

文章编号 1004-924X(2023)02-0234-12

## 并行路径与强注意力机制遥感图像建筑物分割

杨坚华<sup>1,2,3,4</sup>, 张 浩<sup>1,2,3,4</sup>, 花海洋<sup>1,2,3\*</sup>

(1. 中国科学院光电信息处理重点实验室, 辽宁 沈阳 110016;

2. 中国科学院沈阳自动化研究所, 辽宁 沈阳 110016;

3. 中国科学院机器人与智能制造创新研究院, 辽宁 沈阳 110169;

4. 中国科学院大学, 北京 100049)

**摘要:** 遥感图像建筑物分割广泛应用于城市规划及军事领域, 是当前遥感领域的研究热点。针对遥感图像中建筑物之间尺度变化较大、建筑物遮挡、建筑物阴影与建筑物边缘相似所导致建筑物分割精度较低的问题, 提出一种并行路径和强注意力机制的卷积神经网络模型。该模型基于 ResNet 网络残差连接的思想, 以 ResNet 为基础网络提高网络深度, 并采用卷积下采样得到并行路径, 提取建筑物的多尺度特征, 以减少建筑物之间尺度变化的影响。然后加入强注意力机制, 增强多尺度信息的融合效果, 增加不同特征之间的区分度, 抑制建筑物遮挡及建筑物阴影的影响。最后, 在多尺度融合特征后加入金字塔空间池化模块, 抑制分割结果中建筑物内部孔洞的出现, 提高分割精度。在 WHU 以及 Massachusetts Buildings 公开数据集进行实验, 分别从 MIoU, Recall, Precision, F1-score 4 个指标对分割结果进行量化比较, 在 Massachusetts Buildings 数据集中 MIoU 达到 72.84%, 相较于 ResUNet-a 提升 1.46%, 能够有效提高遥感影像中建筑的分割精度。

**关键词:** 遥感图像; 建筑物分割; 并行路径; 强注意力机制; 金字塔空间池化

中图分类号: P236 文献标识码: A doi: 10.37188/OPE.20233102.0234

## Parallel path and strong attention mechanism for building segmentation in remote sensing images

YANG Jianhua<sup>1,2,3,4</sup>, ZHANG Hao<sup>1,2,3,4</sup>, HUA Haiyang<sup>1,2,3\*</sup>

(1. Key Laboratory of Opto-Electronic Information Processing, Chinese Academy of Sciences, Shenyang 110016, China;

2. Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China;

3. Institutes for Robotics and Intelligent Manufacturing, Chinese Academy of Sciences, Shenyang 110169, China;

4. University of Chinese Academy of Sciences, Beijing 100049, China)

\* Corresponding author, E-mail: c3i11@sia.cn

**Abstract:** Building segmentation in remote sensing images is widely used in urban planning and military fields, and is a current focus of research in the remote sensing field. To solve the problems of large-scale changes between buildings, building occlusion, and similar building shadows and edges in remote sensing

收稿日期: 2022-03-01; 修订日期: 2022-03-20.

基金项目: 中科院创新基金资助项目 (No. E01Z040101)

images, which result in low building segmentation accuracy, a convolutional neural network with parallel paths and strong attention mechanism was developed. The model was based on the idea of residual connections of a ResNet network, and used ResNet as the basic network to improve the network depth and convolution downsampling to obtain parallel paths to extract multi-scale features of buildings to reduce the influence of scale changes between buildings. A strong attention mechanism was then added to enhance the fusion effect of the multi-scale information and discrimination of different features, and suppress the influence of building occlusion and shadows. Finally, a pyramid space pooling module was added after the multi-scale fusion features to suppress the appearance of holes inside the building in the segmentation result and improve the segmentation accuracy. Experiments were conducted on the WHU and Massachusetts Buildings public datasets, and the segmentation results were quantitatively compared using four indicators, namely MIOU, recall, precision, and F1-score. In the Massachusetts Buildings dataset, MIOU reaches 72.84%, which is 1.46% higher than the MIOU obtained with ResUNet-a. Thus, the model effectively improved the segmentation accuracy of buildings in remote sensing images.

**Key words:** remote sensing images; building segmentation; parallel path; strong attention mechanism; pyramid space pooling

## 1 引言

随着遥感卫星技术的不断发展,遥感数据随之受到了越来越多的关注,在许多领域都得到了广泛应用。例如在军事领域上被应用于地面目标的识别、在国土资源领域上被应用于环境变化等方面。其中城市区域是遥感数据中应用较为广泛的部分,建筑物作为城市区域的主要元素,对遥感影像中建筑的分割,有助于人口居住变化分析、目标建筑检测及城市规划等。由于建筑物与非建筑物之间的差异较小,且建筑物之间存在尺度大小变化较大、建筑物阴影与建筑物边缘难以区分、建筑物遮挡、建筑之间密集度较大等问题,因此如何提高建筑物的分割精度是一个重要难点。

当前提取遥感影像中建筑的方法可以大致分为两类:传统的图像处理方法和基于深度学习的方法。传统的图像处理方法提取遥感影像中建筑,往往是利用遥感图像本身的光谱、几何形状、纹理等<sup>[1-3]</sup>特征来组合,设计合理的特征值,通过阈值或机器学习的方法对建筑及背景进行分类,提取建筑。文献[4]通过滑动窗口进行计算地物的特征,提取后的特征图通过拟合分类算法进行预测每个窗口内的地物类别信息,进而从分类后的遥感图像中进行对建筑的提取。此方法提取建筑的效率较高,但精度方面仍有较大的改

善空间。文献[5]通过对光学图像、地理信息系统及数字地表模型3个数据源进行结合,以此来区分非建筑物和建筑物,多源数据大大提高了建筑物提取的鲁棒性,但是多源数据相较于单一来源的遥感图像,获取成本较高。

