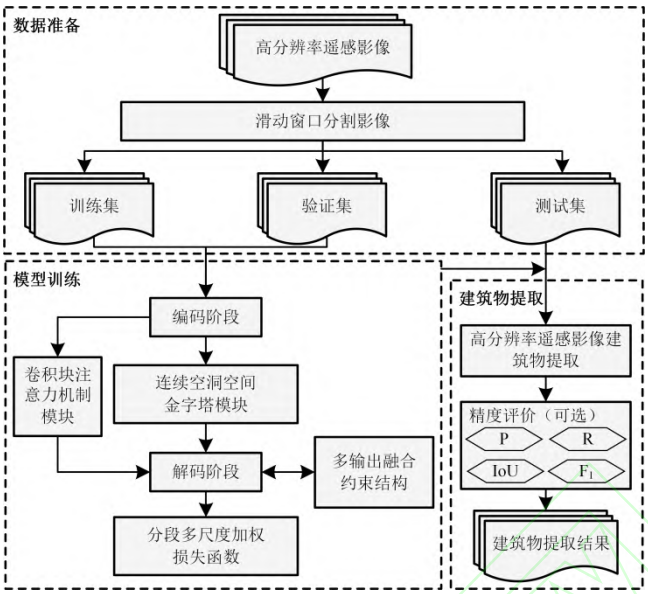

图1 建筑物提取技术流程图

Fig.1 Technical flowchart of building extraction

多特征融合与对象边界联合约束网络结构如图 2 所示，其以改进后的 U-Net 作为主体架构，主要包含编码阶段、多尺度特征提取融合阶段以及解码阶段。编码阶段与 U-Net 保持一致，但为减少计算量，将最后层级的特征图通道数降低为 512。连续空洞空间金字塔模块（continuous-atrous spatial pyramid module, CSPM）紧跟编码阶段，通过连续空洞卷积和通道注意力机制的组合完成多尺度特征提取融合。解码阶段与编码阶段相对应，每一层级通过双线性插值的方式与下一层级进行关联，且在每一层级末端进行建筑物对象和边界的预测输出，并将多层级输出结果反向融入网络，实现基于对象和边界多输出融合约束结（multi-output fusion constraint structure, MFCS）。此外，在编码阶段与解码阶段的横向跳级连接中引入了卷积块注意力机制模块（convolutional block attention module, CBAM），增强网络对重要的特征的学习能力。最后，利用多层级输出结果与标准参照结果共同构建分段多尺度加权损失函数，进一步优化网络训练过程。

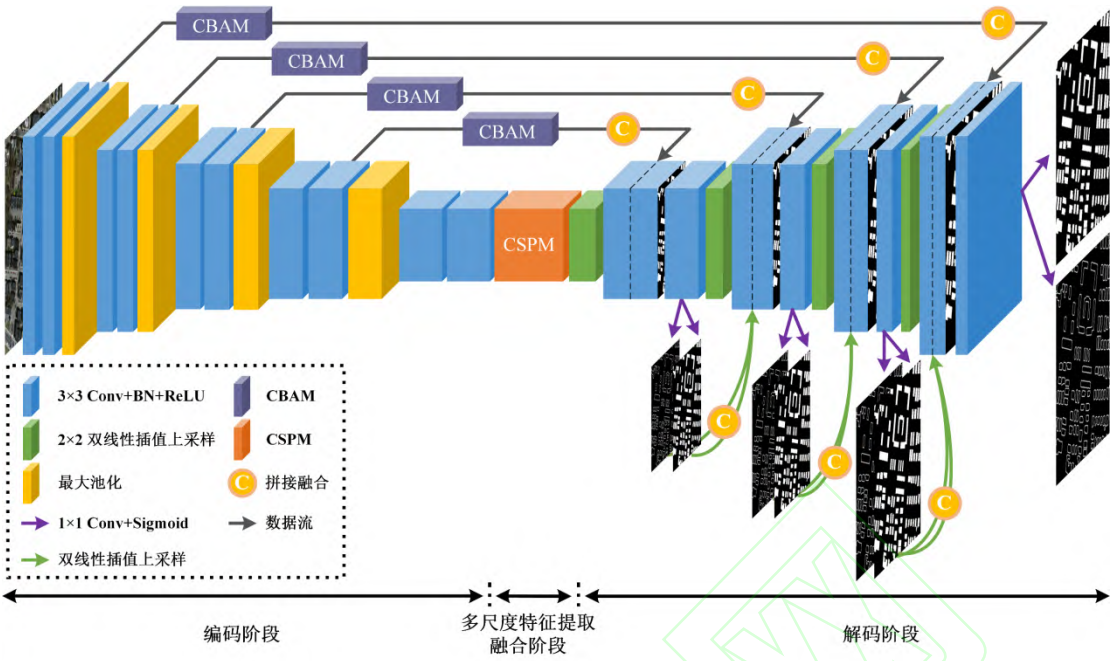

图2 多特征融合与对象边界联合约束网络

Fig.2 Multi-feature fusion and object boundary joint constraint network

2.2. 基于 CBAM 的横向跳级连接

在具有编解码结构的全卷积神经网络中，输入影像在编码阶段的多次池化下采样操作过程中，特征图所蕴含的空间位置、边缘轮廓等细节信息被逐步损失，且解码阶段的上采样操作也难以实现相关信息的完全恢复。因此，通过横向跳级连接将编码与解码阶段的特征进行结合，构建特征金字塔，有助于提高感兴趣目标检测精度。但在建筑物提取任务中，由于遥感影像中包含众多干扰地物噪声信息，单一的特征图串联会导致网络对编码阶段特征的注意力分配产生混淆，未能充分关注并提取有效特征和细节信息。为实现对编码阶段特征中有效特征信息的增强以及无效特征信息的抑制，本文在横向跳级连接中引入 CBAM 模块[19]，其基本结构如图 3 所示。CBAM 由通道注意力模块和空间注意力模块两部分串联而成，分别聚焦于全局信息（‘What’）和局部信息（‘Where’）的权重分配。设输入特征为$\boldsymbol{F} \in \mathbb{R}^{C\times H\times W}$，输出为$\boldsymbol{F}'$，$\otimes$代表哈达玛积，则 CBAM 可表示为：

$$\boldsymbol{F}' = (\boldsymbol{F} \otimes \boldsymbol{M}_C) \otimes \boldsymbol{M}_S \tag{1}$$

