基于注意力金字塔网络的航空影像建筑物变化检测

田青林[1*]，秦凯[1]，陈俊[2]，李瑶[3]，陈雪娇[1]

[1]核工业北京地质研究院遥感信息与图像分析技术国家级重点实验室，北京 100029；

[2]讯飞智元信息科技有限公司，安徽 合肥 230094；

[3]德州农工大学土木与环境工程系，德克萨斯州 77843

摘要 针对遥感图像语义分割中存在对多尺度目标的漏检和分割边界粗糙等问题，提出了一种基于注意力金字塔网络的航空影像建筑物变化检测方法。该方法采用编码-解码结构，在编码阶段使用 ResNet101 作为基础网络来提取特征，并在部分残差模块应用空洞卷积增大感受野，同时将金字塔池化结构作为编码网络的最后一层，以提取图像多尺度特征；在解码阶段的横向连接过程中引入注意力机制以突出重要特征，并采用自上而下的密集连接方式计算特征金字塔，有效融合不同阶段、不同分辨率的特征。在大型建筑物变化检测数据集上进行验证实验，实验结果表明所提方法在对不同尺寸建筑物目标的变化检测中展现出了良好的适应性，相比于经典语义分割网络具有一定的优势。

关键词 图像处理；变化检测；注意力机制；特征金字塔；空洞卷积

中图分类号 TP751.1 **文献标志码** A **doi**：10.3788/AOS202040.2110002

Building Change Detection for Aerial Images Based on Attention Pyramid Network

Tian Qinglin[1*], Qin Kai[1], Chen Jun[2], Li Yao[3], Chen Xuejiao[1]

[1] *National Key Laboratory of Remote Sensing Information and Image Analysis Technology, Beijing Research Institute of Uranium Geology, Beijing* 100029, *China*;

[2] *Iflytek Intelligent Information Technology Co., Ltd., Hefei, Anhui* 230094, *China*;

[3] *Zachry Department of Civil and Environmental Engineering, Texas A&M University, Texas* 77843, *USA*

Abstract Aiming at the problems in the semantic segmentation of remote sensing images, such as missed detection of multi-scale targets and rough segmentation boundary, we propose a method of building change detection for aerial images based on an attention pyramid network. The method adopts an encoding-decoding configuration. In the encoding phase, we utilize ResNet101 as the basic network to extract the features and apply dilated convolutions to improve the receptive field in partial residual modules. Meanwhile, the pyramid pooling structure is selected as the last layer of the encoding network to extract multi-scale features of the images. In the decoding phase, the attention mechanism is employed in lateral connection to highlight significant features, and the procedure of top-down dense connection is used to calculate the feature pyramid and then to fuse the features with different resolutions at different phases. Furthermore, the verification experiments are performed on the dataset of building change detection, and the results indicate that our method has good adaptability to different-size-building change detection and has certain advantages in comparison with the classical semantic segmentation networks.

Key words image processing; change detection; attention mechanism; feature pyramid; dilated convolution

OCIS codes 100.4996; 100.2000; 100.3008

1 引 言

变化检测是遥感领域重要的研究方向之一，在军事和民用领域（如军事目标动态监测、国土督察、灾害评估、城市规划等领域）都有着广泛的应用[1]。建筑物与人类生活密切相关，是组成城市的关键要素之

收稿日期：2020-06-30；**修回日期**：2020-07-06；**录用日期**：2020-07-15

基金项目：国家自然科学基金（41602333）、遥感信息与图像分析技术国家级重点实验室基金项目（ZJ2019-2）、国家重点实验室稳定支持科研项目（ZS1901）

* **E-mail**：736924158@qq.com

—[2]。伴随着城市化发展水平的不断提高，针对建筑物新建、拆除、改建、扩张等更新状况的动态监测对于国土部门督察意义重大。现实工作大多以人工实地调研、巡查取证等方式对城市建筑物开展调查，虽然精度有所保障但效率低下，耗费大量人力、物力和财力，又无法全方位、实时监管国土资源现状。随着遥感技术的快速发展，利用遥感影像直接进行建筑物变化检测能够大幅降低成本，并提高检测效率[3]。

基于传统图像处理的建筑物变化检测方法从最初基于像素的方法逐渐转向面向对象的方法，之后出现了形态学指数计算[4]、马尔可夫随机场[5]等诸多模型，此类方法主要以分割对象作为变化处理单元，分割对象比单个像素包含更多的光谱、空间和纹理信息，提高了变化检测的准确性和完整性，但此类方法是以图像分割为基础，分割效果的好坏直接影响变化检测的精度。

伴随着人工智能在近几年的兴起，深度学习技术得到了迅猛发展，并在语音识别、计算机视觉等领域广泛应用并取得了较好的效果。传统的卷积神经网络如 AlexNet[6]、GoogleNet[7]和 VGGNet[8]通常包含全连接层，要求输入图像大小固定，针对早期网络存在存储开销大、计算效率低等问题，Long 等[9]于 2015 年提出了一种可在任意大小图像上进行语义分割的全卷积网络(FCN)，以提高处理效率。然而，基于 FCN 进行语义分割的结果不够精细，特征图分辨率的降低牺牲了像素的空间位置信息，并且该方法只对单个像素进行独立分类，未能充分考虑丰富的空间信息，无法平衡利用图像局部和全局特征。针对上述问题，很多学者又提出了一系列新方法，例如：基于 FCN 的改进方法(以 DeepLab[10]系列为代表)通过带孔卷积、带孔空间金字塔池化等技术增大感受野，获得图像的多尺度信息，但对小尺度建筑物分割效果仍不理想；基于优化卷积结构的方法[以空洞卷积(DC)[11]和可变形卷积[12]为代表]，通过优化卷积结构增大感受野，减缓特征图分辨率的降低速度，但在一定程度上打破了建筑物像素的连续性，分割边界粗糙；基于编码-解码的方法(以 UNet[13]和 SegNet[14]为代表)通过反卷积或上池化等操作还原像素的空间位置信息，尽可能地避免特征图分辨率降低的问题，其缺点是计算量较大，从而影响建筑物语义分割的速度；基于特征融合的方法(以 RefineNet[15]和 PSPNet[16]为代表)通过空间金字塔池化和级联模型等手段捕获上下文信息，融合图像中不同尺度特征，使分割结果更为精细，但该方法容易出现建筑物部分边界丢失的现象。高分辨率遥感影像中不同建筑物的差异明显，尺度多样，已有学者提出在建筑物的提取中需要融合不同层次特征以确定该像素是否为建筑物[17]。如何有效提取并融合不同层次特征、精确捕捉建筑物边界、减少小尺度目标漏检现象，是本文重点关注的问题。