相较于传统提取遥感影像中建筑的方法,深度学习主要通过其网络模型自动去学习建筑与非建筑之间的深层次特征,对其进行分类提取建筑。MFDC-ResNet<sup>[6]</sup>以ResNet为基础,通过结合空间金字塔模型及空洞卷积,获取建筑不同尺度的特征并进行融合,其分割效果相较于ResNet表现更好。全局局部一体化D-LinkNet<sup>[7]</sup>通过改进条件随机场,增加了网络对全局局部细节的感知,避免了传统条件随机场丢失边界信息的情况,并且通过局部的边缘分割先验,增强建筑边缘提取的准确度。CU-Net<sup>[8]</sup>在FCN的基础上增加了多重约束来增强多尺度特征表示,增加多重约束后网络在建筑物分割方面比FCN表现更好。SiU-Net<sup>[9]</sup>针对多尺度输入设计了两个权重共享的分支,通过此方法可以有效提高分割精度,在大型建筑物分割方面尤为突出,但其计算复杂度将会大大增加。SRI-Net<sup>[10]</sup>基于ResNet101进行修改,提出空间残差模块生成多尺度特征并进行融合,网络通过融合高分辨率的特征,保留了较多的局部细节信息,能有效提高网络性能。EU-Net<sup>[11]</sup>提出了深度空间金字塔结

构,这个结构通过提取多尺度特征增加了对多尺度建筑的提取及边缘定位的精度。并且网络中采用了 focal loss 损失函数,有效提高了训练的稳定性。HRNetv2<sup>[12]</sup>采用多路径方式提取不同尺度的建筑特征,通过中间不断融合不同尺度信息,增加单一路径中的多尺度信息,提高了分割的准确性。

上述模型在遥感图像建筑物分割领域表现较好,但其在遥感影像中建筑之间尺度变化较大时,由于其对建筑物多尺度特征提取不充分,往往只对其中尺度居中的建筑物分割较为精确,对像素占比超过训练图像 1/5 以上的大型建筑分割不够完整、对像素占比小于训练图像的 1/100 的小型建筑边缘分割不够精细。且当建筑物被其他非建筑物遮挡时,无法准确分割出被建筑被遮挡的部分,而建筑物阴影也易被分割为建筑。针对上述问题,提出了一种基于并行多尺度加强注意力机制的语义分割网络提取目标建筑物的轮廓。提出的模型以 ResNet 为基础网络,采用多路径同时提取不同范围的建筑物特征,适应不同尺度建筑物的特征提取,以减少建筑物之间尺度变化较大的影响。在多路径提取不同尺度特征后,引入路径之间强关联性注意力机制,对弱特征进行增强,抑制建筑物被遮挡的影响,增强建筑物与建筑物阴影之间的区分度。并在融合后的单尺度特征中加入空间池化金字塔结构,减少提取建筑物内部空洞的产生,提升分割的性能。

## 2 并行路径与强注意力机制网络 PPA-Net

### 2.1 并行路径网络

当前获取高分辨率特征图的方法主要有两种:(1)编码器-解编码器网络:由网络(例如 ResNet)输出的低分辨率特征图上采样获得高分辨率特征图,或从可选的中间层中等分辨率特征图上采样并反卷积获得高分辨率特征图,如 Hourglass, SegNet<sup>[13]</sup>, DeconvNet<sup>[14]</sup>, U-Net<sup>[15]</sup>, 编码器-解码器(encoder-decoder)等。即通过对下采样后获取的特征图进行上采样来获得高分辨率特征图;(2)并行多路径网络:主路径对高分辨率

特征直接卷积不进行下采样保持高分辨率,并通过并行路径下采样获取低分辨率特征图,提取多尺度信息,例如 HRNetv2。

第一种网络由于对高分辨率特征存在下采样操作,导致后续恢复高分辨率过程中可能出现边缘定位不清晰,而第二种网络在主路径中并未对高分辨率特征图进行下采样,边缘定位相对较好,因此本文基于 HRNetv2 网络进行改进,采用并行路径网络得到多尺度信息,以减少下采样过程中细节信息的损失。

### 2.2 并行路径与强注意力机制网络

基于编码器-解编码器的网络在编码器阶段由于过度下采样操作损失了较多的空间细节信息,尽管解编码器阶段可能通过融合浅层特征图进行空间细节信息恢复,但融合的浅层特征具备的空间细节信息属于低级特征,与高级特征一同进行特征提取效果可能较差,最终导致建筑物边缘细节定位不够精确。考虑到建筑物分割对边缘定位要求较高,相较于编码器-解编码器类型的网络,多路径网络在保留较多空间细节信息的同时,能够提取多尺度信息,较为适合遥感影像中建筑物分割领域。

本文所提的并行路径与强注意力机制网络首先通过两个阶段逐步生成三个并行路径,以 ResNet 残差思想为基础组成卷积块提取网络特征,每个路径中的特征空间分辨率皆为定值,当路径分为三个后,在对各个路径卷积的同时,也进行互相融合,聚合各个路径中不同尺度的信息。然后引入强注意力机制,从多条路径中自适应的学习各个路径特征的重要程度,以提高多尺度特征融合的效果。随后进行金字塔空间池化操作<sup>[16]</sup>,增强特征的全局信息,减少建筑物内部孔洞的产生。

本文提出的 PPA-Net (Parallel Path Attention Net) 主要包括以下两部分:

(1)并行多路径网络,用于提取多尺度高级语义特征,同时保留空间细节信息,且在提取高级语义特征的同时路径之间互相融合,加强各个路径包含的信息,有助于分割尺度变化较大的遥感影像建筑物;

