基于监督对比学习正则化的高分辨率SAR图像建筑物提取方法

康 健[①] 王智睿[*②⑤] 祝若鑫[③] 孙 显[②④⑤]

[①](苏州大学电子信息学院 苏州 215006)

[②](中国科学院空天信息创新研究院 北京 100190)

[③](西安测绘研究所地理信息工程国家重点实验室 西安 710054)

[④](中国科学院大学电子电气与通信工程学院 北京 100190)

[⑤](中国科学院网络信息体系技术科技创新重点实验室 100190)

摘 要：近年来，高分辨合成孔径雷达(SAR)图像的智能解译技术在城市规划、变化监测等方面得到了广泛应用。不同于光学图像，SAR图像的获取方式、图像中目标的几何结构等因素制约了现有深度学习方法对SAR图像地物目标的解译效果。该文针对高分辨SAR图像城市区域建筑物提取，提出了基于监督对比学习的正则化方法，其主要思想是增强同一类别像素在特征空间中的相似性以及不同类别像素之间的差异性，使得深度学习模型能更加关注SAR图像中建筑物与非建筑物区域在特征空间中的区别，从而提升建筑物识别精度。利用公开的大场景SpaceNet6数据集，通过对比实验，提出的正则化方法，其建筑物提取精度相比于常用的分割方法在不同网络结构下至少提升1%，分割结果证明了该文方法在实际数据上的有效性，可以对复杂场景下的城市建筑物区域进行有效分割。此外，该方法也可以拓展应用于其他SAR图像像素级别的地物分割任务中。

关键词：合成孔径雷达；SAR建筑物提取；深度学习；语义分割；对比学习

中图分类号：TP753 **文献标识码：**A **文章编号：**2095-283X(2022)01-0157-11

DOI: 10.12000/JR21124

引用格式：康健, 王智睿, 祝若鑫, 等. 基于监督对比学习正则化的高分辨率SAR图像建筑物提取方法[J]. 雷达学报, 2022, 11(1): 157–167. doi: 10.12000/JR21124.

Reference format: KANG Jian, WANG Zhirui, ZHU Ruoxin, *et al*. Supervised contrastive learning regularized high-resolution synthetic aperture radar building footprint generation[J]. *Journal of Radars*, 2022, 11(1): 157–167. doi: 10.12000/JR21124.

Supervised Contrastive Learning Regularized High-resolution Synthetic Aperture Radar Building Footprint Generation

KANG Jian[①] WANG Zhirui[*②⑤] ZHU Ruoxin[③] SUN Xian[②④⑤]

[①](*School of Electronic and Information Engineering*, *Soochow University*, *Suzhou* 215006, *China*)

[②](*Aerospace Information Research Institute*, *Chinese Academy of Sciences*, *Beijing* 100190, *China*)

[③](*State Key Laboratory of Geo-Information Engineering*, *Xi'an Research Institute of Surveying and Mapping*, *Xi'an* 710054, *China*)

[④](*School of Electronic*, *Electrical and Communication Engineering*, *University of Chinese Academy of Sciences*, *Beijing* 100190, *China*)

[⑤](*Key Laboratory of Network Information System Technology* (*NIST*), *Aerospace Information Research Institute*, *Chinese Academy of Sciences*, *Beijing* 100190, *China*)

收稿日期：2021-09-07；改回日期：2021-11-12；网络出版：2021-12-03

*通信作者： 王智睿 zhirui1990@126.com *Corresponding Author: WANG Zhirui, zhirui1990@126.com

基金项目：国家自然科学基金(62101371, 62076241)，江苏省青年基金项目(BK20210707)

Foundation Items: The National Natural Science Foundation of China (62101371, 62076241), Jiangsu Province Science Foundation for Youths (BK20210707)

责任主编：刘畅 Corresponding Editor: LIU Chang

Abstract: Over the recent years, high-resolution Synthetic-Aperture Radar (SAR) images have been widely applied for intelligent interpretation of urban mapping, change detection, etc. Different from optical images, the acquisition approach and object geometry of SAR images have limited the interpretation performances of the existing deep-learning methods. This paper proposes a novel building footprint generation method for high-resolution SAR images. This method is based on supervised contrastive learning regularization, which aims to increase the similarities between intra-class pixels and diversities of interclass pixels. This increase will make the deep learning models focus on distinguishing building and nonbuilding pixels in latent space, and improve the classification accuracy. Based on public SpaceNet6 data, the proposed method can improve the segmentation performance by 1% compared to the other state-of-the-art methods. This improvement validates the effectiveness of the proposed method on real data. This method can be used for building segmentation in urban areas with complex scene background. Moreover, the proposed method can be extended for other types of land-cover segmentation using SAR images.

Key words: Synthetic Aperture Radar (SAR); SAR building footprint generation; Deep learning; Semantic segmentation; Contrastive learning

1 引言

合成孔径雷达(Synthetic Aperture Radar, SAR)是一种具备全天时、全天候观测能力的主动式微波成像雷达，在军事和民用对地观测领域中具有广阔的应用前景。随着SAR技术的快速发展，SAR图像在空间分辨率及质量上得到了显著提升，这进一步提升了高分辨率SAR图像解译技术在城市规划、城市变化监测等方向的重要应用价值[1–4]。

不同于光学图像，SAR图像的成像机理使得被观测地物目标具有独特的几何特性，比如透视收缩、叠掩等，而且SAR图像主要反映地物目标对微波的后向散射特性，并不能充分显示出目标的纹理结构及颜色特征，这些因素使得SAR图像解译一直面临较大的挑战。随着对地观测技术对大范围、精细化SAR图像的地物目标识别精度的要求不断提升，大场景、高精度、智能化的SAR图像解译技术是领域内近年来重要的研究方向。