受上述研究启发，本文提出了一种基于注意力金字塔网络的航空影像建筑物变化检测方法，采用语义分割的思路检测建筑物的变化。模型整体采用编码-解码结构，在编码阶段应用空洞卷积改进特征提取效果，使用金字塔池化模块(PPM)提取图像多尺度特征；在解码阶段的横向连接过程中引入注意力机制(AM)以突出重要特征，并采用自上而下的密集连接方式计算特征金字塔，充分融合不同阶段、不同分辨率的特征，用于多尺度建筑物目标的变化检测；最后，通过与经典语义分割网络进行对比分析，验证了本文所设计网络的有效性。

2　相关工作

得益于计算机硬件的不断发展，在图像目标检测、场景分类的相关研究中，深度神经网络得到了迅速发展，如空洞卷积、PPM、注意力机制等一系列网络结构和方法得以应用。

2.1　空洞卷积

在语义分割领域，经典的网络模型需要经过多层卷积和池化进行特征提取，从而找到分类目标，但在这个过程中图像尺寸逐渐减小，导致图像细节结构和边缘信息丢失，分割结果不够精细。针对这一问题，目前主要有两种解决方法：一种是编码-解码结构，通过编码逐步减小特征图，从而学习抽象特征，通过解码逐渐恢复图像大小和目标细节，但通过解码恢复丢失图像细节信息的效果有限；另一种方法是取消池化层并使用空洞卷积，在不增加额外参数、不降低特征图分辨率的情况下扩大卷积核感受野，保留多尺度特征和图像细节信息，但随着网络层次和特征图数量的增加，保持特征图尺寸不变势必导致计算量变大[10]。结合两种解决方法的优势，本文通过在编码-解码结构网络中加入空洞卷积，获取更多的细节结构信息，从而更好地提取建筑物的多尺度特征。

空洞卷积是在正常连续卷积操作中加入不同的间隔(间隔大小由扩张率决定)，可以在不损失分辨率、不增加网络参数的情况下增大感受野。普通卷积和空洞卷积的区别如图 1 所示，普通卷积即扩张率(记为 r)为 1 的空洞卷积。

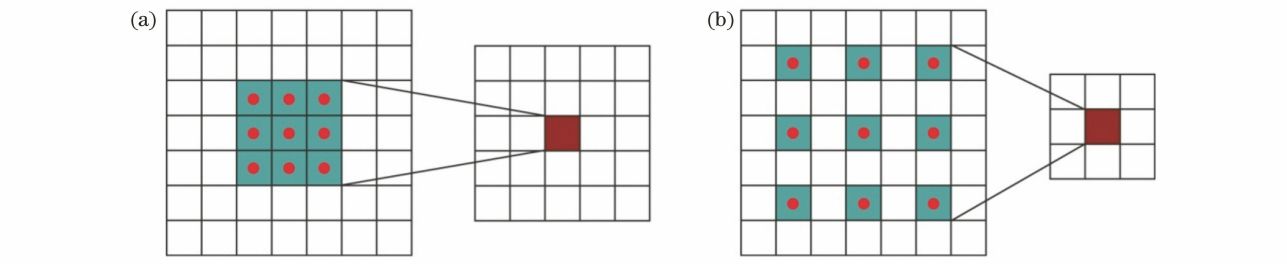

图 1　普通卷积和空洞卷积的示意图。(a)普通卷积;(b)空洞卷积

Fig. 1　Diagrams of (a) traditional convolution and (b) dilated convolution. (a) Traditional convolution; (b) dilated convolution

2.2　金字塔池化模块

遥感图像中建筑物类型多样,尺寸大小也各不相同,而卷积神经网络中连续的池化操作会导致空间信息丢失,传统的池化操作只在某个固定大小的窗口内进行下采样,且使用的上下文信息(即感受野)有限,导致部分不同尺度建筑物的分割效果欠佳。针对上述问题,本文在特征提取网络的最后一层引入 PPM,提取图像不同层次特征,以提高多尺度建筑物的分割精度。PPM 由一组不同尺度的多级池化块组成,可以融合不同子区域之间的多尺度信息,其结构如图 2 所示[16]。

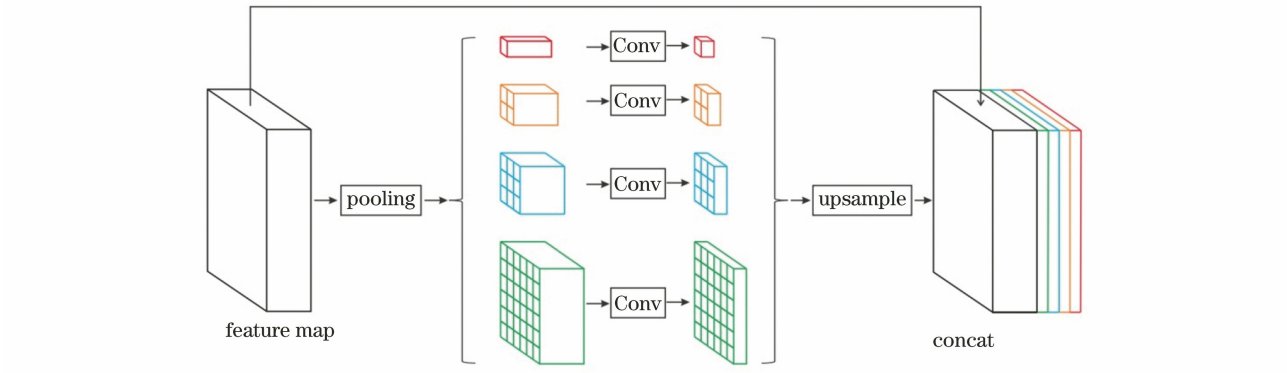

图 2　PPM 结构示意图

Fig. 2　Structural diagram of PPM

2.3　注意力机制

深度学习领域的注意力机制的核心目标是从丰富的目标提取特征中选择出对当前任务更为关键的显著特征,广泛应用于自然语言处理、目标检测以及语音识别等任务。卷积块注意模块(CBAM)是一种简单有效的注意力模块,在通道和空间两个维度上引入注意力机制,其中通道注意力应用在全局信息,而空间注意力集中在局部信息,通过二者的结合能更有效地获取目标显著特征,在不明显增加参数和运算量的前提下提升网络性能[18]。