(2)在并行网络提取末端,加入路径之间强

相关的多尺度注意力机制,使网络更加注重于建筑物的特征及尺度信息,有助于建筑物尺度识别及抑制建筑物遮挡、阴影等影响。在注意力机制

后加入金字塔空间池化结构,抑制干扰特征,减少建筑物分割结果内部出现孔洞的现象。网络结构如图1所示。

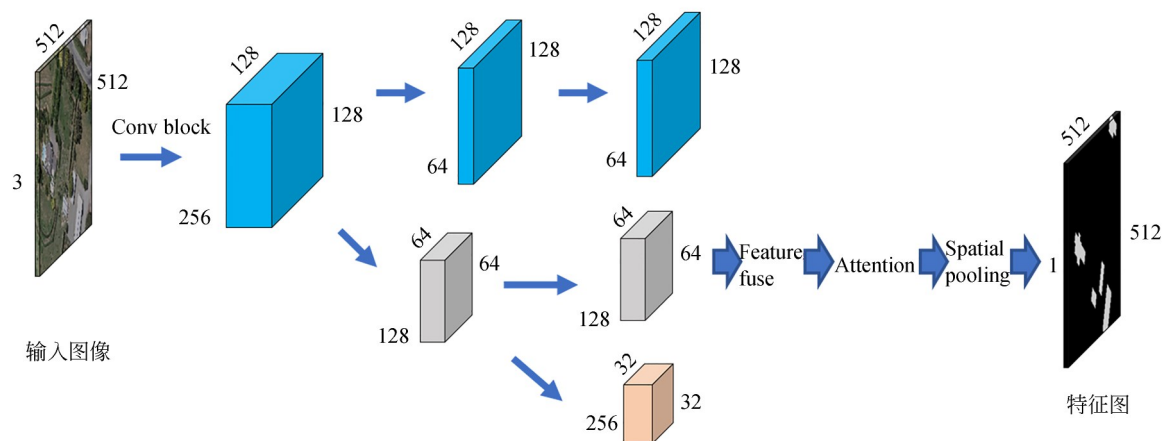


图1 PPA-Net网络结构图

Fig. 1 PPA-Net network structure

### 2.3 多路径及特征融合机制

多路径特征提取网络相较于基于编码器-解编码器的网络,由于路径的增多网络的参数量相较于同样深度的编码器-解编码器网络,会产生明显提升。因此所提模型仍采取了一定的最大值池化下采样,以减少网络的复杂度。由于模型只对原输入图像进行了1/4下采样,损失的定位信息较少,因此相较于编码器-解编码器网络对边缘定位精度影响较低。在多路径提取特征的同时,卷积只提取了单个尺度的信息,缺失了其他尺度的信息,因此在提取过程中加强各个路径之间的融合,能够有效增加对各个尺度信息的提取。所提模型的并行路径为3个,提取尺度分为1,1/2,1/4,经过并行路径之间的相互融合,提取的尺度更加细分,对建筑的识别效果也更好。其中多路径融合过程如图2所示。

### 2.4 强尺度注意力机制及金字塔空间池化

从不同路径提取的特征图具有不同的空间分辨率,通过注意力机制对不同路径特征聚合需要对低空间分辨率的特征图进行上采样恢复空间分辨率。如图3所示,低分辨率特征通过线性插值将特征上采样至原始图像的1/4,再通过 $3 \times 3$ 卷积将特征转化为64通道,并建立相邻路径之间的注意力机制,对相邻路径特征进行逐层

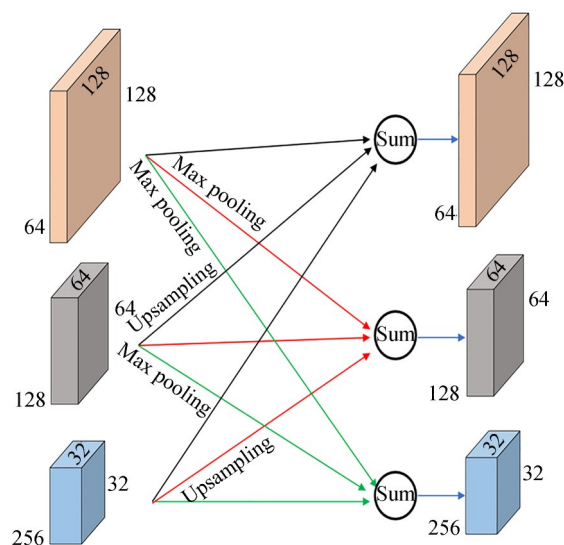


图2 多路径融合机制

Fig. 2 Multi-path fusion mechanism

融合。

从图3可以看出,注意力机制首先通过卷积学习得到最小尺度的空间注意力权重,再根据这个权重,对其进行处理,得到一个相关的权重,将其注入到相邻的并行路径中去,形成相邻并行尺度之间空间注意力的强相关。最终实现对目标建筑的尺度自适应识别分割,且能有效增强网络对阴影及建筑相似物的识别能力,减少建筑物遮



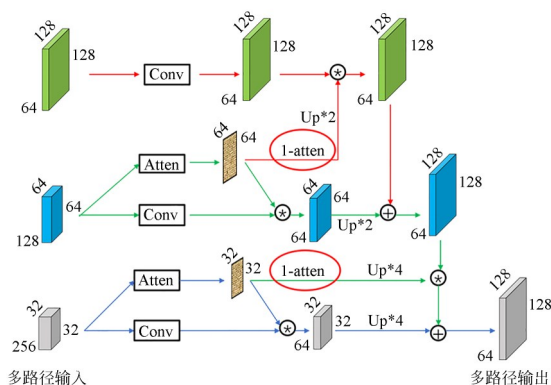


图3 强注意力机制

Fig. 3 Strong attention mechanism

挡的影响。本文所提强注意力机制基于传统空间注意力机制改进,相较于传统多路径空间注意力机制,强注意力机制结合并行路径中多尺度特征互补特性,路径之间所学注意力权重图相加和为1,以本文三个并行路径为例具体描述为:最小分辨率路径所学注意力权重图为A,中间路径所学注意力权重图为B,则最终中间路径权重图为 $(1-A) \times B$ ,最高分辨率路径注意力权重图可由另外并行路径权重图得到为 $(1-A) \times (1-B)$ 。并行路径之间注意力权重图相加和为1,可以更好地控制空间注意力的学习趋势。因为图像中每个像素的尺度是固定的,多尺度信息必然有所侧重而不存在每个尺度信息重要程度一样,而本文所提强注意力机制中某一尺度注意力权重越大,其相邻尺度注意力权重必然会越小,可以更好区分多尺度信息的重要程度,进而提高像素尺度信息识别。

其中Conv结构与atten结构相似,区别在于最后atten结构输出结果为单通道,Conv结果输出为多通道,这样可以尽可能使得两个结构提取出的特征属于同一水平深度特征,权重图可以更好地表示对应位置特征的重要性。其结构如图4所示。