城市地区的建筑物区域自动提取属于SAR图像解译技术的主要任务之一[5]。SAR图像建筑物提取旨在从获取到的SAR图像中分离出建筑物区域与背景区域。在城市地区，建筑物高度参差不齐且密集排布，存在相互遮挡的现象，而且背景目标丰富，电磁散射情况复杂，这些原因均影响了高分辨率SAR图像中建筑物区域的提取精度。

经典的SAR图像中提取建筑物区域的方法主要是基于人为手工设计的建筑物区域特征，如线段与点特征。通过利用回波在建筑物与地面之间的二次散射(double bounce)现象，Tupin等人[6]设计了线段检测器用以识别出SAR图像中建筑物区域的轮廓。Xu等人[7]采用恒虚警率(Constant False Alarm Rate, CFAR)线段检测及霍夫(Hough)变换等方法对建筑物区域两端的平行线段进行检测。Michaelsen等人[8]提出了基于感知分组的方法提取城市地区中建筑物的细节结构。通过提取SAR图像的一系列底层特征，Ferro等人[9]将建筑物区域中不同类别的散射进行分类，从而实现了在单张SAR图像中对建筑物区域的自动检测。其他特征如矩形、L型结构也被应用于提取SAR图像中特定形状的建筑物[10,11]。虽然传统的方法已经应用于城市地区SAR图像建筑物提取，但人工设计的特征较适用于具有规则形状的建筑物区域建模，并不能对复杂形状的建筑物进行有效提取，而且上述方法通常应用于小范围场景图像中的建筑物区域识别，很难应用于城市级别的建筑物提取任务中。

近年来，数据驱动下的深度学习技术已经成为图像处理及视觉领域的主流方法。在海量训练数据的前提下，多层的卷积神经网络(Convolutional Neural Network, CNN)可以自适应地调节各层卷积核权重，使其能准确地挖掘目标从底层到高层的语义特征[12]。鉴于其特征提取的优越性能，CNN方法在SAR图像中的地物目标提取中获得了广泛关注。Wang等人[13]运用开放地图(Open Street Map)作为真值训练轻量级建筑物分割模型，并在高分辨率SAR图像城市地区建筑物提取中取得了良好的效果。杜康宁等人[14]结合时间序列图像的特点，利用多层神经网络的方法，提出了一种基于时间序列的建筑物区域提取方法。Shermeyer等人[15]提供了包括高分辨率SAR图像在内的鹿特丹城市地区建筑物提取的多源遥感图像数据集。通过融合层析SAR(TomoSAR)得到的建筑物点云以及OpenStreetMap信息，Shahzad等人[16]研究了大尺度SAR图像建筑物提取基准数据并利用全卷积神经网络(Fully

Convolutional Neural Network, FCNN)实现了建筑物分割模型。Jing等人[17]提出了有选择性的空间金字塔膨胀神经网络模型以及L型权重损失函数用来对SAR图像建筑物区域进行识别。Chen等人[18]利用多张SAR图像的干涉信息以及提出的复卷积神经网络模型对建筑物区域进行了有效分割。通过利用数字高程模型(Digital Elevation Model, DEM)以及地理信息系统(Geographic Information System, GIS)产生的建筑物区域真值，Sun等人[19]提出了多尺度特征融合的SAR图像建筑物提取方法。

现有的基于深度学习的SAR图像建筑物提取方法大部分利用交叉熵(Cross Entropy, CE)或者Dice系数作为损失函数训练CNN模型，这些基于分类或分割效果设计的损失函数并没有充分利用建筑与背景像素在特征空间中的语义关系，这使得训练模型对于复杂城市地区散射点的辨识能力不强，从而制约了模型对于大范围建筑物提取上的效果以及泛化能力。针对上述问题，本文设计了基于监督对比学习正则化的方法，在分割损失函数的基础上，利用同一类别像素在特征空间的距离近、不同类别像素在特征空间的距离远的性质，进一步在网络训练过程中约束不同像素之间的语义相似性，从而提升CNN模型对于建筑和背景像素的分辨能力，模型对于SAR图像建筑物的提取精度得到有效提高。

2　监督对比正则化的SAR图像建筑物提取

所提出的对比正则化的SAR图像建筑物提取方法主要包括：(1)图像分割网络模块；(2)像素级特征提取模块。与一般的图像分割所用网络一致，图像分割网络模块主要采取常用的分割网络，如DeepLabV3+[20]等，用来提取建筑物区域的预测结果，并利用分割损失函数进行优化学习，像素级特征提取模块主要用来学习一个特征空间，使得建筑物与非建筑物区域像素在其中更好地进行分离。图1展示了所提出方法的主要模型结构。

2.1　对比学习

对比学习的核心思想是在特征空间中缩小同一类别特征之间的距离，增加不同类别特征之间的距离。这一朴素思想近年来广泛应用于图像特征的自监督学习以及模型参数的预训练等方向[21,22]。基于给定的标签信息，Khosla等人[23]运用监督对比学习方法在ImageNet数据集上取得了比传统基于交叉熵损失函数训练模型更好的效果。根据这一结果，Liu等人[24]将图像级别的监督对比学习机制拓展到像素级别，用来增强不同类别像素在特征空间中的可分辨性。该方法主要在分割输出层之前引入卷积层，得到非线性特征投影，再利用图像的真值信息对特征投影进行监督对比学习，即约束类内及类间的特征投影距离，使得同一类别的特征投影在特征空间中距离近，不同类别的特征投影之间的距离远。在多类别场景图像语义分割中，该方法取得了良好的效果。受到这一思想的启发，本文将像素级别的监督对比学习方法引入到SAR图像的建筑物提取中。如图2所示，与光学遥感图像相比，SAR图像的建筑物区域的特征不明显，且受到相干斑噪声的影响，难以将其与周围地物进行区分。现有基于深度学习的SAR图像建筑物分割方法大多采用常用的CE或者Dice损失函数，这些损失函数均直接作用于类别的分割图，没有对不同类别像素对应的深度特征进行约束，对于散射机制较为相近的地物目标，如楼房与街道，上述损失函数并不能取得很好的分类效果，因此，本文拟采用监督对比学习方法，