式中，$\boldsymbol{M}_C \in \mathbb{R}^{C\times 1\times 1}$代表通道注意力图，其计算过程如式(2)；$\boldsymbol{M}_S \in \mathbb{R}^{1\times H\times W}$代表空间注意力图，其计算过程如式(3)：

$$\boldsymbol{M}_C = \sigma\left(MLP\left(AvgPool(\boldsymbol{F})\right) + MLP\left(MaxPool(\boldsymbol{F})\right)\right) \tag{2}$$

$$\boldsymbol{M}_S = \sigma\left(f^{7\times 7}([AvgPool(\boldsymbol{F}); MaxPool(\boldsymbol{F})])\right) \tag{3}$$

式中，$AvgPool$和$MaxPool$分别代表全局平局池化和全局最大池化，MLP为共享权重卷积，σ代表 Sigmoid 函数激活。

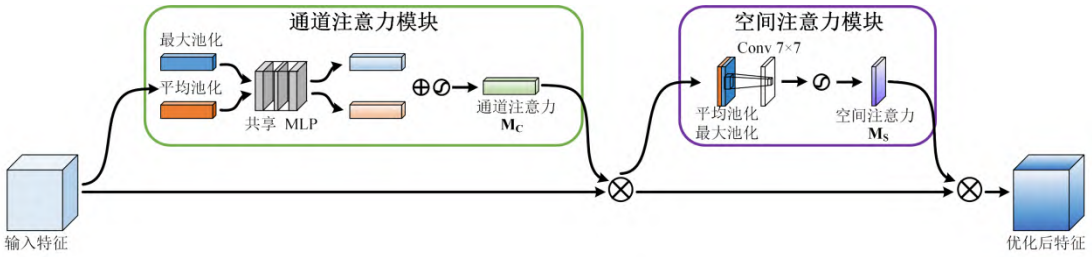

图3 卷积块注意力机制模块

Fig. 3 Convolutional block attention module (CBAM)

2.3. 连续空洞空间金字塔模块

在城市场景的高分辨率遥感影像中，建筑物往往密集分布，尺寸大小差异明显，为实现模型对不同尺度建筑物的自适应提取，则需要进行多种尺度的特征提取及融合。基于混合空洞卷积（hybrid dilated convolution, HDC）思想[20]，本文设计了一种多尺度特征提取融合模块 CSPM，其结构如图 4 所示。CSPM 由并行连续空洞卷积模块、残差模块、通道注意力模块组成。连续空洞卷积模块主要思想是通过并行符合 HDC 约束的连续小尺度空洞卷积运算，实现对输入特征图中大、中、小三种不同尺度特征进行提取，使每一条支路末端层级感受野完全覆盖原始输入特征，有效减缓多尺度特征提取过程中有效信息的损失，并增强远距离信息间的连续性。设 CSPM 的输入特征为$\boldsymbol{F} \in \mathbb{R}^{C\times H\times W}$，并行连续空洞卷积模块输出特征为$\boldsymbol{F_P} \in \mathbb{R}^{C\times H\times W}$，则并行连续空洞卷积模块计算过程可表示为：

$$\boldsymbol{F_P} = \left((\boldsymbol{F} * \Phi_1^1) * \left((\Phi_3^1\delta * \Phi_3^2\delta * \Phi_3^3\delta) + (\Phi_3^1\delta * \Phi_3^3\delta * \Phi_3^5\delta) + (\Phi_3^1\delta * \Phi_3^3\delta * \Phi_3^9\delta)\right)\right) * (\boldsymbol{F} * \Phi_1^1) \tag{4}$$

式中，$\Phi_k^d \in \mathbb{R}^{\frac{C}{2}\times C\times H\times W}$表示空洞卷积，$k$代表卷积核大小，$d$代表空洞率，$\delta$代表批归一化（BN）处理和 Sigmoid 函数激活，‘$*$’代表卷积运算。

残差模块为恒等映射，旨在增强对原始输入信息的重用，并利于误差反向传播，避免梯度消失。此外，由于卷积核的大小有限，卷积运算窗口通常只能提取聚合局部感受野区域中的空间和通道信息，难以从全局的角度去考虑各个通道的关联性。因此，本文在 CSPM 中引入 DANet[21]中的通道注意力机制，通过对特征图通道间的依赖关系进行建模，实现对具有特定语义信息通道的强化，并对部分无用特征通道进行抑制。设通道注意力模块的输出图为$\boldsymbol{F_C} \in \mathbb{R}^{C\times H\times W}$，CSPM 的最终输出特征$\boldsymbol{F_{Out}} \in \mathbb{R}^{C\times H\times W}$,则 CSPM 的计算过程可以表示为：

$$\boldsymbol{F_{Out}} = \varphi_{0.3}(\boldsymbol{F} + \boldsymbol{F_P} + \boldsymbol{F_C}) \tag{5}$$

式中，φ_r代表 DropOut 操作，r表示神经元消亡率。

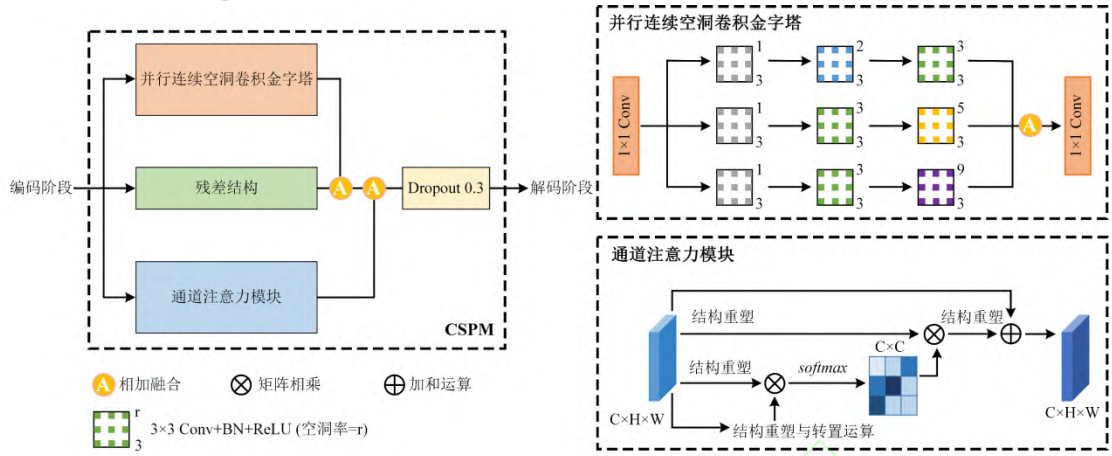

图4 连续空洞空间金字塔模块

Fig. 4 Continuous-atrous spatial pyramid module (CSPM)