CBAM 结构如图 3 所示,分别在网络的通道维度及空间维度上进行特征压缩和生成权重并重新进行加权。首先给定输入特征图 $\boldsymbol{F}\in\mathbf{R}^{C\times H\times W}$,通过通道注意力机制得到通道注意力图 $\boldsymbol{M}_C\in\mathbf{R}^{C\times 1\times 1}$,将 $\boldsymbol{F}$ 与 $\boldsymbol{M}_C$ 两个矩阵的对应元素相乘后得到经通道注意力机制的特征图 $\boldsymbol{F}_C\in\mathbf{R}^{C\times H\times W}$;然后将 $\boldsymbol{F}_C$ 经空间注意力机制得到空间注意力图 $\boldsymbol{M}_S\in\mathbf{R}^{1\times H\times W}$;最后将 $\boldsymbol{F}_C$ 与 $\boldsymbol{M}_S$ 两个矩阵的对应元素相乘后得到显著特征图 $\boldsymbol{F}_R\in\mathbf{R}^{C\times H\times W}$。CBAM 的计算过程为[19]

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

$$\boldsymbol{F}_R=\boldsymbol{M}_S(\boldsymbol{F}_C)\otimes\boldsymbol{F}_C,\tag{2}$$

式中:⊗为哈达玛积,其定义为两个矩阵对应元素的乘积。

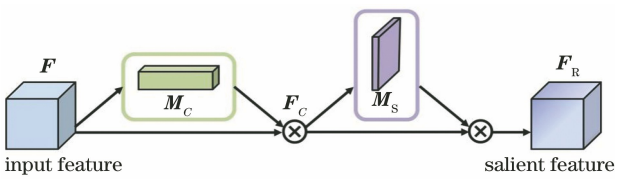

图 3　CBAM 结构示意图

Fig. 3　Structural diagram of CBAM

3　本文方法

本文算法的网络结构如图 4 所示,可分为编码

和解码两个部分。在编码部分采用 ResNet101 作为基础网络提取图像的多层特征，将具有相同尺寸特征图的结构称为阶段，如表 1 所示。首先把原始 ResNet101 网络中的卷积层分成 5 个阶段，分别表示为 conv1、conv2、conv3、conv4 和 conv5，并将 conv2～conv5 阶段最后一个残差模块输出的特征图$\{C_2,C_3,C_4,C_5\}$组成自下而上的前向网络，再通过 1×1 大小的卷积层将特征图进行通道维度上的压缩。此外，在 conv4 和 conv5 阶段的残差模块中应用空洞卷积以增大感受野，使得在特征提取时保留较大的特征尺寸。ResNet101 网络中 conv2～conv5 阶段输出的特征尺寸大小分别为原始图像的 1/4，1/8，1/16，1/32，本文分别使用扩张率为 2 和 4 的空洞卷积代替 conv4 和 conv5 阶段的普通卷积，将 conv2～conv5 阶段输出的特征尺寸大小增大为 1/4，1/8，1/8，1/8，更大的特征尺寸有利于后续的特征融合，能够弥补视野上的缺陷，使网络较少丢失特征的空间信息，从而提升网络的整体性能。此外，为了减少分割时多尺度建筑物目标存在的漏检现象，对特征编码阶段最后一个特征图 C_5 进行金字塔池化操作得到特征图 C_5'，以更好地提取不同大小局部图像的上下文信息，进一步获取多尺度特征，优化建筑物分割结果。

表 1　编码阶段特征提取网络结构

Table 1　Architecture of feature extraction network in encoding stage

ResNet101 convolutional layer	Stage name	Output feature	Output scale
7×7，64，stride 2	conv1	C_1	1/2
3×3，max pooling，stride 2			
$\begin{pmatrix}1\times1,64\\3\times3,64\\1\times1,256\end{pmatrix}\times3$	conv2	C_2	1/4
$\begin{pmatrix}1\times1,128\\3\times3,128\\1\times1,512\end{pmatrix}\times4$	conv3	C_3	1/8
$\begin{pmatrix}1\times1,256\\3\times3,256\\1\times1,1024\end{pmatrix}\times23$(DC)	conv4	C_4	1/8
$\begin{pmatrix}1\times1,512\\3\times3,512\\1\times1,2048\end{pmatrix}\times3$ (DC)	conv5	C_5	1/8

自下而上的前向网络中，高层特征包含丰富的语义信息，但缺少空间分辨率信息，低层特征具有较高的分辨率，但缺少语义信息。在多尺度建筑物变化检测中所提取的特征既要包含足够的空间分辨率信息用于定位小尺度建筑物，也要包含丰富的语义信息以有效区分建筑物目标与其他干扰信息。因此，有效融合高层语义信息和低层空间分辨率信息十分重要。在解码部分的横向连接过程中，引入 CBAM 可使网络更好地区分特征之间的重要程度，突出有用特征，对特征图$\{C_2,C_3,C_4,C_5\}$分别进行 CBAM 操作，得到逐级增强后的特征图$\{M_2,M_3,M_4,M_5\}$。再将特征图 C_5' 与$\{M_2,M_3,M_4,M_5\}$进行逐级叠加，使用自上而下的密集连接方式计算特征金字塔。具体过程如图 4 所示，首先采用 concatenate 方式将金字塔池化操作输出的特征图 C_5' 与经 CBAM 操作得到的特征图 M_5 进行叠加，之后通过一次 1×1 大小卷积操作将通道数量减少为原来的 1/2，得到融合后的特征图 P_5，类似地，计算 P_4 和 P_3。而 P_2 的计算方式如图 5 所示，需要将特征图 P_5、P_4 和 P_3 分别进行 2 倍上采样，然后将