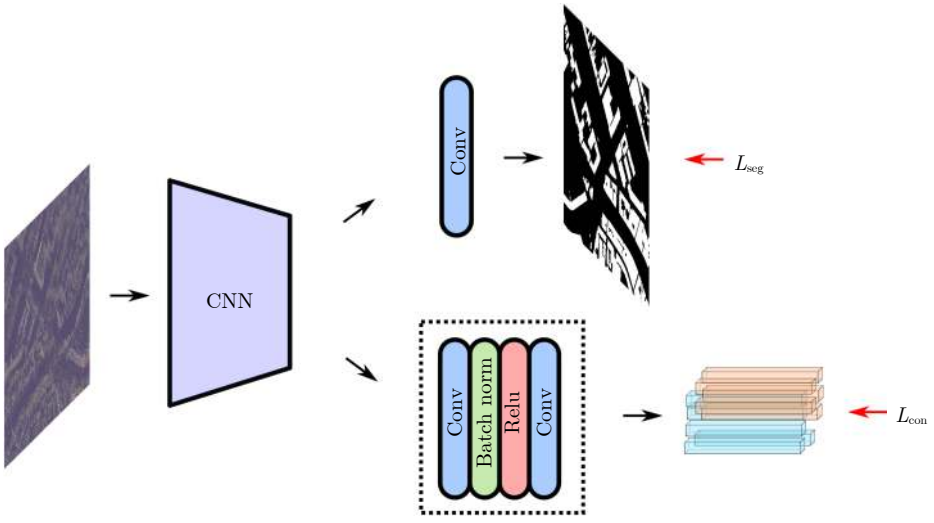


图 1 监督对比学习正则化的SAR图像建筑物提取模型示意图

Fig. 1　Supervised contrastive learning regularized SAR building footprint segmentation model

(a) 城市地区多极化SAR图像
(a) Multi-polarization SAR image of one urban area

(b) 同一地区RGB图像 (来源：SpaceNet6数据[15])
(分辨率均为0.5 m)
(b) The RGB optical image of the same area (SpaceNet6 data[15]) (Both resolutions are 0.5 m)

图 2 城市地区的多模态遥感图像

Fig. 2 Multi-modality remote sensing images for urban areas

使CNN模型在训练过程中充分考虑到建筑物与非建筑物像素在特征空间中的特征相似性，迫使模型能更好地对不同类别像素在特征空间中加以分辨，从而进一步提升SAR图像中建筑物提取精度。

为了实现上述效果，像素级别的监督对比学习旨在优化如下损失：

$$L_{\mathrm{con}}=\sum_{c\in\{0,1\}}\sum_{\boldsymbol{f}_{\mathrm{q}}\in\mathcal{R}_{\mathrm{q}}^{c}}-\lg\frac{\exp\left(\boldsymbol{f}_{\mathrm{q}}^{\mathrm{T}}\boldsymbol{f}_{\mathrm{k}}^{c,+}/\tau\right)}{\exp\left(\boldsymbol{f}_{\mathrm{q}}^{\mathrm{T}}\boldsymbol{f}_{\mathrm{k}}^{c,+}/\tau\right)+\sum\limits_{\boldsymbol{f}_{\mathrm{k}}^{-}\in\mathcal{R}_{\mathrm{k}}^{c}}\exp\left(\boldsymbol{f}_{\mathrm{q}}^{\mathrm{T}}\boldsymbol{f}_{\mathrm{k}}^{-}/\tau\right)} \tag{1}$$

$$\mathcal{R}_{\mathrm{q}}^{c}=\bigcup_{(u,v)}1(\boldsymbol{Y}[u,v]=c)\boldsymbol{F}[u,v,:] \tag{2}$$

$$\boldsymbol{f}_{\mathrm{k}}^{c,+}=\frac{1}{|\mathcal{R}_{\mathrm{q}}^{c}|}\sum_{\boldsymbol{f}_{\mathrm{q}}\in\mathcal{R}_{\mathrm{q}}^{c}}\boldsymbol{f}_{\mathrm{q}} \tag{3}$$

$$\mathcal{R}_{\mathrm{k}}^{c}=\bigcup_{(u,v)}1(\boldsymbol{Y}[u,v]\neq c)\boldsymbol{F}[u,v,:] \tag{4}$$

其中，c表示类别(0为背景，1为建筑物区域)，$\boldsymbol{f}_{\mathrm{q}}$表示对比学习损失函数所作用的查询(query)特征向量，其维度为D，$\boldsymbol{f}_{\mathrm{k}}^{c,+}$表示所有正例特征的平均，$\boldsymbol{f}_{\mathrm{k}}^{-}$表示负例特征，$\tau$为温度参数，$\mathcal{R}_{\mathrm{q}}^{c}$为查询特征向量选取集合，$\mathcal{R}_{\mathrm{k}}^{c}$表示与查询向量做对比的键(key)特征向量$\boldsymbol{f}_{\mathrm{k}}^{-}$选取集合，$\boldsymbol{Y}$表示类别的真值矩阵，$\boldsymbol{F}$为三维的特征张量。考虑到计算复杂性，本文从正例以及负例像素中选取一部分进行对比，M_{q}与M_{k}分别为查询特征向量与键特征向量的个数。在对比学习过程中，需要特别关注的是能引起较大损失的查询特征，而本文对已经能很好进行分辨的特征不必过多关注。如图3所示，能引起较大损失的特征通常在特征空间中离相应的类别特征均值比较远，属于难以分辨的特征，因此，查询特征需要从这些特征向量中进行选择，这样可以使模型收敛速度更快并且得到充分学习。