2.4. 基于对象和边界的多输出融合约束结构

在现有语义分割模型中，为实现对输入数据逐像素的精确分类，通常需要构建从粗推理到精推理的编解码结构，并在解码阶段进行预测输出。受特征金字塔网络（feature pyramid networks, FPN）[22]的启发,本文改进了网络的解码阶段，提出了基于建筑物对象和边界的多输出融合约束结构，其结构如图 5 所示。首先，将原始 U-Net 网络解码阶段的卷积层划分为 4 个层级，自顶向下，标记为$\{L_1, L_2, L_3, L_4\}$，并使用 1×1 大小的卷积层级 Sigmoid 激活函数对其进行预测输出，分别得到建筑物对象预测图$\{O_1, O_2, O_3, O_4\}$和建筑物边界预测图$\{B_1, B_2, B_3, B_4\}$。然后，将$\{O_2, O_3, O_4\}$和$\{B_2, B_3, B_4\}$利用 Concatenate 的方式与下一层级起始特征$\boldsymbol{F}_i \in \mathbb{R}^{C\times H\times W}(i \in [1,3])$进行融合，得到新的起始特征$\boldsymbol{F}_{i_new} \in \mathbb{R}^{(C+2)\times H\times W}(i \in [1,3])$，以增强网络对来自前一层级有效特征的聚合能力，融入更多准确的建筑物特征，并实现对建筑物边界的细化。最后，利用双线性插值的方式对$\{O_1, O_2, O_3, O_4\}$和$\{B_1, B_2, B_3, B_4\}$分别上采样 1、2、4、8 倍至原图大小，并分别与建筑物对象和建筑物边界标准参照结果构建损失函数$\{l_{ia}^1, l_{ia}^2, l_{ia}^3, l_{ia}^4\}$和$\{l_{ba}^1, l_{ba}^2, l_{ba}^3, l_{ba}^4\}$，对网络进行多输出联合约束，实现对网络参数精确、有效的更新。

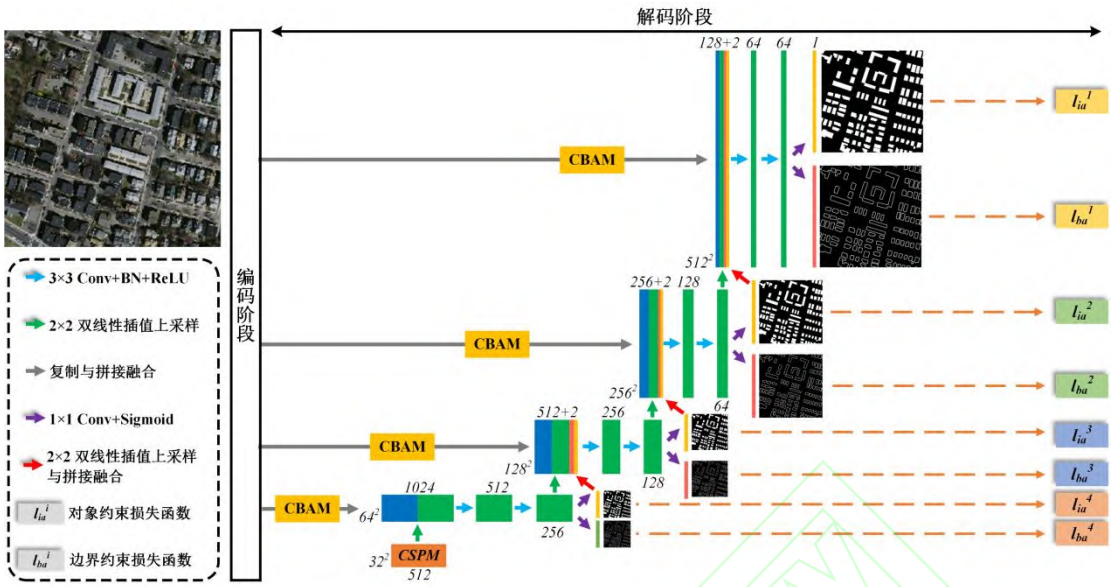

图5 多输出融合约束结构

Fig. 5 Multi-output fusion constraint structure (MFCS)

2.5. 分段多尺度加权损失函数

在 CNN 的训练过程中，损失函数被用来计算预测值与标准参照值之间的差异，并以这种差异作为导向进行网络参数的更新与优化。如图 5 所示，在网络的解码阶段包含多个建筑物对象和边界的预测输出,对于每一层级的建筑物对象损失约束，本文采用了在医疗影像分割领域中常用的 Dice Loss[23]作为损失函数，其能更好的平衡样本中建筑物像素与背景像素的关系，避免陷入局部最优，其计算过程如式(6)所示。此外，考虑到建筑物边界像素在影像中所占比例较低，本文采用了类平衡交叉熵损失函数[24]，其计算过程如式(7)所示：

$$l_{\mathrm{ia}} = 1 - \frac{2\times\sum_{i=1}^{N}(g_i\times p_i)}{\sum_{i=1}^{N} g_i+\sum_{i=1}^{N} p_i} \tag{6}$$

$$l_{\mathrm{ba}} = -\frac{1}{N}\sum_{i=1}^{N}(\beta g_i \times log\, p_i + (1-\beta)(1-g_i)\times log(1-p_i)) \tag{7}$$

式中，N表示影像总的像素数目，g_i表示标准参照结果中第i个像素是否属于建筑物，若属于建筑物则$g_i = 1$，否则$g_i = 0$，P_i表示的是预测图中第i个像素为建筑物的概率，β表示非边界像素数目在建筑物标准参照结果中所占的比例。