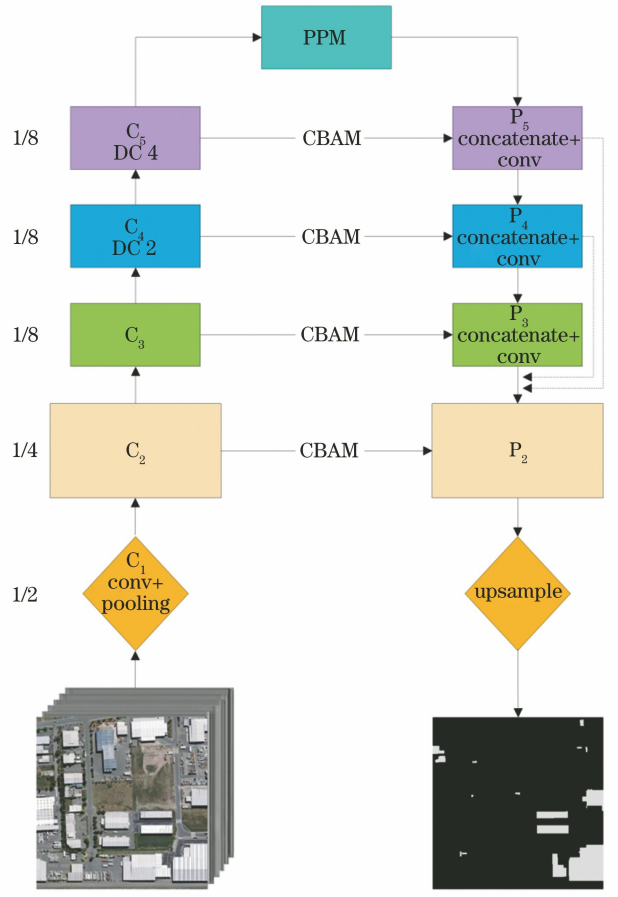

图 4　网络结构图

Fig. 4　Illustration of network architecture

3 个层次上采样结果以 concatenate 方式进行叠加，再将合并结果与 M_2 进行连接，得到融合后的特征图 P_2。至此，通过计算得到融合后的自上而下的网络分支{P_2，P_3，P_4，P_5}，使得每一层次特征图都充分融合了不同分辨率、不同语义强度的特征，提高了对多尺度建筑物目标的变化检测效果；最后，将特征金字塔底端的特征图 P_2 上采样至原始图像大小，将其输入分类器，得到检测结果。

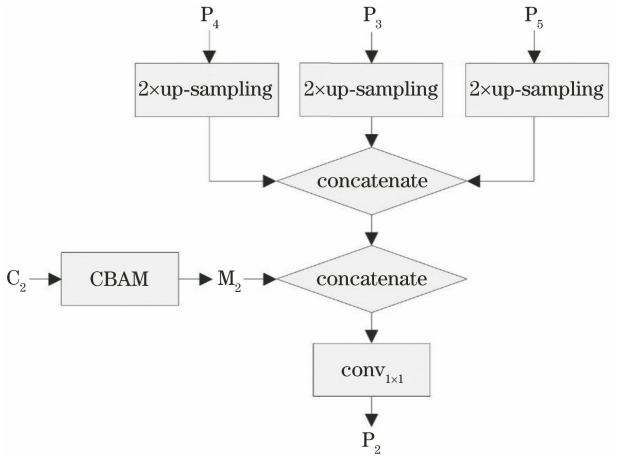

图 5　不同尺度特征的融合

Fig. 5　Fusion of features with different scales

4　实验过程及结果分析

4.1　数据集与预处理

数据集选用 WHU building dataset[20]，该数据集由 1 个航空影像子数据集、2 个卫星影像子数据集和 1 个建筑物变化检测子数据集组成。其中，用于本文网络训练和测试的是建筑物变化检测子数据集，该数据集包含 2012 年和 2016 年拍摄的两个时相影像的覆盖面积达 20.5 km^2 的新西兰地区航空影像及其对应的地面建筑物变化真值图。2012 年和 2016 年拍摄的影像均为 15354 pixel×32507 pixel 的红、绿、蓝 3 通道图，空间分辨率为 0.2 m，配准精度为 1.6 pixel，分别包括了 12796 和 16077 个建筑物对象。

将数据集中的两个时相影像进行通道合并得到 6 通道数据，由于网络参数量较大，为了降低对计算机性能的要求，将通道合并后的影像滑动切割成 256 pixel×256 pixel 的图像块，再按照 7∶3的比例将其随机划分为训练数据和测试数据用于网络训练与测试。

深度学习中的神经网络包含大量参数，为确保这些参数的正确性，需要利用充足的样本数据进行训练，样本数量越多，网络训练的效果越好，泛化能力越强，从而可以有效防止过拟合现象的发生。为此，对前面切割得到的小尺寸训练样本图像及其标签进行数据增强，分别将图像进行水平变换，垂直翻转变换，以及逆时针旋转 90°、180°、270°，部分变换效果如图 6 所示。

网络预测结果中会存在部分孤立点、空洞等噪声，为了使结果更准确，同时为了保证对比的一致性，本文对后续实验中所有方法的结果均采用腐蚀、膨胀和空洞填充等形态学操作来去除干扰、平滑边界，从而进一步细化预测结果。预测结果经本文方法处理前后的对比如图 7 所示。