为此，本文采用了如下难分辨特征的采样方法：

$$\mathcal{R}_{\mathrm{q}}^{c,\mathrm{hard}}=\bigcup_{(u,v)}1\left(\boldsymbol{Y}[u,v]=c,\hat{\boldsymbol{Y}}[u,v]\leq\delta\right)\boldsymbol{F}[u,v,:] \tag{5}$$

$\hat{\boldsymbol{Y}}$表示模型预测的类别概率矩阵，δ为类别阈值，式(5)表示对于查询特征，本文选取类别预测概率较低的特征向量，而忽略已经能很好被分类且可信度较高的特征向量。对于键特征向量，本文从$\mathcal{R}_{\mathrm{k}}^{c}$中进行随机选取。

2.2 建筑物提取的联合损失函数

除了上述对比损失之外，本文采用常用的分割损失函数对建筑物区域进行学习，本文的分割损失包括焦点损失[25](Focal Loss)和Dice损失，其中Focal Loss为

$$L_{\mathrm{focal}}=\sum_{M}-(1-p_{\mathrm{t}})^{\gamma}\lg(p_{\mathrm{t}}) \tag{6}$$

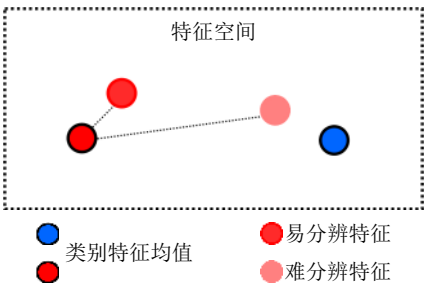


图 3 对比学习特征空间中的难易特征

Fig. 3 The easy and hard query feature vectors for contrastive learning

$$p_{\mathrm{t}}=\begin{cases}p, & \text{if } y=1\\ 1-p, & \text{otherwise}\end{cases} \tag{7}$$

其中，p为模型预测的类别概率，γ为超参数。与交叉熵损失函数相比，Focal Loss对于容易分类的像素点得到的损失小，从而避免模型由于“过度自信”所造成泛化能力下降的问题。Dice损失函数为

$$L_{\mathrm{Dice}}=1-\frac{2yp+1}{y+p+1} \tag{8}$$

其主要作用是提升模型预测结果的F1精度。至此，结合上述几种损失函数项，本文所提出用于SAR图像建筑物区域提取的联合损失函数为

$$L=L_{\mathrm{seg}}+L_{\mathrm{con}}=L_{\mathrm{focal}}+L_{\mathrm{Dice}}+L_{\mathrm{con}} \tag{9}$$

2.3 分割网络

本文提出的联合损失函数适用于任何分割网络结构，在实验中选取了比较常用的DeepLabV3+[20]以及UNet[26]，其中的特征提取网络分别选取了ResNet34与ResNet50[27]。DeepLabV3+与UNet网络结构如图4所示。

3 实验与分析

3.1 数据集

本文采用的数据集为2020年EarthVision竞赛发布的SpaceNet6[15]，其包括荷兰鹿特丹港120 km^{2}的3401张X波段全极化(HH, HV, VH和VV) SAR图像，其空间分辨率为0.5 m，每张大小为900×900，共有大约48,000个带标注的建筑物区域，如图5所示。为了图像显示方便，文中的SAR图像举例均选取HH, HV和VV 3个通道并将其转化为uint8格式。本文选取2696张图像作为训练，其余705张作为测试。