在网络解码阶段，建筑物边界预测结果中目标像素极度稀疏，且自底向上每一层级特征图分辨率逐渐升高，所蕴含语义信息愈加丰富与准确。为平衡各层级输出约束，本文提出一种分段多尺度加权损失函数，其计算过程如式(8)所示。在网络训练初期，预测结果与标准参照结果差异明显，此时采用基于建筑物对象的多层级损失约束，进行粗放式参数更新，实现网络向目标特征的快速拟合。随着训练的进行，网络参数的更新逐步趋向于精细化，此时

向网络中添加边界损失，并逐渐降低或舍弃低分辨率层级损失约束，以达到精细化更新网络参数的目的。

$$L_{total}=\begin{cases}\sum_{n=1}^{4}\omega_n l_{\mathrm{ia}}^{n}, & (0<e<m)\\ \sum_{n=1}^{3}\omega_n(\lambda_n {l_{\mathrm{ia}}}^{n}+\mu_n {l_{\mathrm{ba}}}^{n}), & (m\le e<k)\\ \sum_{n=1}^{2}\omega_n(\lambda_n {l_{\mathrm{ia}}}^{n}+\mu_n {l_{\mathrm{ba}}}^{n}), & (k\le e\le t)\end{cases} \tag{8}$$

式中，ω_n为第n个层级损失约束权重值，且$\omega_1+\quad+\omega_n=1$，$\lambda_n$和$\mu_n$分别代表第$n$个层级建筑物对象和边界损失约束权重值，且$\lambda_n+\mu_n=1$，$e$为当前训练迭代次数，$m$和$k$为自定义条件下的迭代次数，$t$为网络训练总迭代次数。

3. 实验过程及结果分析

3.1. 数据集

WHU 建筑物数据集由季顺平等[25]在 2019 年开源公布，覆盖了新西兰克赖斯特彻奇约 450km^2 的区域。该航空数据集由包含 4736 幅影像的训练集、1036 幅影像的验证集、2416 幅影像的测试集共同组成，且所有影像大小均为 512×512 像素，空间分辨率为 0.3m。图 6（a）展示了部分影像与对应建筑物标签数据。

Massachusetts 建筑物数据集由 Mnih[26]在 2013 年开源公布，一共包含了 155 幅波士顿地区的航拍影像及建筑物标签图。影像的空间分辨率为 1m，大小为 1500×1500 像素。受限于 GPU 显存大小，本文采用 512×512 像素大小的滑动窗口对原始影像进行裁剪，并剔除部分不完整影像，最后得到的训练集包含 1066 幅影像，验证集包含 36 幅影像，测试集包含 90 幅影像。图 6(b)展示了部分裁剪后的影像与对应建筑物标签数据。

图6 航空数据集影像及对应建筑物标签数据。（a）WHU建筑物数据集；（b）Massachusetts建筑物数据集

Fig. 6 Aerial datasets image and corresponding building label images. (a) WHU Building Dataset; (b) Massachusetts Buildings Dataset

3.2. 实验细则和实验条件

本文所有的实验均在装有 64 位 Windows10 专业工作站版操作系统的工作站进行。工作

站配备了 AMD Ryzen5 5600X 6-Core Processor 处理器，32GB 的内存，1TB 固态硬盘以及 NVIDIA GeForce RTX 309024GB 显存的显卡用以加快模型训练。所有网络模型均在基于 Python3.8+Pytorch1.8.0+CUDA11.1 的深度学习环境下进行实现。在模型训练过程中，以 Adam 作为模型优化器进行梯度下降与参数更新，学习率设置为 0.0001，迭代次数设置为 50，每次迭代的批量大小为 8。此外，在实验中还采用了自动混合精度(automatic mixed precision, AMP)和梯度缩放模式策略，以减少对显存的占用率。

3.3. 评价指标

为准确评估所提模型性能，本文选取了准确率(P)、召回率(R)、F_1 分数和交并比(intersection over union, IoU)对实验结果进行评价。其中，P 指的是被正确分类为正类的像素在所有被分为正类的像素中所占的比例，其计算过程如式(9)；R 指的是被正确分类为正类的像素在所有正类的像素中的比例，其计算过程如式(10)；F_1 分数则是 P 与 R 的调和平均结果，其计算过程如式(11)；IoU 是所有预测为正类的像素与真实正类像素的交集比上他们的并集，其计算过程如式(12)：

$$P = \frac{TP}{TP+FP} \times 100\% \tag{9}$$

$$R = \frac{TP}{TP+FN} \times 100\% \tag{10}$$

$$F_1 = \frac{2\times P\times R}{P+R} \times 100\% \tag{11}$$

$$IoU = \frac{TP}{TP+FP+FN} \times 100\% \tag{12}$$

式中，TP 表示被正确分类的建筑物像素个数；FP 表示被错误分类的背景像素个数；FN 表示被错误分类的建筑物像素个数；TN 表示被正确分类的背景像素个数。

3.4. 实验结果与分析

3.4.1. 对比模型

为进一步评估本文模型的有效性，本文选取了 U-Net、U 型 CNN、MA-FCN、MFCNN、BRRNet 作为对比模型。U-Net 拥有简洁明了的编解码结构，且作为本文模型的基础架构，其首先被选为基准模型。U 型 CNN、MA-FCN 和 MFCNN 与本文模型类似，都具有多层级输出约束结构，其中 U 型 CNN 为双重约束，MA-FCN、MFCNN 与本文模型为多重约束，且 MFCNN 具备多尺度特征融合模块，将他们作为对比模型，能有效的验证本文模型的优

异性。此外，考虑到残差结构和空洞卷积在神经网络中的广泛应用，本文还选取了具有残差细化结构的 BRRNet 作为对比模型。

3.4.2. WHU 建筑物数据集实验结果及分析

各对比模型在 WHU 建筑物数据集的定量评价结果如表 1 所示。由表可知，本文模型在除准确率外的其余指标上均优于对比模型，其中 IoU 和 F_1 分数比基准 U-Net 模型分别高出了 1.88%和 1.05%。将本文模型与其余四种建筑物提取模型相比，本文模型在准确率和召回率上相较于 F_1 分数第二高的 MA-FCN 分别高出了 0.37%和 0.4%，这表明本文模型对建筑物提取的正确率和完整度均有提升。

表 1 WHU 建筑物数据集上各模型定量评价结果

Table 1 Quantitative evaluation result of several modules on the WHU Building Dataset