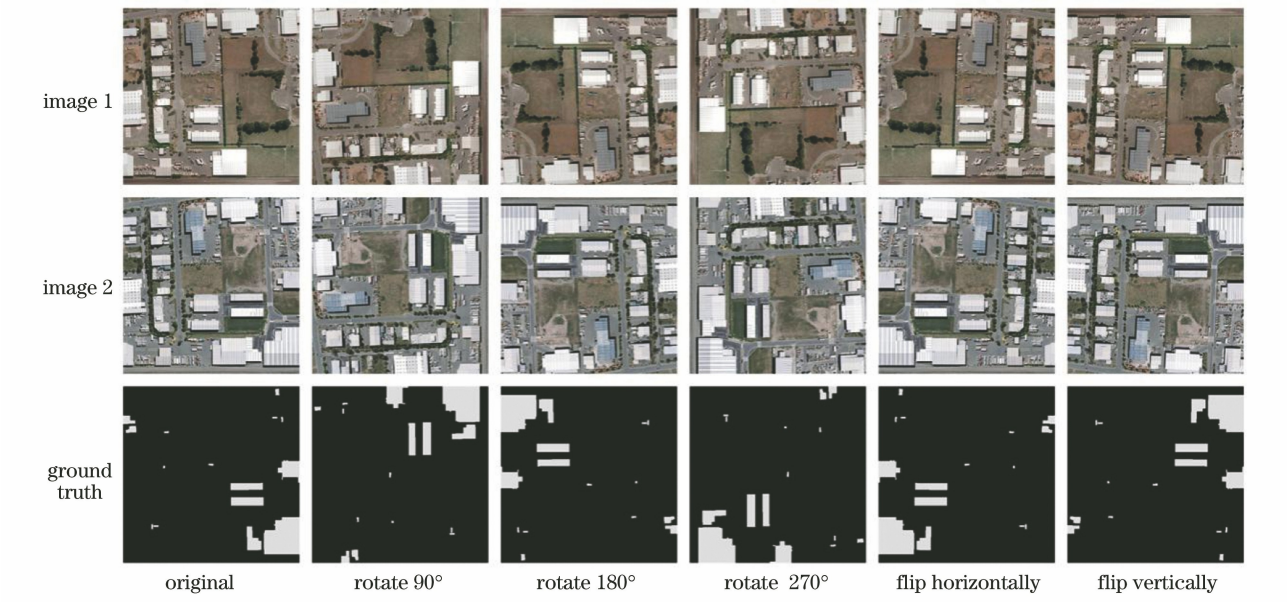

图 6　数据增强示意图

Fig. 6　Illustration of data augmentation

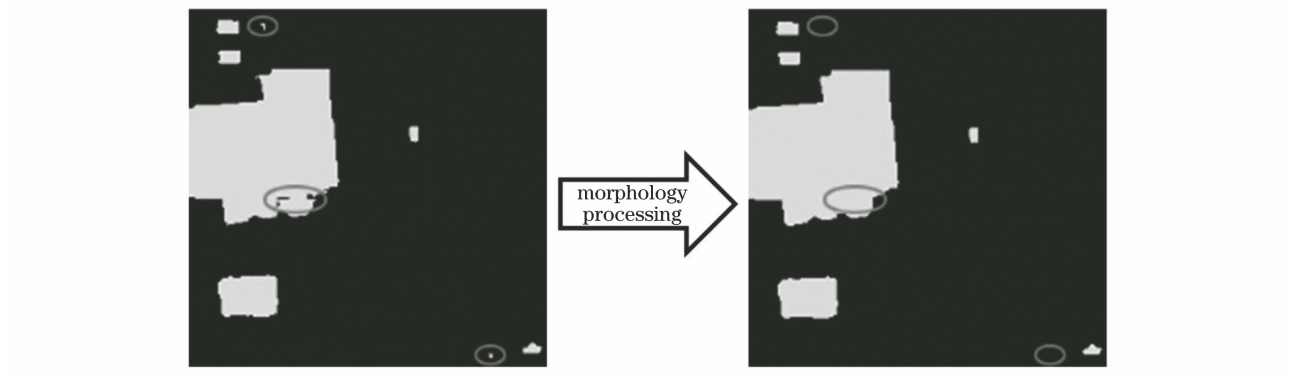

图 7　预测结果经本文方法处理前后的对比

Fig. 7　Comparison of results before and after post-processing by proposed method

4.2　参数设置及实验平台

实验环境设置：CPU 为 i7-8700K 处理器，显卡为 NVIDIA GeForce GTX 1080Ti，内存容量为 64 GB，采用 PyTorch 深度学习框架完成模型的构建。训练时使用 Adam 作为优化器进行迭代求解，动量设置为 0.9，衰减因子为 0.0005，基础学习率为 0.0001，损失函数为 Dice Loss，不断进行正反向传播，满足迭代次数后停止训练。

4.3　评价指标

本文采用准确率(P)、召回率(R)和 F_1 值作为建筑物变化检测结果的评价指标。其中，准确率表示预测结果中被预测为变化的像素中实际为变化像素的比例，召回率表示预测结果中被预测为变化的像素占所有实际为变化像素的比例。F_1 值是准确率与召回率的加权调和平均值，其计算公式为

$$F_1=\frac{2PR}{P+R},\tag{3}$$

式中：P 为准确率；R 为召回率。

4.4　实验结果与分析

针对本文所设计的网络，为了验证空洞卷积、CBAM 和 PPM 对网络整体性能的影响，采用如下方式进行消融实验分析：1）删除本文网络中的 DC 模块，保留 PPM 和 CBAM，将该消融模型标记为 A；2）删除本文网络中的 CBAM，保留 DC 和 PPM，将该消融模型标记为 B；3）删除本文网络中的 PPM，保留 DC 和 CBAM，将该消融模型标记为 C。表 2 给出了消融实验的定量结果，图 8 展示了该情况下消融实验的定性结果。

从表 2 可以看出，本文网络中使用到的三个模块在多项评价指标中都发挥了重要作用，删除任何一个模块都会导致网络的检测精度下降。其中，F_1 分值综合了准确率和召回率两项指标因子，更能全面体现网络的变化检测性能，当消融实验中缺少 DC 模块时，网络性能下降最明显，F_1 值降低了 2.14 个百分点，该情况对精度的影响最大；当缺少 CBAM 时，F_1 值降低了 0.78 个百分点，该情况对精度的影响较小；当缺少 PPM 时，F_1 值降低了 0.64 个百分点，该情况对精度的影响最小。消融实验结果说明了 DC、CBAM 和 PPM 三个模块在多尺度建筑物变化检测任务中的必要性，且三个模块对网络整体性能影响的程度依次递减。

表 2　网络的消融实验分析

Table 2　Ablation experiment analysis of network

Ablation experiment	DC	CBAM	PPM	P /%	R /%	F_1/%
A	×	✓	✓	85.16	83.10	84.11
B	✓	×	✓	83.94	87.07	85.47
C	✓	✓	×	84.54	86.71	85.61
Ours	✓	✓	✓	84.47	88.10	86.25