3.2 实验设计

对于输入数据，本文采用如表1的数据增强方法，并运用随机梯度下降(SGD)方法对网络模型进行优化，训练过程中的其他参数如表2所示。为了

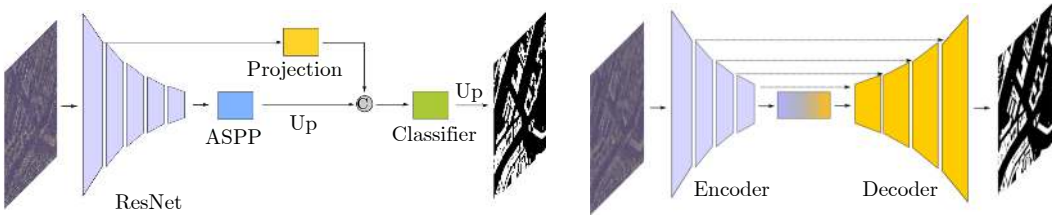


图 4 本文所采用的常用的DeepLabV3+与UNet网络结构

Fig. 4 The CNN architectures of DeepLabV3+ and UNet exploited in this paper

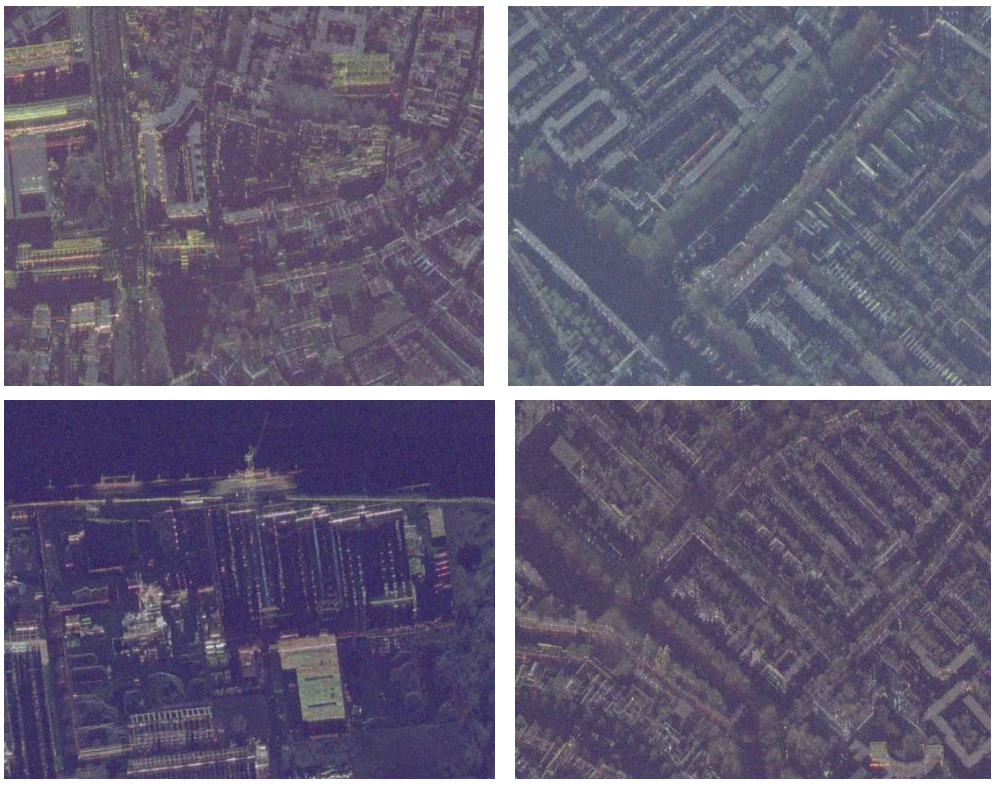


图 5 SpaceNet6城市地区全极化SAR图像(分辨率：0.5 m)

Fig. 5 Examples of SpaceNet6 full polarization SAR images (Resolution: 0.5 m)

衡量所提出方法的有效性，本文运用了包括F1分数在内的4种度量指标。

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{10}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{11}$$

$$\text{IoU} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{TN}} \tag{12}$$

$$\text{F1} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{TN}} \tag{13}$$

其中，TP, FP, FN, TN分别表示真正例、假正例、假负例和真负例，Precision与Recall分别表示精准度与召回率，即度量正例结果中有多少是真正例以及真正例像素有多少被挑选出来，IoU与F1主要衡量候选区域与真值区域的交叠程度，如果交叠度越高，则说明分割结果越精确。

表 1 训练过程采用的数据增强方法

Tab. 1 Adopted data augmentation methods for training

数据增强	数值
随机裁剪	512×512
水平翻转	0.5
归一化	均值：128 方差：32

表 2 训练过程中的参数设定

Tab. 2 Other parameters for training

参数	数值
M_q	256
M_k	512
τ	0.1
D	128
δ	0.97
Learning rate	1×10^{-3}
Batch size	16
Epoch	200

3.3 实验结果

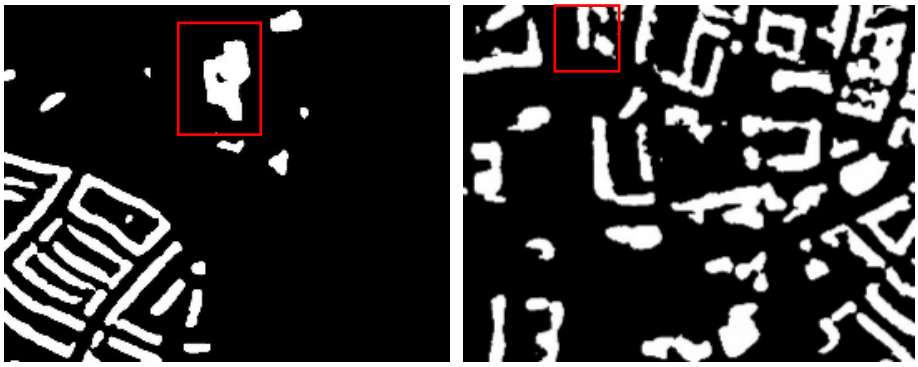
为了验证算法的有效性，在DeepLabV3+与UNet框架内，本文分别采用了ResNet34与ResNet50作为特征提取主干网络，并且测试并对比常用的分割损失函数包括Dice以及Focal+Dice下取得的分割精度，除此之外，本文也列出文献[15]所采用的方法在建筑物分割上取得的结果。针对每个模型，本文分别进行了3次独立重复实验，并且计算各个度量指标的均值与均方差。如表3所示，提出的联合损失函数分别在不同特征提取主干网络上均取得了最好的效果，相比于Focal+Dice，引入的对比损失正则化项可以将4项度量指标分别至少提升了1%，这表明在对建筑物区域进行分割的同时，需要考虑不同类别像素在特征空间之间的相互关系，通过优化像素级特征之间的紧凑性以及分离性，可以增强模型在测试数据上的泛化能力。与基准方法[15]相比，DeepLabV3+与UNet均取得了更好的效果，可能的原因是本文采用了ResNet系列的特征提取主干网络，相比于U-Net中采用的VGG网络能对图像中的语义特征刻画得更加有效。另外，本文模型选用了ImageNet预训练模型参数作为初始条件，相比于文献[15]中的随机初始化，其能更好地迁移到基于SAR图像的建筑物分割任务中。虽然提出的方法能明显提升建筑物识别精度，但是总体上来看，从高分辨率SAR中提取建筑物区域的精度不算很好，主要原因在于相比于光学图像，SAR图像的纹理及颜色特征不明显，而且成像过程中存在的相干斑噪声均对分割结果造成了影响。除此之外，基于SAR的成像机理，斜视视角下观测到的地表物体几何结构具有透视收缩效应，这与光学图像下的相同目标的几何结构大有不同，这也解释了表3中精确率(Precision)数值较高而召回率(Recall)数值较低的原因。为了能进一步明显地分析不同方法的建筑物分割效果，图6展示了DeepLabV3+