为了获得完整性更好的建筑物分割结果,在网络末端加入金字塔空间池化模块,以增强特征的全局相关性。其中金字塔空间池化模块主要由4个具有不同大小的最大值池化层组成,这些最大值池化层根据建筑物尺寸进行设计,以提取建筑物的全局特征。通过池化层得到的特征图与原始特征图进行级联以实现全局空间增强。

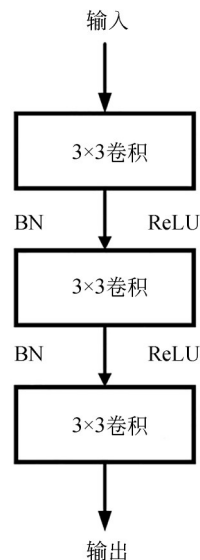


图4 Conv结构示意图

Fig. 4 Schematic diagram of Conv structure

在网络末端增加金字塔空间池化模块可以避免全局信息在网络提取过程中被抑制,能更好地提高建筑物的分割完整性,减少建筑物内部孔洞的产生。金字塔空间池化结构如图5所示。

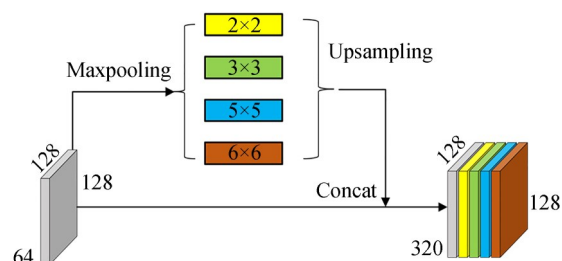


图5 金字塔空间池化模块

Fig. 5 Pyramid space pooling module

### 3 实验设置与结果

#### 3.1 实验数据及实验平台

本文研究是基于TensorFlow框架,在具有12 GB显存的单个2080Ti GPU上进行的。网络训练采用Adam优化器,设置初始学习率为0.001,并将beta1和beta2设置为默认值,batch\_size设置为4,迭代次数为80次。通过对数据集进行随机旋转及翻转操作进行数据集扩充。对比模型为ResUNet-a<sup>[15]</sup>, PSPNet<sup>[17]</sup>, ResNet101<sup>[18]</sup>, HRNetv2和SCAttNet<sup>[19]</sup>。

因为涉及到像素级二分类,所以采用 sigmoid 交叉熵函数作为损失函数,计算方式如公式(1)所示:

$$Loss = -[y \log \hat{y} + (1 - y) \log (1 - \hat{y})], \quad (1)$$

其中: $\hat{y}$ 代表当前像素的真实值, $y$ 代表当前像素的预测值, $Loss$ 为损失量。

实验数据集选择 WHU 建筑数据集<sup>[9]</sup>及 Massachusetts Buildings 数据集<sup>[20]</sup>,两个数据集都为已标注数据集。其中 WHU 数据集样本来自新西兰土地信息服务网站,数据集有约 22 000 个独立建筑,图像分辨率为 0.3 m。图像为 RGB 遥感影像,每个图像尺寸为 512×512 pix。数据集共有 8 188 个图像,其中训练数据集为 4 736 个图像,验证集为 1 036 个图像,测试集为 2 416 个图像。

Massachusetts Buildings 数据集由波士顿地区的 151 个遥感图像组成,每张遥感图像的像素为 1 500×1 500,覆盖面积为 2.25 km<sup>2</sup>。其中训练集包括 137 个图像,测试集包括 10 个图像,验证集为 4 个图像。相较于 WHU 数据集,Massachusetts Buildings 数据集阴影干扰及建筑物密集程度较高,Massachusetts Buildings 中每个 512×512 输入图像中建筑物个数普遍存在 200 个左右,WHU 数据集中则大多只存在 40 个左右,且 Massachusetts Buildings 数据集中多数建筑存在阴影干扰,且阴影面积占建筑面积接近 1/3 影响较大,WHU 数据集中阴影干扰较少,难以判断网络在阴影干扰条件下的分割效果。因此采用 Massachusetts Buildings 数据集可以验证本文算法在不同场景下的性能。WHU 数据集与 Massachusetts Buildings 数据集部分区域对比如图 6 所示。

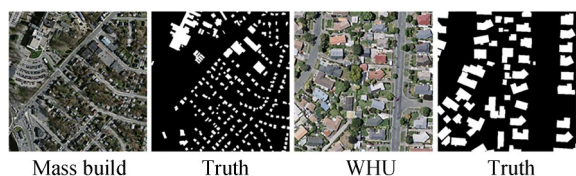


图6 WHU数据集与Massachusetts Buildings数据集对比  
Fig. 6 Comparison of WHU dataset and Massachusetts Buildings dataset

### 3.2 分割性能评价指标

本文采用精确率(precision),召回率(recall),F1-score和平均交并比(MIoU)像素级度量指标来评估PPA-Net和其他不同方法的性能,具体公式如下:

$$precision = \frac{TP}{TP + FP}, \quad (2)$$

$$recall = \frac{TP}{TP + FN}, \quad (3)$$

$$F1 = \frac{2 \times precision \times recall}{precision + recall}, \quad (4)$$

$$IoU = \frac{TP}{TP \cup FN \cup FP}, \quad (5)$$

其中: $TP$ 指对正样本的正确预测, $FN$ 指对正样本的错误预测, $TN$ 指对负样本的正确预测, $FP$ 指对负样本的错误预测。 $precision$ 表示预测为正样本的数据中预测正确的样本数量比例, $recall$ 表示总正样本中预测正确的样本数量比例, $F1-score$ 是 $precision$ 和 $recall$ 的调和平均值,可以更好的反映模型的分割性能, $IoU$ 表示预测为正样本的数据及标签中正样本数据的交集除以两个数据集的并集, $MIoU$ 为总数据集的平均 $IoU$ 。