从图 8 中定性结果可以看出，去除 DC 模块的消融模型 A 的预测结果最差，尤其是对小尺度建筑物的变化检测存在严重漏判现象，检测结果不够精细，这说明 DC 模块对网络整体性能的影响最大。导致上述现象的可能原因是删除 DC 模块的特征提取网络在进行卷积、池化的过程中，图像尺寸减小，导致建筑物的细节结构和边缘信息丢失，尤其对中小型建筑物的分割结果不够精细。去除 CBAM 的消融模型 B 相对消融模型 A 有所改进，尤其是对于大尺度建筑物的变化的检测面积更完整，但对于小型建筑物变化区域的漏检现象依然存在，其可能原因是 CBAM 通过对提取的特征进行筛选加权，突出了与建筑物变化目标相关的显著特征而舍弃了其他无用信息，使得变化检测面积更完整，但未能有效捕获小尺度建筑物的细节结构和边缘信息。相对而言，去除 PPM 的消融模型 C 效果更好，检测面积不足、漏检等现象均得到较大改善，这说明 PPM 对网络整体性能的影响最小，也从另一个角度验证了

DC 和 CBAM 能够增强网络的特征提取能力，尤其是可以提高对中小尺度建筑物目标的变化检测精度。本文网络同时引入三个模块，显著提高了网络对多尺度建筑物目标变化的检测精度，改善了建筑物变化检测结果不完整、不连续及漏检等现象，F_1 值综合指标达到最优，这证明了三个模块结合的有效性。此外，值得注意的是，图 8(d)中方框区域有一处目标被本文网络检测为建筑物变化区域，但其在图 8(c)标注的地面真值中却未显示，通过对比图 8(h)中原始影像发现，此处为建筑物变化位置，属于数据集中真值标签漏标注对象，从而进一步证明了本文网络对建筑物变化检测的准确性和可靠性。

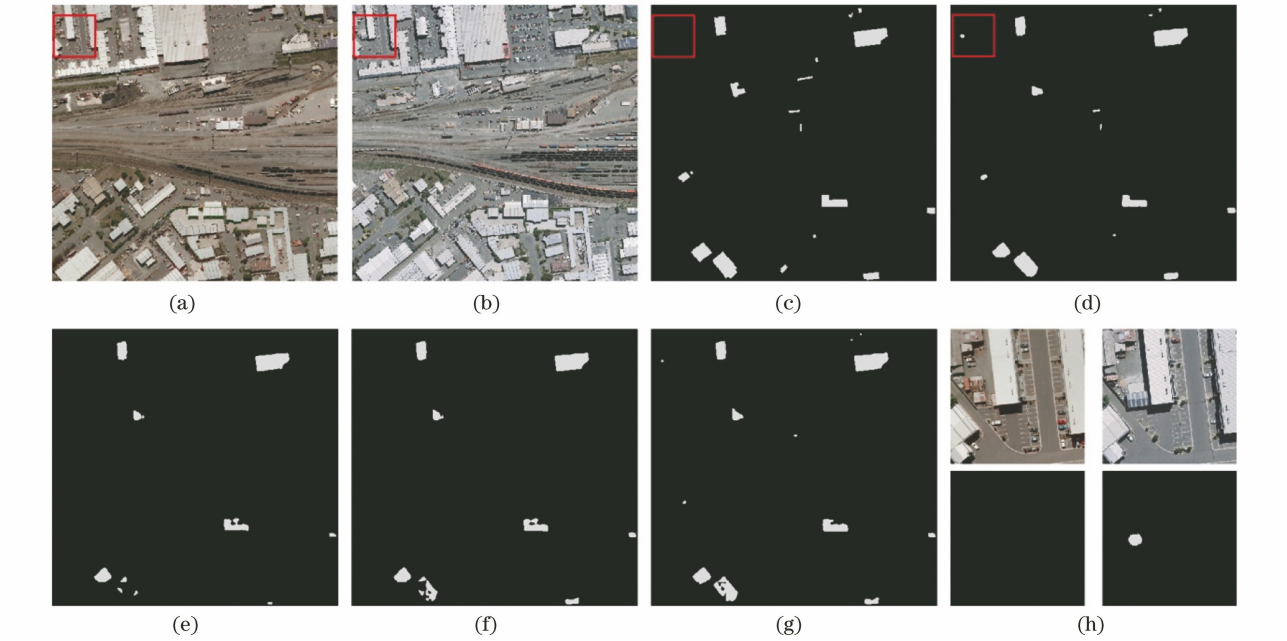

图 8　网络的消融模型对比实例图。(a)图像 1；(b)图像 2；(c)地面真值；(d)所提网络；(e)消融模型 A；(f)消融模型 B；(g)消融模型 C；(h)方框部分放大图

Fig. 8　Comparison of examples of network ablation models. (a) Image 1; (b) image 2; (c) ground truth; (d) proposed network; (e) ablation model A; (f) ablation model B; (g) ablation model C; (h) enlarged drawing in box

为了进一步验证本文网络的效果，将本文网络与 UNet[13]、DeepLab[10]、CSCDNet[21]、UPerNet[22]等经典语义分割网络的检测精度进行对比，精度如表 3 所示。从表中可以看出，与上述语义分割网络相比，本文所设计网络的准确率、召回率和 F_1 值等多项指标均有明显提高，检测效果更好。

表 3　不同方法建筑物变化检测精度评价

Table 3　Accuracy assessment of building change detection by different methods

Method	P /%	R /%	F_1 /%
UNet	73.09	42.84	54.02
DeepLab	77.73	51.41	61.89
CSCDNet	81.08	69.60	74.90
UPerNet	78.66	71.99	75.18
Ours	84.47	88.10	86.25

实验数据集中包含了居住区中小型建筑物和商业区大型建筑物等多种类型，用于衡量网络对于多尺度建筑物目标的检测性能。图 9 展示了不同网络对航空影像中小型规律性建筑物变化的检测结果，通过将其与地面真值标签比较后发现，5 种方法基本都能检测出建筑物的变化范围，但是 UNet 网络只能大致指出建筑物的变化区域，且边界粗糙，检测面积明显不足，存在较严重的误检、漏检现象。DeepLab 网络由于采用了 DC，相较 UNet 有一定改进，检测面积更完整，图 9(e)右下角区域的两处变化对象被检测出来，误检、漏检现象有所减少。CSCDNet 网络使用 Siamese ResNet18 作为编码阶段的特征提取网络，并引入相关联层克服传感器视点的影响，对建筑物变化的检测效果更好，但对小尺度建筑物目标的检测依然存在困难。UPerNet 网络由于同时使用了 PPM 和特征金字塔结构，检测效果得到明显改善，对建筑物边缘的提取更为准确，但同上述方法一样对小尺度建筑物的变化仍存在较多的误检、漏检现象，本文网络在此基础上引入 DC 和 CBAM，并采用密集连接的方式计算特征金字塔的融合多尺度特征，进一步改善了对中小尺度建筑物变化的检测能力，检测精度得到明显提高。