表 3 不同网络模型及损失函数下的建筑物提取性能比较(单位：%)

Tab. 3 Performance comparison of the building segmentation based on different methods (Unit: %)

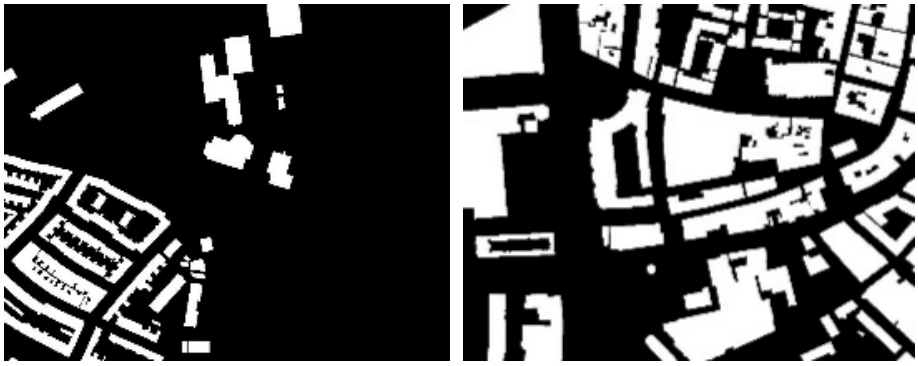
网络模型	损失函数	IoU	Dice	Precision	Recall
U-Net[15]	Focal+Dice	34.98[1.48]	51.81[1.61]	62.30[1.50]	44.42[2.38]
DeepLabV3+[ResNet34]	Focal+Dice	48.40[0.14]	65.23[0.13]	76.95[0.37]	56.61[0.03]
DeepLabV3+[ResNet34]	Focal+Dice+**CL**	49.81[0.31]	66.50[0.28]	78.32[0.48]	57.78[0.68]
DeepLabV3+[ResNet50]	Focal+Dice	48.72[0.48]	65.51[0.43]	76.55[0.31]	57.26[0.49]
DeepLabV3+[ResNet50]	Focal+Dice+**CL**	50.15[0.62]	66.79[0.54]	78.30[0.48]	58.25[0.87]
U-Net[ResNet34]	Focal+Dice	49.09[0.12]	65.85[0.11]	77.36[0.13]	57.33[0.20]
U-Net[ResNet34]	Focal+Dice+**CL**	50.62[0.58]	67.21[0.52]	76.32[0.25]	60.05[0.97]
U-Net[ResNet50]	Focal+Dice	49.24[0.30]	65.99[0.27]	77.37[0.23]	57.53[0.34]
U-Net[ResNet50]	Focal+Dice+**CL**	50.99[1.18]	67.53[1.03]	78.21[1.32]	59.50[2.36]

(a) 输入的全极化SAR图像
(a) The input full-polarization SAR images



(b) DeepLabV3+[ResNet50]运用Focal+Dice得到的结果
(b) The results obtained via DeepLabV3+[ResNet50] and Focal+Dice

(c) DeepLabV3+[ResNet50]在Focal+Dice+**CL**上得到的结果
(c) The results obtained via DeepLabV3+[ResNet50] and Focal+Dice+**CL**



(d) 两幅图像对应的真值
(d) The ground-truth masks

图 6 不同方法的建筑物提取结果

Fig. 6　Different methods for building footprint extraction of SAR images

[ResNet50]网络基于Focal+Dice以及Focal+Dice+**CL**在两张SAR城市图片上的建筑物提取结果。对比所框出区域，运用对比学习正则化项得到的建筑物区域更加精确，相反，采用Focal+Dice损失函数得到的结果较容易将建筑物像素以及背景像素相混淆，这进一步证明了本文所提出的对比学习正则化项在区分SAR图像中建筑物及背景像素上的有效性。

为了进一步分析所提出的对比学习正则化项，图7展示了难查询像素的所在区域，可以看出，分类结果置信度较低的区域一般属于建筑物边界区域，通过对这些区域像素点的采样，使得对比学习损失函数更加“关注”建筑物边界与背景像素的特征对比，从而使模型能更精准地对建筑物边界像素进行分类。图8展示了DeepLabV3+[ResNet50]模型在有与没有对比损失正则化下得到建筑物区域像素特征之间的相似性直方图，其中橙色表示监督对比正则化下得到的特征相似性，蓝色表示没有正则化下的特征相似性。在有对比损失函数的情况下，建筑物区域像素之间的特征相似性明显高于没有对比损失函数得到的特征预测结果，高的特征相似性可以使建筑物区域像素的特征更加一致，从而使其更容易被分类而且精度更高。本文所提出方法所涉及的主要参数包括查询及键像素的数量，即M_{q}与M_{k}，为了测试不同参数下所提方法对于建筑物区域分割精度的影响，本文在M_{q}=128, M_{k}=256以及M_{q}=512, M_{k}=1024两种不同参数下分别对所得结果进行评价，得到的F1数值如图9所示。由此可见，在M_{q}与M_{k}取较大值的情况下，训练得到的模型在建筑物提取上能取得更好的效果。不过大量的查询及键像素会增加训练计算量，降低模型训练速度，在实际应用中需要根据速度及精度指标要求灵活选取参数值。为了验证所提出方法在复杂城市地区建筑物提取上的效果，图10 给出了两块城市中心地区的全极化SAR图像以及本文方法取得的预测结果。从预测结果可以看出，所提出方法能将复杂