方法	P /%	R /%	IoU/%	F_1/%
U-Net[12]	93.69	94.18	88.56	93.93
MA-FCN[16]	94.26	94.93	89.75	94.60
MFCNN[17]	94.70	93.94	89.25	94.32
BRRNet[14]	**94.95**	94.13	89.64	94.54
U 型 CNN[18]	94.22	94.14	89.00	94.18
本文方法	94.63	**95.33**	**90.44**	**94.98**

* 最佳定量评价值以加粗和下划线的形式突出显示

图 7 展示了各对比模型在 WHU 建筑物数据集上的部分可视化提取结果，其中影像 1、2 为大型建筑物，影像 3、4 为小型密集建筑物，而影像 5、6、7 为不同大小的混合型建筑物集群。从图中可以看出本文模型所提建筑物对象更为连续完整，过度提取现象较少，边界也更为精确清晰。如影像 1、3、5、6、7 所示，当建筑物屋顶与地面光谱特征相似时，对比模型会对建筑物与道路、裸地的判别产生混淆，造成建筑物的漏检与误检。在影像 2 中，由于大量汽车存在，影响了同一建筑物光谱特征连续表达，造成了对比方法所提建筑物存在大量空洞和缺失。针对空洞漏检的情况，本文模型通过构建 CSPM 减缓了空洞卷积过程中有效信息的损失，使所提取建筑物的颜色、纹理等信息具备更强的连续性，有效的避免了大型建筑物内部空洞以及不连续现象的产生。而对于过度提取现象，本文模型将 MFCS 的多层级预测输出结果以及经过注意力分配后的浅层特征融入解码阶段，补充并强化了有效特征，提高了建筑物检测精度。此外，对于如影像 4 中密集连续分布的建筑物以及影像 5 中被部分被树木所遮挡的建筑物，相较于对比模型，本文模型通过对建筑物边界进行约束，获得了更好的区分度以及更为完整的提取结果。

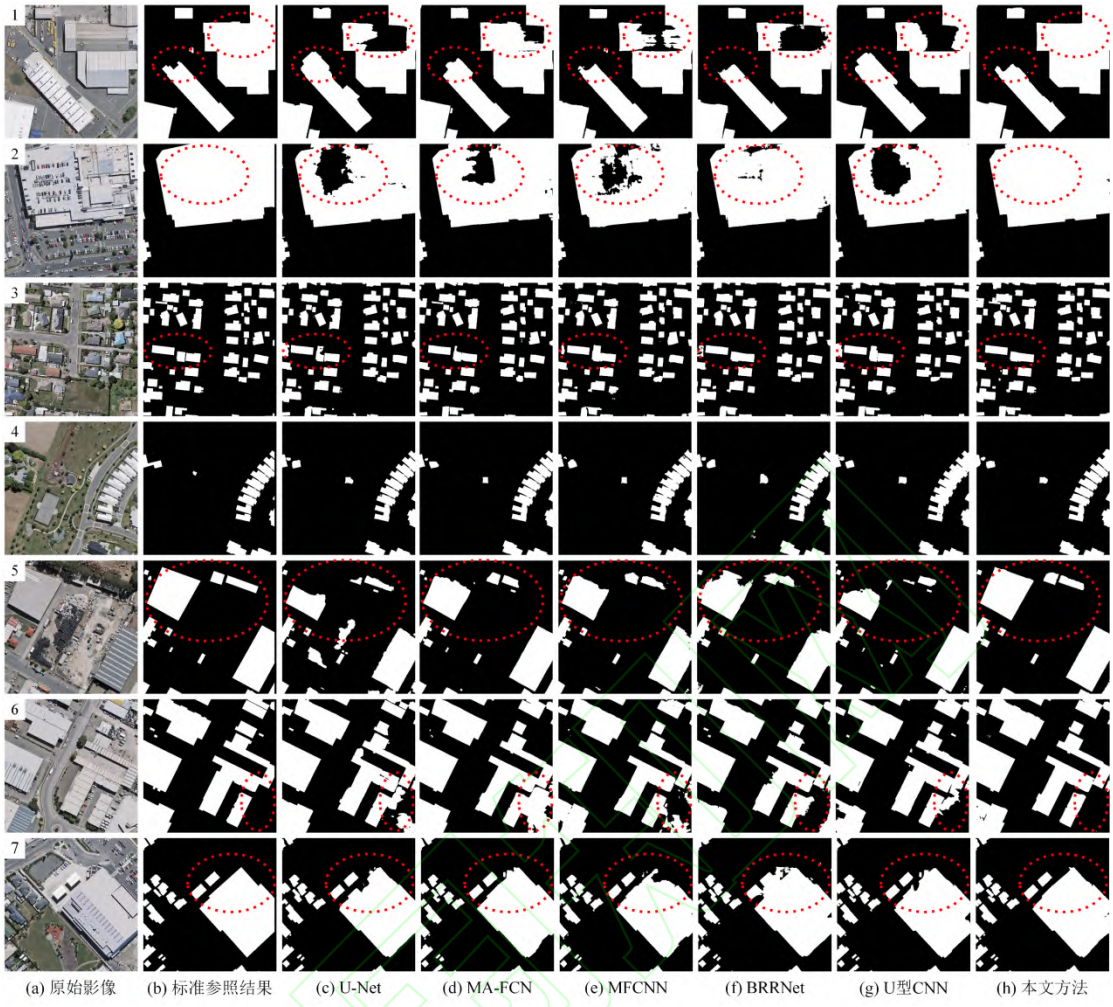

图7 WHU建筑物数据集上各方法的典型建筑物提取结果

Fig.7 Typical building extraction results of various methods on WHUbuilding dataset

3.4.3. Massachusetts 建筑物数据集实验结果及分析

各对比模型在 Massachusetts 建筑物数据集的定量评价结果如表 2 所示。由于 Massachusetts 建筑物数据集分辨率低于 WHU 建筑物数据集，且建筑物场景更为复杂，因此其总体定量评价结果低于 WHU 建筑物数据集。与对比模型相比，本文模型依旧在除准确率外的其余指标上表现最优，相较于 F_1 分数第二高的 BRRNet，IoU 和 F_1 分数分别高出了 1.17% 和 0.78%。此外，MA-FCN 模型在 WHU 建筑物数据集上表现较好，而在 Massachusetts 建筑物数据集上却仅高于基准 U-Net 模型，表明其特征提取能力鲁棒性较差。而本文模型在两个数据集上的定量评价结果均为最优，显示了本文模型对不同类型数据具有一定的通用性。