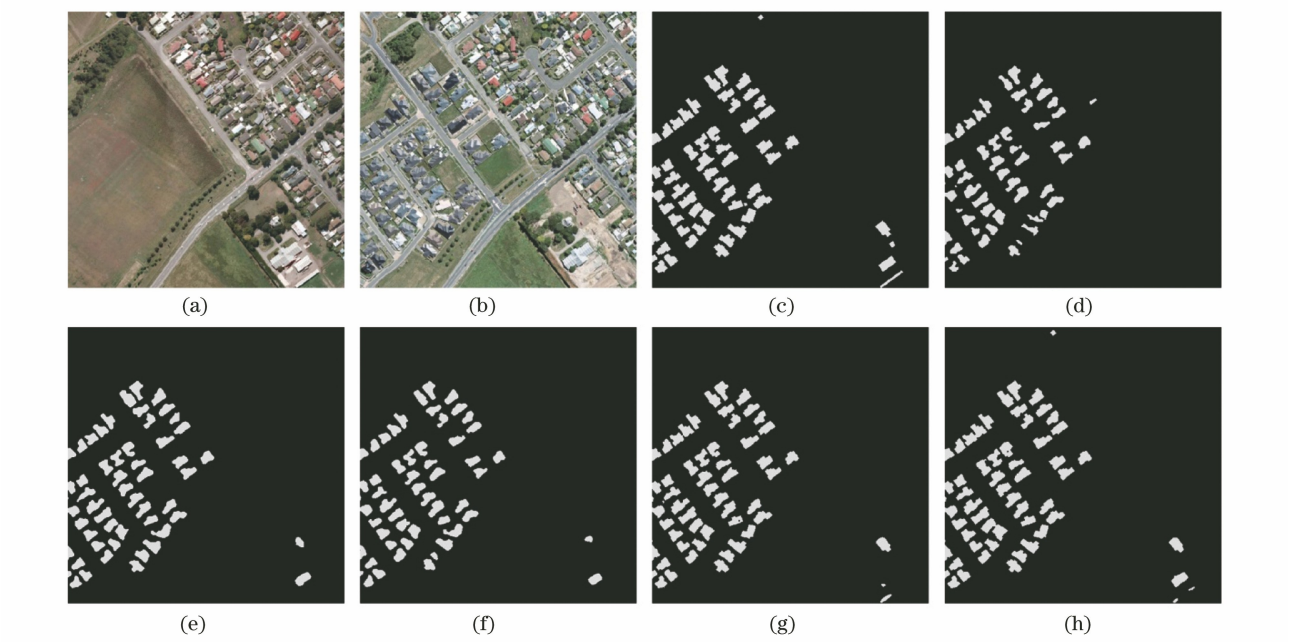

图 9　对规律性建筑物变化的检测结果对比。(a)图像 1;(b)图像 2;(c)地面真值;(d) UNet;(e) DeepLab;(f) CSCDNet;(g) UPerNet;(h)所提网络

Fig. 9　Results of detection for ordered building change. (a) Image 1; (b) image 2; (c) ground truth; (d) UNet; (e) DeepLab; (f) CSCDNet; (g) UPerNet; (h) proposed network

图 10 展示了不同网络对多尺度建筑物变化的检测结果,通过与地面真值标签进行比较后发现,本文网络对不同尺度建筑物丰富特征的提取能力较强,大、中、小多尺度建筑物变化目标均可以被很好