### 3.3 模型对比试验结果分析

在 WHU 数据集上,本文模型与其他对比模型的分割结果对比如图 7 所示。

从图 7 可以看出,第一个场景中,在建筑物被附近植被部分遮挡的情况下,本文算法相较于其他算法,对被遮挡部分的建筑物分割效果较好,能够有效抑制建筑物遮挡的影响。第二个场景中,本文算法相较于其他算法,在小型建筑之间的间隙识别中具有更好的性能,其他算法则将相邻建筑之间的背景识别为了建筑。第三个场景中,本文算法精确识别出了边缘的小型建筑,而其他算法则识别不全或识别不出。第四个场景中,本文算法对大型建筑边缘分割较为精确,且大型建筑内部不存在孔洞。从图 7 整体的结果可以看出,HRNetv2 由于在并行多路径提取特征后只是简单的将多尺度特征级联到一起,没有很好地结合多尺度特征,导致分割结果并不好。ResUNet-a 模型以 U-Net 为基础架构,基于残差连接的思想提出 Resunet-a 模块增加网络深度及多尺度信息,但其仍将编码器对应阶段的特征复制

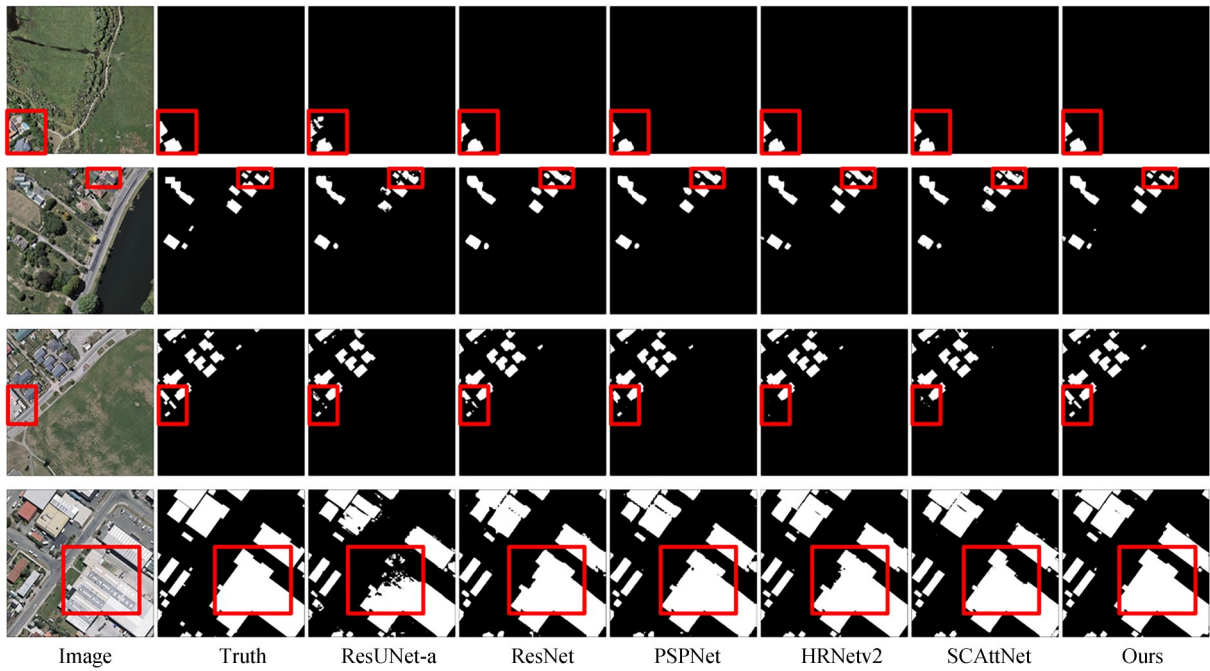


图 7 在 WHU 数据集分割结果对比

Fig. 7 Comparison of segmentation results in WHU datasets

融合到解码器中相应阶段中,引入了浅层特征中的大量噪声,最终导致分割效果较差。ResNet 模型网络层数较深,能够提取到深层的语义特征,但由于下采样操作导致信息的缺失,最终导致分割边缘精度的下降。PSPNet 模型基于 ResNet50,引入金字塔结构,但只是简单地提取了多尺度的信息,对多尺度信息融合较为简单,导致分割结果并不好。SCAttNet 模型以 SegNet 为基础网络,在网络上采样恢复分辨率前加入空间注意力及通道注意力模块以提高分割精

度,但其模型仍存在下采样过程中损失过多细节信息的问题。相较于这些算法,PPA-Net 在提取多尺度特征之后,采用强相关多尺度注意力机制,很好地融合了多尺度信息,在这之后采用金字塔空间池化结构,增强网络分割建筑边缘的精确度和孔洞的减少,在建筑物被遮挡及建筑物之间尺度变化较大的情况下,分割效果较好。

表 1 给出了各个模型对 WHU 数据集分割的指标对比,其中参数量及计算量以百万为单位 (Million,M)。

表 1 在 WHU 数据集分割指标对比

Tab. 1 Comparison of segmentation indicators in WHU dataset

| 模型性能对比    | MIoU   | Precision | Recall | F1-score | 参数量/M | 计算量/M  |
|-----------|--------|-----------|--------|----------|-------|--------|
| ResUNet-a | 84.08% | 92.63%    | 90.11% | 91.35%   | 43.16 | 86.36  |
| PSPNet    | 89.1%  | 94.5%     | 93.97% | 94.24%   | 46.72 | 93.48  |
| ResNet101 | 88.41% | 93.98%    | 93.71% | 93.85%   | 52.11 | 104.31 |
| HRNetv2   | 90.27% | 95.23%    | 94.53% | 94.88%   | 29.54 | 59.20  |
| SCAttNet  | 88.1%  | 94.63%    | 92.73% | 93.67%   | 29.46 | 39.65  |
| Ours      | 90.45% | 94.9%     | 95.07% | 94.98%   | 26.43 | 52.95  |

从表 1 中不同算法再 WHU 数据集上所,分割后的结果的定量分析可以看出,本文所提模型 PPA-Net 在 MIoU,Recall 及 F1-score 表现皆为最

优,而在指标 Precision 上并不是最优,表明本文所提模型在整体分割性能上是要优于其他算法,但在预测结果为建筑部分的准确率仍有待提高,



仍需降低其误分类概率。从参数量及计算量可以看出,由于本文所提模型 PPA-Net 是基于 HRNetv2 进行了改进的,对比 HRNetv2 本文所提模型参数量及计算量皆有所下降,且分割性能