表 2 Massachusetts 建筑物数据集上各方法定量评价结果

Table 2 Quantitative evaluation result of several methods on the Massachusetts Building Dataset

方法	P /%	R /%	IoU/%	F_1/%
U-Net[12]	83.94	78.38	68.16	81.07
MA-FCN[16]	**87.90**	76.51	69.22	81.81
MFCNN[17]	83.69	81.24	70.14	82.45
BRRNet[14]	84.33	82.33	71.40	83.32
U 型 CNN[18]	80.18	82.43	68.47	81.29
本文方法	85.31	**82.93**	**72.57**	**84.10**

* 最佳定量评价值以加粗和下划线的形式突出显示

不同模型在 Massachusetts 建筑物数据集上的部分可视化结果如图 8 所示，其中影像 1、2 为密集小型建筑物，1-1 和 2-1 分别为影像 1 和影像 2 的局部细节放大图，影像 3 为大型建筑物，影像 4，5 为混合型建筑物集群。从视觉上观察分析，与对比模型相比，本文模型的提取结果与地面真实值更加接近。对于建筑物屋顶被阴影遮蔽或与周围伴生阴影光谱相近，造成光谱特征混淆的影像 1 和 2，四种对比模型提取结果中均出现了背景像素被错分为建筑物，建筑物边界粘连，以及部分建筑物缺失的现象。而本文模型通过增强网络对来自前一层级有效特征的聚合能力并构建多层级对象边界约束结构，为网络融入更多准确的建筑物特征以及边界信息，从而获得更为精确、规整的提取结果。此外，如图 8 中的影像 3、4、5 所示，当对受屋顶材质、阴影或其他人造物等因素影响而造成具有“同物异谱”现象的建筑物进行提取时，对比模型提取结果出现了较多空洞缺失现象。而本文模型通过 CSPM 中的连续小尺度空洞卷积从不同的尺度分别提取局部和全局特征，降低了提取过程中有效信息的损失，增强所提特征之间的关联程度，有效避免了因光谱特征不一致所带来的干扰，所提取的建筑物更加完整连续。

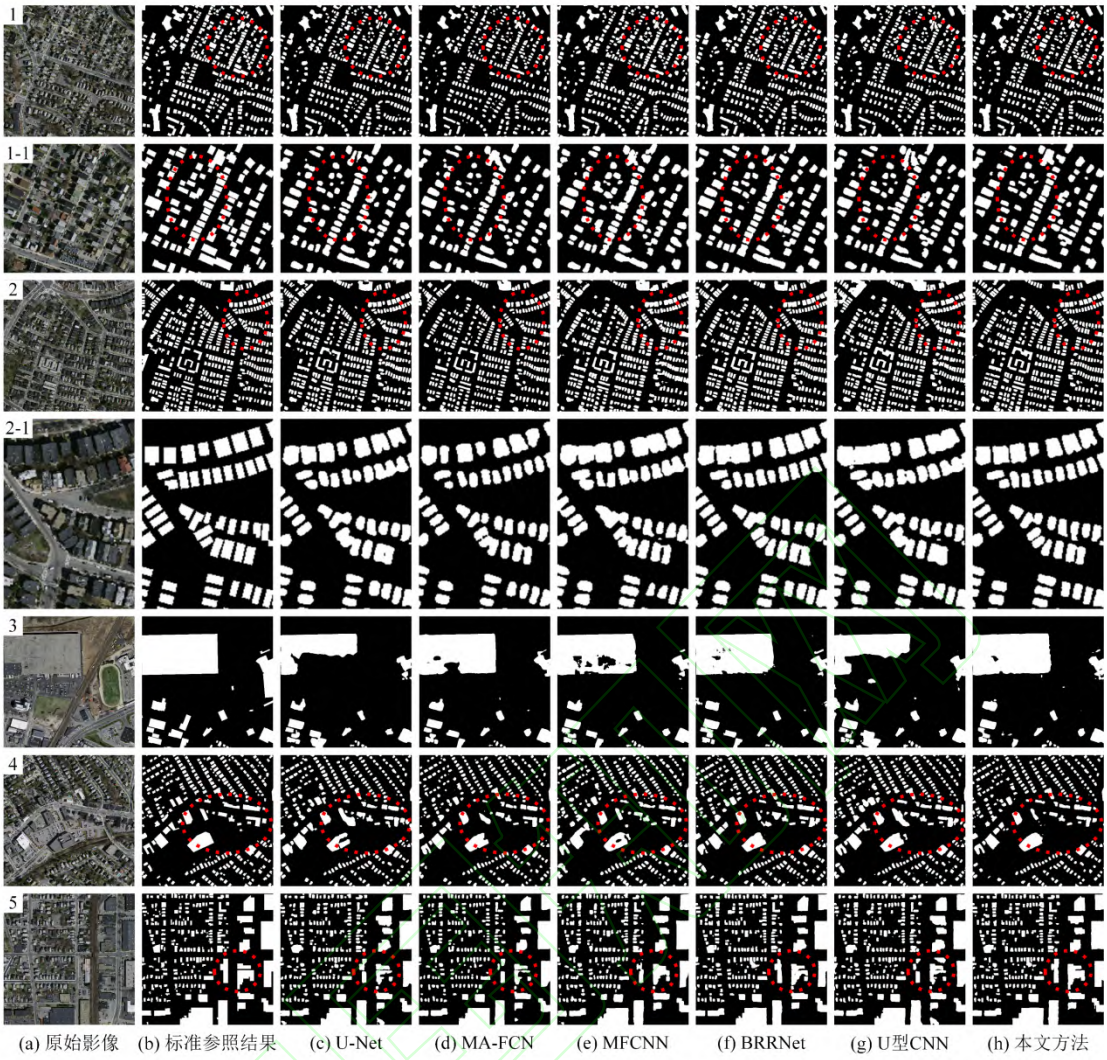

图8 Massachusetts建筑物数据集上各方法的典型建筑物提取结果

Fig.8 Typical building extraction results of various methods on Massachusetts building dataset