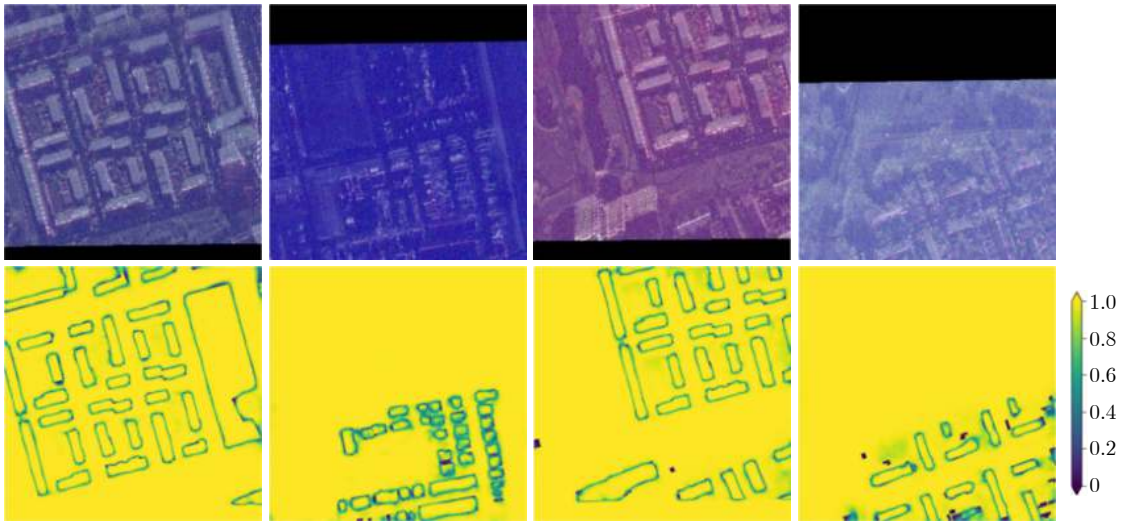


图 7 提出算法选取到的难查询像素：第1行为输入SAR图像，第2行为像素预测类别的置信度，颜色越深表示置信度越低

Fig. 7 Selection of hard query pixels: The first row is the input SAR images and the second row shows the classification confidences

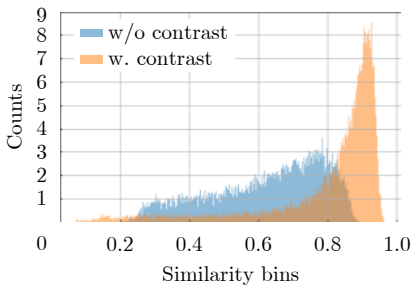


图 8 DeepLabV3+[ResNet50]模型在有与没有对比损失正则化下得到建筑物区域像素特征之间的相似性直方图

Fig. 8 The histogram of the feature similarities among the building pixels obtained by the trained DeepLabV3+[ResNet50] models with and without the contrastive loss

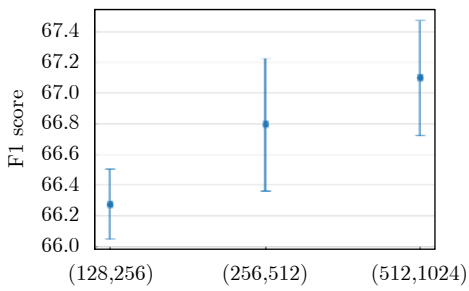
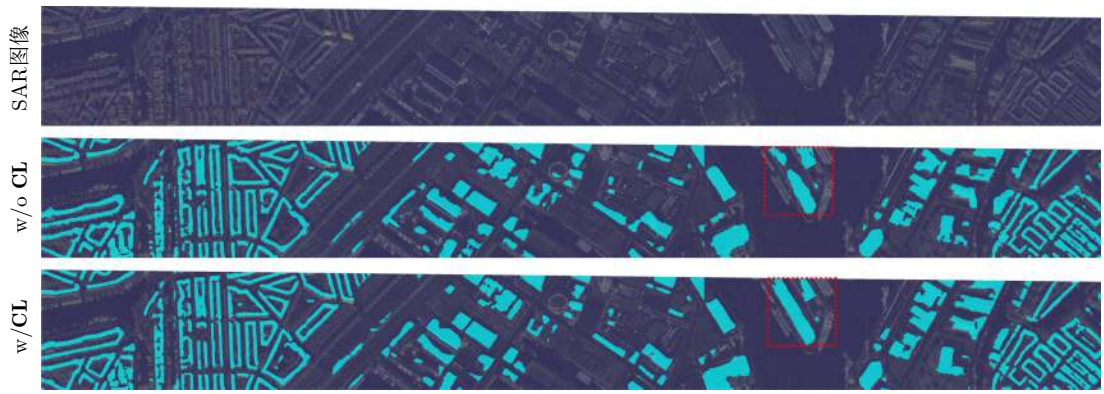
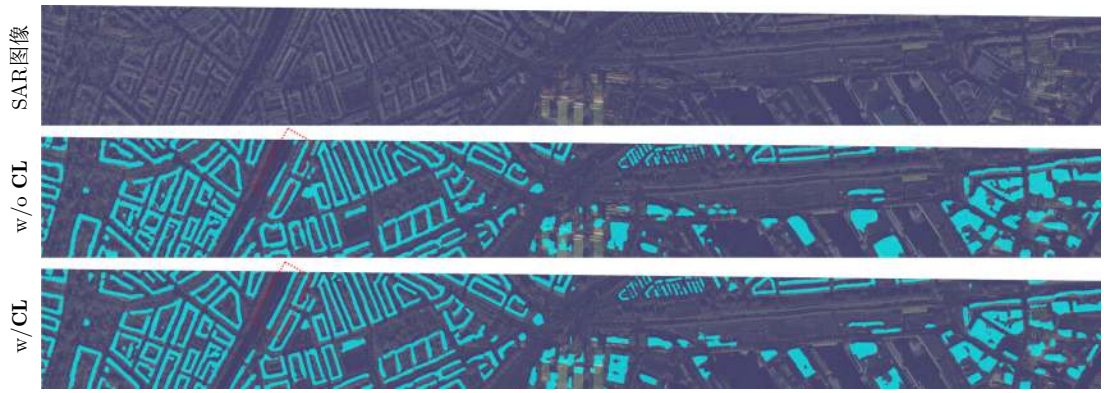