得到了一定提高。

在 Massachusetts Buildings 数据集上,本文模型与其他对比模型的分割结果如图 8 所示。

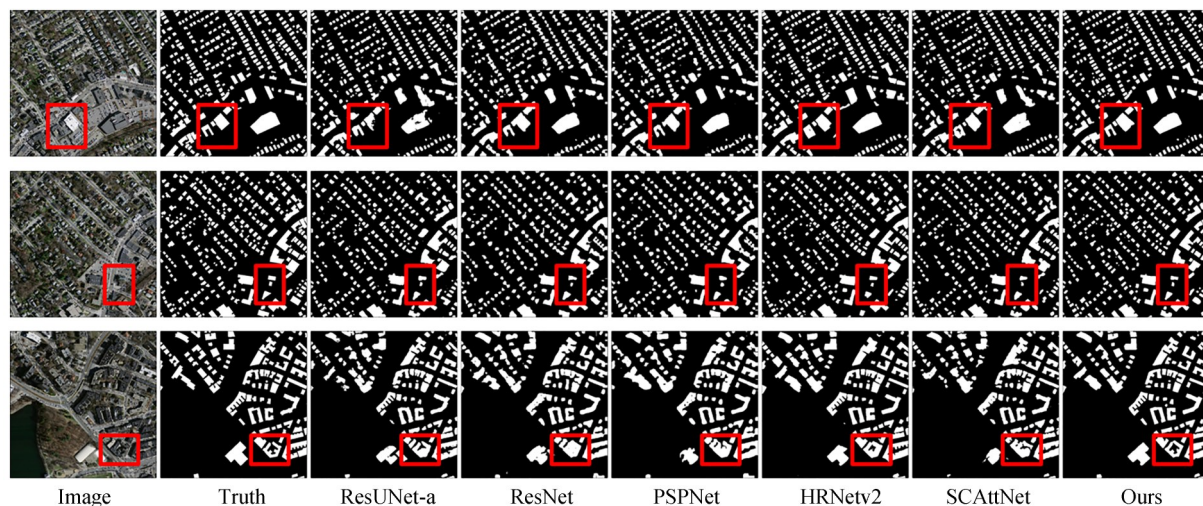


图8 在 Massachusetts Buildings 数据集分割结果对比

Fig. 8 Comparison of segmentation results in Massachusetts Buildings dataset

从图8中,第一个场景的分割结果可以看出,ResUNet-a对大型建筑分割非常差,缺失了一大部分建筑,ResNet101和PSPNet则未识别出两个建筑物之间的间隙,HRNetv2和SCAttNet则对叉形建筑存在误分割现象,相对而言本文所提模型 PPA-Net 在两个建筑之间间隔较小时,可以准确分割出其间隙,且未对叉形建筑内部非建筑物产生误分割现象。从第二个场景的分割结果可以看出,在方框区域内,PPA-Net对小型建筑及大型建筑分割效果较好,对其边缘定位较为精确,而HRNetv2未分割出一部分三角形建筑,其他对比模型则对小型建筑分割效果较差,边缘定位不够准确。从第三个场景的分割结果可以看出,方框区域内的建筑物属于圈型建筑,其内部存在阴影的干扰,对比的模型对建筑内部分割较差,存在严重的误分割现象,而本文所提 PPA-Net 在阴影干扰下对建筑内部分割较好。从数据集 Massachusetts Buildings 的分割结果来看,本文算法对阴影抗干扰性较强,且能在密集型建筑物中准确分割出其边缘。

表2给出了各个模型对 Massachusetts Buildings 数据集分割指标的对比。

表2 在 Massachusetts Buildings 数据集分割指标对比

Tab. 2 Comparison of segmentation indicators in Massachusetts Buildings dataset

| 模型性能对比    | MIoU   | Precision | Recall | F1-score |
|-----------|--------|-----------|--------|----------|
| ResUNet-a | 71.38% | 86.23%    | 80.56% | 83.3%    |
| PSPNet    | 67.74% | 82.26%    | 79.32% | 80.76%   |
| ResNet101 | 67.46% | 81.93%    | 79.25% | 80.57%   |
| HRNetv2   | 70.28% | 84.69%    | 80.50% | 82.54%   |
| SCAttNet  | 70.36% | 85.57%    | 79.83% | 82.61%   |
| Ours      | 72.84% | 86.81%    | 81.91% | 84.29%   |

从表2指标对比可以看出,本文模型相较于对比模型在各个指标上都有较大提升,MIoU指标相较于 ResUNet-a 提升 1.46%,分割性能更佳,通过两个数据集的指标提升对比,可以发现 PPA-Net 在建筑物密集且阴影干扰较多的数据集上表现更好。

### 3.4 与其他现有模型对比

本部分将所提模型与目前一些较新的建筑物分割模型进行对比,由于相关论文代码并未开源,因此在 WHU 数据集上的性能指标直接采用



相应文章的结果,对比如表3所示。SRI-Net<sup>[10]</sup>采用修改后的 ResNet-101 编码器生成多级特征,以空洞卷积拓宽感受野,并提出空间残差初始(SRI)模块以融合多尺度信息。DE-Net<sup>[21]</sup>以编码解编码为主要架构,在下采样过程中采用最大池化层及跨部卷积级联,并用密集型上采样卷积获得建筑分割结果。DS-Net<sup>[22]</sup>采用深层次监督子网络对模型的多尺度输出进行监督学习,并采用多尺度注意力模块对多尺度信息进行融合输

出分割结果。AGEDNet<sup>[23]</sup>以修改后的 ResNet-50 为基础网络,加入空间注意力及空间金字塔池化模块以提高建筑物分割精度。RBUNet[24]在编码部分加入了自注意力模块以提高模型对显著区域的关注度,在解编码部分采用多个大型卷积进行上采样以加强模型解编码能力。由表3中数据可以看出,本文所提模型在 MIoU, Recall 及 F1-score 指标中均优于所比较的建筑物分割模型。