3.4.4. 模型复杂度及效率分析

在对网络模型性能进行综合评估时，除了精度外，模型的复杂度与效率同样是重要的表征指标。利用“ptflops”工具包对六种模型进行计算量和参数量的统计，并分别记录训练和预测耗时，结果如表 3 所示。由实验结果可知，本文模型计算量最小，参数量在六种方法中位列倒数第三，比最少的 BRRNet 多了约 5.82 百万，比最多的 MFCNN 少了 43.73 百万。综合计算量与参数量的统计结果，本文模型复杂度低于 U-Net、MFCNN、BRRNet 与 U 型 CNN。在引入注意力机制和 CSPM 后，本文模型的内存访问成本(MAC)升高，并行度降低，导致在 WHU 和 Massachusetts 建筑物数据集上训练与推理时间的花销高于基准 U-Net 模型，但远低于 MFCNN 与 BRRNet。与在 WHU 和 Massachusetts 建筑物数据集上训练和预测时间均

最短的MA-FCN相比，本文模型的IoU和F_1分数分别高出了0.69%、0.38%和3.35%、2.29%。此外，MA-FCN在WHU和Massachusetts建筑物数据集上的定量评价排名差异较大。综合上述分析来看，本文方法在牺牲了一定复杂度与效率的前提下，取得了更精确的建筑物提取结果，且具有更强的尺度鲁棒性。

表3 不同方法复杂度及计算效率对比

Table 3 Comparison of complexity and computational efficiency of different methods

方法	计算量（GFLOPs）	参数量/百万	WHU数据集		Massachusetts数据集	
			训练时间/(min/epoch)	预测时间/s	训练时间/(min/epoch)	预测时间/s
U-Net[12]	192.99	28.94	3.42	71	0.80	5.46
MA-FCN[16]	179.25	22.68	3.10	70	0.72	5.33
MFCNN[17]	302.42	66.88	5.26	136	1.20	7.17
BRRNet[14]	254.80	17.33	5.42	131	1.22	6.91
U型CNN[18]	192.99	28.94	3.32	98	0.86	5.21
本文方法	166.95	23.15	4.04	100	0.98	5.94

4. 结论

为减缓因光谱混淆而造成建筑物提取结果存在误分、漏分，以及边界粘结缺失的问题，提出了基于多特征融合与对象边界联合约束网络的高分辨率遥感影像建筑物提取方法。该方法以调整后的U-Net作为主体架构，首先在横向跳级连接中引入CBAM，对来自编码阶段的浅层特征进行了注意力分配；然后在编码阶段末端融入CSPM，减缓了多尺度特征提取过程中有效信息的损失；最后在解码阶段构建了基于对象和边界的MFCS，强化了建筑物有效特征的提取能力，细化了建筑物边界，并增强了浅层的特征的学习能力。综合不同方法在WHU和Massachusetts建筑物的实验结果可知，本文方法提高了建筑物检测准确率与完整度，保持了良好的边界，且具有较强尺度鲁棒性。然而，本文参数量依旧偏大，且过分依赖于对大量人工标签数据的训练和学习，构建轻量化的半监督建筑物提取网络将是未来研究的重点。

参考文献

[1] Wang Zhenqing, Zhou Yi, Wang Shixin, et al. House Building Extraction from High-Resolution Remote Sensing Images Based on IEU-Net[J]. *National Remote Sensing Bulletin*, 2021, 25(11): 2245-2254 (王振庆, 周艺, 王世新, 等. IEU-Net高分辨率遥感影像房屋建筑物提取[J]. 遥感学报, 2021, 25(11): 2245-2254)

[2] Wu Wei, Luo Jiancheng, Shen Zhanfeng, et al. Building Extraction from High Resolution Remote Sensing Imagery Based on Spatial-Spectral Method[J]. *Geomatics and Information Science of Wuhan University*, 2012, 37(7): 800-805 (吴炜, 骆剑承, 沈占锋, 等. 光谱和形状特征相结合的高分辨率遥感图像的建筑物提取方法[J]. 武汉大学学报·信息科学版,

2012, 37(7): 800-805)

[3] Lv Fenghua, Shu Ning, Gong Yan, et al. Regular Building Extraction from High Resolution Image Based on Multilevel-Features[J]. *Geomatics and Information Science of Wuhan University*, 2017, 42(5): 656-660 (吕凤华, 舒宁, 龚龑, 等. 利用多特征进行航空影像建筑物提取[J]. 武汉大学学报·信息科学版, 2017, 42(5): 656-660)

[4] Gao Xianjun, Zheng Xuedong, Shen Dajiang, et al. Automatic Building Extraction Based on Shadow Analysis from High Resolution Images in Suburb Areas[J]. *Geomatics and Information Science of Wuhan University*, 2017, 42(10): 1350-1357 (高贤君, 郑学东, 沈大江, 等. 城郊高分影像中利用阴影的建筑物自动提取[J]. 武汉大学学报·信息科学版, 2017, 42(10): 1350-1357)

[5] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA.: 3431-3440

[6] Szegedy C, Liu W, Jia Y Q, et al. Going deeper with convolutions[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA.: 1-9

[7] 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. Honolulu, HI, USA.: 6230-6239

[8] 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

[9] Chen L C, Papandreou G, Schroff F, et al. Rethinking Atrous Convolution for Semantic Image Segmentation[EB/OL]. 2017: arXiv: 1706.05587. https://arxiv.org/abs/1706.05587

[10] Li Daoji, Guo Haitao, Lu Jun, et al. A Remote Sensing Image Classification Procedure Based on Multilevel Attention Fusion U-Net[J]. *Acta Geodaetica et CartographicaSinica*, 2020, 49(8): 1051-1064 (李道纪, 郭海涛, 卢俊, 等. 遥感影像地物分类多注意力融和 U 型网络法[J]. 测绘学报, 2020, 49(8): 1051-1064)

[11] 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

[12] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation[C]// Medical Image Computing and Computer-Assisted Intervention-MICCAI, Munich, Germany, 2015

[13] Diakogiannis F I, Waldner F, Caccetta P, et al. ResUNet-A[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2020, 162: 94-114

[14] Shao Z F, Tang P H, Wang Z Y, et al. BRRNet: A Fully Convolutional Neural Network for Automatic Building Extraction from High-Resolution Remote Sensing Images[J]. *Remote Sensing*, 2020, 12(6): 1050