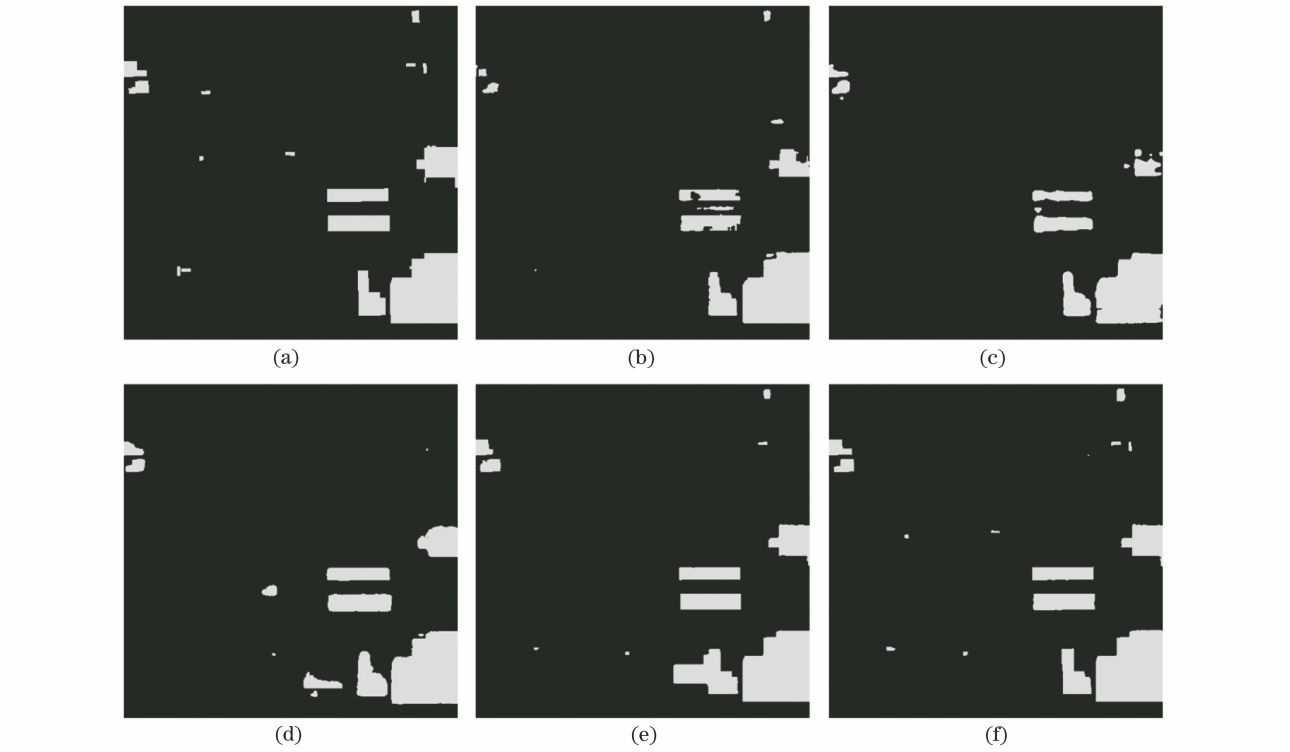

图 10　多尺度建筑物变化检测结果对比。(a)地面真值;(b) UNet;(c) DeepLab;(d) CSCDNet;(e) UPerNet;(f)所提网络

Fig. 10　Results of detection for multi-scale building change. (a) Ground truth; (b) UNet; (c) DeepLab; (d) CSCDNet; (e) UPerNet; (f) proposed network

地检测出来，并且本文网络改善了检测大型建筑物变化时存在的空洞、不完整等问题，准确度、边缘光滑明晰度方面明显优于 UNet、DeepLab、CSCDNet 和 UPerNet 等网络。UNet、DeepLab、CSCDNet 和 UPerNet 网络的特征表达能力欠缺，对多尺度建筑物目标分割不够精细，边缘错误较多，尤其是对小尺度建筑物变化目标存在较明显的误检、漏检现象。

5 结　论

针对遥感图像语义分割中存在对多尺度目标的漏检和分割边界粗糙等问题，提出一种基于注意力金字塔网络的航空影像建筑物变化检测方法。在编码阶段的特征提取部分加入空洞卷积和 PPM，一方面可在扩大建筑物图像特征映射感受野的同时保留特征尺寸，另一方面通过多尺度提取图像特征保留了中小尺度建筑物的更多特征。在解码阶段的横向连接过程中引入 CBAM 可以有效恢复中小尺度建筑物的特征信息，并采用自上而下的密集连接方式计算特征金字塔，充分融合不同阶段、不同分辨率的特征。对 WHU building dataset 建筑物变化检测数据集进行测试，发现本文方法具有较高的准确率，本文方法与传统的 UNet、DeepLab、CSCDNet 和 UPerNet 等语义分割网络相比整体性能较优，分割结果可观。在后续研究中，将尝试搜集、制作卫星影像建筑物变化检测数据集，以进一步评估网络的泛化能力。

参　考　文　献

[1] Shi W Z, Zhang P L. State-of-the-art remotely sensed images-based change detection methods[J]. Geomatics and Information Science of Wuhan University, 2018, 43(12): 1832-1837.
史文中, 张鹏林. 光学遥感影像变化检测研究的回顾与展望[J]. 武汉大学学报·信息科学版, 2018, 43(12): 1832-1837.

[2] Huang X, Zhu T T, Zhang L P, et al. A novel building change index for automatic building change detection from high-resolution remote sensing imagery[J]. Remote Sensing Letters, 2014, 5(8): 713-722.

[3] Zhu T Y, Dong F, Gong H X. Remote sensing building detection based on binarized semantic segmentation[J]. Acta Optica Sinica, 2019, 39(12): 1228002.
朱天佑, 董峰, 龚惠兴. 基于二值语义分割网络的遥感建筑物检测[J]. 光学学报, 2019, 39(12): 1228002.

[4] Huang X, Zhang L P, Zhu T T. Building change detection from multitemporal high-resolution remotely sensed images based on a morphological building index[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2014, 7(1): 105-115.

[5] Bruzzone L, Prieto D F. Automatic analysis of the difference image for unsupervised change detection [J]. IEEE Transactions on Geoscience and Remote Sensing, 2000, 38(3): 1171-1182.

[6] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks [J]. Communications of the ACM, 2017, 60(6): 84-90.

[7] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions [C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA, USA. New York: IEEE Press, 2015: 1-9.

[8] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2015-04-10)[2019-07-08]. https://arxiv.org/abs/1409.1556.

[9] 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.

[10] 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.

[11] Yu F, Koltun V, Funkhouser T. Dilated residual networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 636-644.

[12] Dai J F, Qi H Z, Xiong Y W, et al. Deformable convolutional networks[C]//2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 764-773.

[13] Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation [M]//Lecture Notes in Computer Science. Cham: Springer International Publishing, 2015: 234-241.

[14] Zhang Z H, Fang W, Du L L, et al. Semantic segmentation of remote sensing image based on encoder-decoder convolutional neural network [J]. Acta Optica Sinica, 2020, 40(3): 0310001.
张哲晗, 方薇, 杜丽丽, 等. 基于编码-解码卷积神经网络的遥感图像语义分割[J]. 光学学报, 2020, 40(3): 0310001.

[15] Lin G S, Milan A, Shen C H, et al. RefineNet: multi-path refinement networks for high-resolution semantic segmentation [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 5168-5177.

[16] 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), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 6230-6239.

[17] Yuan J Y. Automatic building extraction in aerial scenes using convolutional networks [C]// Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016.

[18] Wang Y, Liu L B. Bilinear residual attention networks for fine-grained image classification [J]. Laser & Optoelectronics Progress, 2020, 57(12): 121011.
王阳, 刘立波. 面向细粒度图像分类的双线性残差注意力网络[J]. 激光与光电子学进展, 2020, 57(12): 121011.

[19] Cui Z Y, Li Q, Cao Z J, et al. Dense attention pyramid networks for multi-scale ship detection in SAR images [J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57(11): 8983-8997.

[20] 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.

[21] Sakurada K. Weakly supervised silhouette-based semantic scene change detection[EB/OL]. (2018-11-29) [2020-06-30]. https://arxiv.org/abs/1811.11985v1.

[22] Xiao T T, Liu Y C, Zhou B L, et al. Unified perceptual parsing for scene understanding [C]// Proceedings of the 15th European Conference on Computer Vision. Munich, Gemany: Springer, 2018: 432-448.