表 3 与现有模型在 WHU 数据集上对比  
Tab. 3 Comparison with existing models on the WHU dataset

| 模型性能对比  | MIoU          | Precision     | Recall        | F1-score      |
|---------|---------------|---------------|---------------|---------------|
| SRI-Net | 89.23%        | 95.67%        | 93.69%        | 94.51%        |
| DE-Net  | (90.12±0.24)% | (95.00±0.16)% | (94.60±0.19)% | (94.80±0.18)% |
| DS-Net  | 90.4%         | 94.85%        | 95.06%        | 94.96%        |
| AGEDNet | 90.29%        | 94.97%        | 94.81%        | 94.9%         |
| RBUNet  | 89.39%        | 95.56%        | 93.25%        | 94.4%         |
| Ours    | 90.45%        | 94.9%         | 95.07%        | 94.98%        |

3.5 模型消融实验分析

为了分析强注意力机制和金字塔空间池化模块对分割精度的影响,本文对整体模型在 WHU 及 Massachusetts Buildings 数据集上进行消融学习实验。首先,本文在没有添加任何模块的基础模型(baseline)上进行实验,

然后在基础模型上增加强注意力机制模块(baseline+A)进行实验,最后在基础模型上增加金字塔空间池化模块(baseline+P)进行实验。在 WHU 及 mass build 数据集上的分割指标对比如表4所示,部分区域分割结果对比如图9所示。

表 4 消融实验指标对比  
Tab. 4 Comparison of ablation experimental indexes

| 数据集                     | 模型对比       | MIOU   | Precision | Recall | F1-score | 参数量   | 计算量   |
|-------------------------|------------|--------|-----------|--------|----------|-------|-------|
| WHU                     | baseline   | 89.49% | 94.52%    | 94.38% | 94.45%   | 23.35 | 46.78 |
|                         | Baseline+A | 90.30% | 94.48%    | 95.33% | 94.90%   | 26.39 | 52.88 |
|                         | Baseline+P | 89.86% | 95.76%    | 93.58% | 94.66%   | 23.37 | 46.82 |
|                         | PPA-Net    | 90.45% | 94.9%     | 95.07% | 94.98%   | 26.43 | 52.95 |
| Massachusetts Buildings | baseline   | 71.52% | 86.34%    | 80.64% | 83.39%   |       |       |
|                         | Baseline+A | 72.62% | 87.03%    | 81.43% | 84.14%   |       |       |
|                         | Baseline+P | 72.01% | 86.48%    | 81.15% | 83.73%   |       |       |
|                         | PPA-Net    | 72.84% | 86.81%    | 81.91% | 84.29%   |       |       |

由表4的数据对比可以看出,baseline加入强注意力机制模块及金字塔空间池化模块能够有

效提高建筑分割的准确度,其中金字塔空间池化模块的增加对模型的计算量及参数量增加较少,

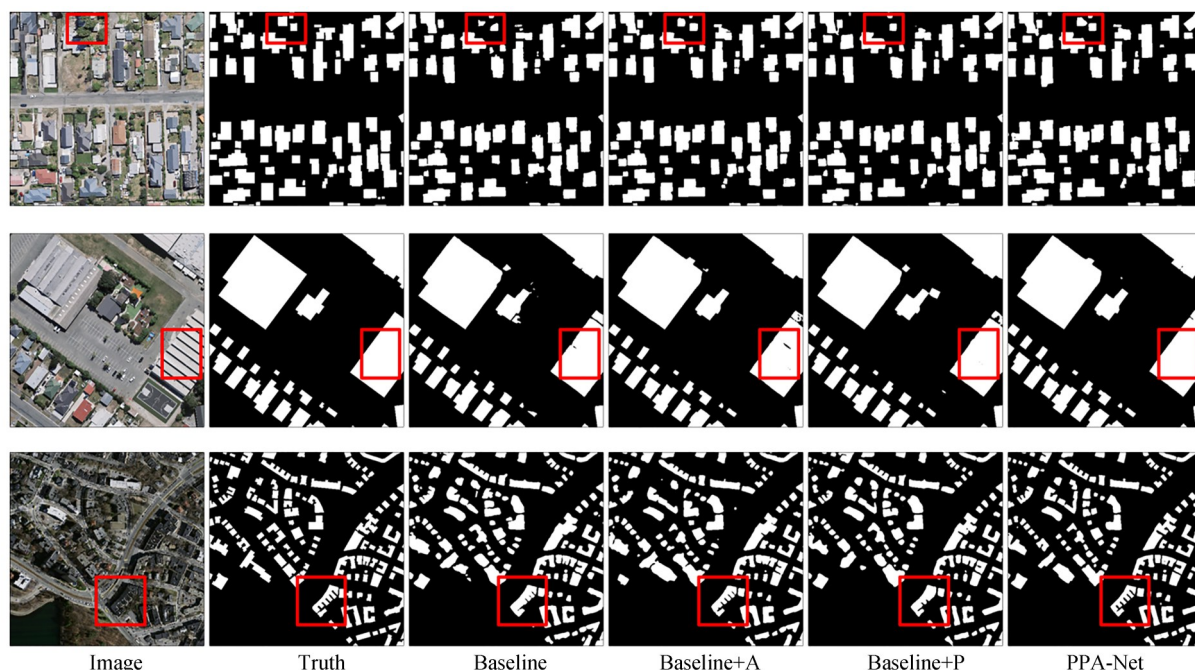


图9 消融实验分割结果对比

Fig. 9 Comparison of ablation experiment segmentation results

但其分割性能上提升也不多,强注意力机制模块对模型的计算量及参数量增加较多,模型的分割性能上提升也较为明显。总的来说,金字塔空间池化模块及强注意力机制模块能够有效提高模型分割精度,且两个模块存在一定的互补关系,整体网络相较于增加单个模块分割性能更优。

由图9的分割结果对比可以看出,场景一方框区域内,当建筑被植被遮挡较多时,增加了强注意力机制模块的模型相较于未增加强注意力机制模块的模型能够识别出更多被遮挡的建筑,分割效果更好。在场景二的方框区域内,当模型未增加金字塔空间池化模块时,对大型建筑分割内部存在孔洞现象,增加金字塔空间池化模块后能够有效抑制孔洞现象的产生。在场景三的区域,建筑存在被阴影干扰的现象,增加了强注意力机制模块的模型在建筑与阴影相交边界的分割效果更好,能够有效增加对阴影的抗干扰作用。由消融实验对比结果可以看出,强注意力机制模块能够有效增加模型在建筑遮挡、阴影干扰现象中的分割精度,金字塔空间池化模块能够有效抑制