[15] Xu Jiawei, Liu Wei, Shan Haoyu, et al. High-Resolution Remote Sensing Image Building Extraction Based on PRCUnet[J]. *Journal of Geo-Information Science*, 2021, 23(10): 1838-1849 (徐佳伟, 刘伟, 单浩宇, 等. 基于 PRCUnet 的高分遥感影像建筑物提取[J]. 地球信息科学学报, 2021, 23(10): 1838-1849)

[16] 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

[17] Xie Y K, Zhu J, Cao Y G, et al. Refined Extraction of Building Outlines from High-Resolution Remote Sensing Imagery Based on a Multifeature Convolutional Neural Network and Morphological Filtering[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13: 1842-1855

[18] Wu Guangming, Chen Qi, Shibasaki R, et al. High Precision Building Detection from Aerial Imagery Using a U-Net Like Convolutional Architecture[J]. *Acta Geodaetica et CartographicaSinica*, 2018, 47(6): 864-872 (伍广明, 陈奇, Ryosuke SHIBASAKI, 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)

[19] Woo S, Park J, Lee J Y, et al. CBAM: Convolutional Block Attention Module[EB/OL]. 2018: arXiv: 1807.06521. https://arxiv.org/abs/1807.06521

[20] Wang P Q, Chen P F, Yuan Y, et al. Understanding convolution for semantic segmentation[C]//2018 IEEE Winter Conference on Applications of Computer Vision. Lake Tahoe, NV, USA.: 1451-1460

[21] Fu J, Liu J, Tian H J, et al. Dual attention network for scene segmentation[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA.: 3141-3149

[22] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA.: 936-944

[23] Milletari F, Navab N, Ahmadi S A. V-net: Fully convolutional neural networks for volumetric medical image segmentation[C]//2016 Fourth International Conference on 3D Vision (3DV). Stanford, CA, USA.: 565-571

[24] He S, Jiang W S. Boundary-Assisted Learning for Building Extraction from Optical Remote Sensing Imagery[J]. *Remote Sensing*, 2021, 13(4): 760

[25] Ji Shunping, Wei Shiqing. Building Extraction via Convolutional Neural Networks from an Open Remote Sensing Building Dataset[J]. *Acta Geodaetica et CartographicaSinica*, 2019, 48(4): 448-459 (季顺平, 魏世清. 遥感影像建筑物提取的卷积神经元网络与开源数据集方法[J]. 测绘学报, 2019, 48(4): 448-459)

[26] Mnih V. Machine Learning for Aerial Image Labeling[D]. Toronto: University of Toronto, 2013

Building Extraction Based on Multi-feature Fusion and Object-boundary Joint Constraint Network

GAO Xianjun[1] *RAN Shuhao*[1] *ZHANG Guangbin*[1] *YANG Yuanwei*[1, 2, 3]

1 School of Geosciences, Yangtze University, Wuhan 430100, China

2 Beijing Key Laboratory of Urban Spatial Information Engineering, Beijing Institute of Surveying and Mapping, Beijing 100045, China

3 Hunan Provincial Key Laboratory of Geo-Information Engineering in Surveying, Mapping and Remote Sensing, Hunan University of Science and Technology, Xiangtan 411201, China

Abstract: Objectives: Accurately and automatically extracting buildings from high-resolution remote sensing images is of great significance in many aspects, such as urban planning, map data updating, emergency response, etc.The problems of missing and wrong detection of buildings and missing boundaries caused by spectrum confusion still exist in the existing full convolution neural networks (FCN). **Methods:** In order to overcome the limitations, a multi-feature fusion and object-boundary joint constraint network was presented in this paper. The method is based on an encoding and decoding structure. In the encoding stage, the continuous-atrous spatial pyramid module (CSPM) is positioned at the end to extract and combine multi-scale features without sacrificing too much useful information. In the decoding stage, more accurate building features are integrated into the network and the boundary is refined, by implementing the multi-output fusion constraint structure (MFCS) based on object and boundary. In the skip connection between the encoding and decoding stages, the convolutional block attention module (CBAM) is introduced to enhance the effective features. Furthermore, the multi-level output results from the decoding stage are used to build a piecewise multi-scale weighted loss function for fine network parameter updating. **Results:** Comparative experimental analysis was performed on the WHU and Massachusetts building datasets. The results show that: (1) The buildings extraction results proposed by the proposed method are closer to the ground truth. (2) The quantitative evaluation result is higher than the other five state-of-the-art approaches. Specifically, IoU and F_1-Score on Massachusetts and WHU building datasets reached 90.44%, 94.98%, and 72.57%, 84.10%, respectively. (3) The proposed model outperforms the MFCNN and BRRNet in both complexity and efficiency. **Conclusions:** The proposed method not only improves the accuracy and integrity of extraction results in spectral obfuscation buildings, but also maintains a good boundary. It has strong scale robustness.

Key words: building extraction; fully convolutional neural network; multi-scale features; attention mechanism; joint constraint

First author: GAO Xianjun, PhD, associate professor, specializes in the theories and methods of automatic objects recognition from high resolution images. E-mail: junxgao@yangtzeu.edu.cn

Corresponding author: RAN Shuhao, master. E-mail: 201500880@yangtzeu.edu.cn

Foundation support: The Open Fund of Beijing Key Laboratory of Urban Spatial Information Engineering (No.20210205);Open Fund of National Engineering Laboratory for Digital Construction and Evaluation Technology of Urban Rail Transit (No.2021ZH02); Open Fund of Hunan Provincial Key Laboratory of Geo-Information Engineering in Surveying, Mapping and Remote Sensing, Hunan University of Science and Technology (No.E22133); the Open Research Fund of Key Laboratory of Earth Observation of Hainan Province (No.2020LDE001); the National Natural Science Foundation of China (No. 41872129).

网络首发：

标题：基于多特征融合与对象边界联合约束网络的建筑物提取

作者：高贤君，冉树浩，张广斌，杨元维

DOI：10.13203/j.whugis20210520

收稿日期：2022-06-20

引用格式：

高贤君，冉树浩，张广斌，等. 基于多特征融合与对象边界联合约束网络的建筑物提取[J]. 武汉大学学报· 信息科学版，2022，DOI：10.13203/j.whugis20210520（GAO Xianjun, RAN Shuhao,ZHANG Guangbin, et al. Building Extraction Based on Multi-feature Fusion and Object-boundary Joint Constraint Network[J]. *Geomatics and Information Science of Wuhan University*, 2022, DOI: 10.13203/j.whugis20210520）