模型对大型建筑分割时内部孔洞的出现。

## 4 结 论

对遥感影像中的建筑物提取,本文提出一个并行多路径网络,在多路径提取特征的过程中对并行路径特征进行融合,加强各个路径之间的多尺度信息;在多路径提取的最后阶段,加入强相关注意力机制,融合各个路径的特征,增强特征之间信息的互补,对特征进行有效加强,提升网络在建筑物被遮挡及建筑物阴影干扰下的分割性能。在最后阶段加入金字塔空间池化结构,减少建筑内部孔洞的产生,并提高建筑分割的精确度。实验证明,本文所提的模型在遥感影像中对建筑分割效果较好,相较于对比的算法,在WHU数据集及Massachusetts Buildings数据集上IoU指标,F1-score,Recall指标都是最好的,但在WHU数据集上Precision不是最优的,说明预测中将背景误认为建筑较高,且边缘并不是很精确,是下一步改进的方向,后续也可将此网络改进推广到遥感图像其他元素的分割中。

## 参考文献:

- [1] DU J L, CHEN D, WANG R S, *et al.* A novel framework for 2.5-D building contouring from large-scale residential scenes[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2019, 57(6): 4121-4145.
- [2] 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.
- [3] GAVANKAR N L, GHOSH S K. Automatic building footprint extraction from high-resolution satellite image using mathematical morphology[J]. *European Journal of Remote Sensing*, 2018, 51(1): 182-193.
- [4] 孙伟, 孙鹏翔, 黄恒, 等. 基于滑动窗口的影像中建筑物特征提取方法研究[J]. *传感技术学报*, 2021, 34(8): 1096-1101.  
SUN W, SUN P X, HUANG H, *et al.* Research on building feature extraction method in image based on sliding window[J]. *Chinese Journal of Sensors and Actuators*, 2021, 34(8): 1096-1101. (in Chinese)
- [5] SOHN G, DOWMAN I. Data fusion of high-resolution satellite imagery and LiDAR data for automatic building extraction[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2007, 62(1): 43-63.
- [6] 徐胜军, 欧阳朴衍, 郭学源, 等. 多尺度特征融合空洞卷积 ResNet 遥感图像建筑物分割[J]. *光学精密工程*, 2020, 28(7): 1588-1599.  
XU SH J, OUYANG P Y, GUO X Y, *et al.* Building segmentation in remote sensing image based on multiscale-feature fusion dilated convolution resnet[J]. *Opt. Precision Eng.*, 2020, 28(7): 1588-1599. (in Chinese)
- [7] 朱祺琪, 李真, 张亚男, 等. 全局局部细节感知条件随机场的高分辨率遥感影像建筑物提取[J]. *遥感学报*, 2021, 25(7): 1422-1433.  
ZHU Q Q, LI ZH, ZHANG Y N, *et al.* Global-Local-Aware conditional random fields based building extraction for high spatial resolution remote sensing images[J]. *National Remote Sensing Bulletin*, 2021, 25(7): 1422-1433. (in Chinese)
- [8] WU G M, SHAO X W, GUO Z L, *et al.* Automatic building segmentation of aerial imagery using multi-constraint fully convolutional networks[J]. *Remote Sensing*, 2018, 10(3): 407.
- [9] 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.
- [10] LIU P H, LIU X P, LIU M X, *et al.* Building footprint extraction from high-resolution images via spatial residual inception convolutional neural network[J]. *Remote Sensing*, 2019, 11(7): 830.
- [11] KANG W C, XIANG Y M, WANG F, *et al.* EU-net: an efficient fully convolutional network for building extraction from optical remote sensing images[J]. *Remote Sensing*, 2019, 11(23): 2813.
- [12] SUN K, XIAO B, LIU D, *et al.* Deep high-resolution representation learning for human pose estimation[C]. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 15-20, 2019, Long Beach, CA, USA. IEEE, 2020: 5686-5696.
- [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] NOH H, HONG S, HAN B. Learning deconvolution network for semantic segmentation[C]. 2015 *IEEE International Conference on Computer Vision (ICCV)*. December 7-13, 2015, Santiago, Chile. IEEE, 2016: 1520-1528.
- [15] 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.
- [16] 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.
- [17] 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. IEEE, 2017: 6230-6239.
- [18] 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. IEEE, 2016: 770-778.
- [19] LI H F, QIU K J, CHEN L, *et al.* SCAttNet: semantic segmentation network with spatial and channel attention mechanism for high-resolution remote sensing images [J]. *IEEE Geoscience and Remote Sensing Letters*, 2021, 18(5): 905-909.
- [20] MNH V. *Machine Learning for Aerial Image Labeling* [M]. Canada: University of Toronto, 2013.
- [21] LIU, LUO, HUANG, *et al.* DE-net: deep encoding network for building extraction from high-resolution remote sensing imagery [J]. *Remote Sensing*, 2019, 11(20): 2380.
- [22] GUO H N, SU X, TANG S K, *et al.* Scale-robust deep-supervision network for mapping building footprints from high-resolution remote sensing images [J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2021, 14: 10091-10100.
- [23] DENG W J, SHI Q, LI J. Attention-gate-based encoder-decoder network for automatic building extraction [J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2021, 14: 2611-2620.
- [24] CHEN Z Y, LI D L, FAN W T, *et al.* Self-attention in reconstruction bias U-net for semantic segmentation of building rooftops in optical remote sensing images [J]. *Remote Sensing*, 2021, 13(13): 2524.

## 作者简介:



杨坚华(1998—),男,浙江台州人,硕士研究生,2019年于郑州大学获得学士学位,现为中国科学院沈阳自动化研究所硕士研究生,从事景象匹配地面任务规划研究。E-mail: yjh993879413@163.com

## 通讯作者:



花海洋(1978—),男,项目研究员,硕士,2001年于东北大学获学士学位;2006年于中国科学院沈阳自动化研究所获硕士学位,现为中国科学院沈阳自动化研究所项目研究员,主要从事光电系统性能评估理论与方法、光电仿真、目标光学特性分析与建模等领域的研究。E-mail: c3i11@sia.cn