图 9 不同查询及键像素数量对所提出方法的敏感性分析(采用DeepLabV3+网络结构)

Fig. 9 The sensitivity analysis of the proposed method under different numbers of query and key pixels (The CNN architecture of DeepLabV3+ is adopted)



(a) 城市中心区域A
(a) Urban center A



(b) 城市中心区域B
(b) Urban center B

图 10 所提出方法得到的大范围城市地区建筑物提取结果(DeepLabV3+[ResNet50])

Fig. 10 The large-scale urban building extraction result based on the proposed method (DeepLabV3+[ResNet50])

城市地区的大部分建筑物识别出来，并且在红色框选出的建筑物区域，方法能更好地将建筑识别出来，而没有对比正则化项的损失函数在这些区域预测结果的一致性较差，从而产生整个建筑物被割裂的现象，因此，所提出的方法能较好地对大范围城市地区建筑物区域进行识别提取。

4　结束语

本文针对高分辨率SAR图像的建筑物提取，提出了基于对比学习正则化的联合损失函数，通过对图像中建筑物与背景区域像素在特征空间中的语义关系建模，增强同一类别像素的特征相似性，同时减弱不同类别像素的特征相似性，可以有效地对城市地区建筑物以及复杂背景区域像素进行分离，从而提升训练模型对于建筑物与背景像素的判别能力以及对新数据的泛化能力。在SpaceNet6数据上取得了比常用分割损失函数如Dice, Focal+Dice等更加有效的建筑物提取结果，分割精度提升至少1%，这对于需要精确度量建筑物区域面积的任务中，如恶劣天气条件下评估受灾建筑物面积，有着重要应用价值。然而，本团队在实验过程中发现小型建筑物(别墅等)以及高楼(写字楼等)的分割精度还有待提高，后续工作将着眼于利用多源异构图像融合技术来提升SAR图像建筑物分割精度。

参 考 文 献

[1] 徐丰, 王海鹏, 金亚秋. 深度学习在SAR目标识别与地物分类中的应用[J]. 雷达学报, 2017, 6(2): 136–148. doi: 10.12000/JR16130.
XU Feng, WANG Haipeng, and JIN Yaqiu. Deep learning as applied in SAR target recognition and terrain classification[J]. *Journal of Radars*, 2017, 6(2): 136–148. doi: 10.12000/JR16130.

[2] 王雪松, 陈思伟. 合成孔径雷达极化成像解译识别技术的进展与展望[J]. 雷达学报, 2020, 9(2): 259–276. doi: 10.12000/JR19109.
WANG Xuesong and CHEN Siwei. Polarimetric synthetic aperture radar interpretation and recognition: Advances and perspectives[J]. *Journal of Radars*, 2020, 9(2): 259–276. doi: 10.12000/JR19109.

[3] 丁赤飚, 仇晓兰, 徐丰, 等. 合成孔径雷达三维成像——从层析、阵列到微波视觉[J]. 雷达学报, 2019, 8(6): 693–709. doi: 10.12000/JR19090.

DING Chibiao, QIU Xiaolan, XU Feng, *et al*. Synthetic aperture radar three-dimensional imaging—from TomoSAR and array InSAR to microwave vision[J]. *Journal of Radars*, 2019, 8(6): 693–709. doi: 10.12000/JR19090.

[4] 李宁, 牛世林. 基于局部超分辨重建的高精度SAR图像水域分割方法[J]. 雷达学报, 2020, 9(1): 174–184. doi: 10.12000/JR19096.
LI Ning and NIU Shilin. High-precision water segmentation from synthetic aperture radar images based on local super-resolution restoration technology[J]. *Journal of Radars*, 2020, 9(1): 174–184. doi: 10.12000/JR19096.

[5] ZHAO Lingjun, ZHOU Xiaoguang, and KUANG Gangyao. Building detection from urban SAR image using building characteristics and contextual information[J]. *EURASIP Journal on Advances in Signal Processing*, 2013, 2013: 56. doi: 10.1186/1687-6180-2013-56.

[6] TUPIN F and ROUX M. Detection of building outlines based on the fusion of SAR and optical features[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2003, 58(1/2): 71–82.

[7] XU Feng and JIN Yaqin. Automatic reconstruction of building objects from multiaspect meter-resolution SAR images[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2007, 45(7): 2336–2353. doi: 10.1109/TGRS.2007.896614.

[8] MICHAELSEN E, SOERGEL U, and THOENNESSEN U. Perceptual grouping for automatic detection of man-made structures in high-resolution SAR data[J]. *Pattern Recognition Letters*, 2006, 27(4): 218–225. doi: 10.1016/j.patrec.2005.08.002.

[9] FERRO A, BRUNNER D, and BRUZZONE L. Automatic detection and reconstruction of building radar footprints from single VHR SAR images[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2012, 51(2): 935–952.

[10] ZHANG Fengli, SHAO Yun, ZHANG Xiao, *et al*. Building L-shape footprint extraction from high resolution SAR image[C]. 2011 Joint Urban Remote Sensing Event, Munich, Germany, 2011: 273–276.

[11] WANG Yinghua, TUPIN F, HAN Chongzhao, *et al*. Building detection from high resolution PolSAR data by combining region and edge information[C]. IGARSS 2008-2008 IEEE International Geoscience and Remote Sensing Symposium, Boston, MA, USA, 2009: IV–153.

[12] GOODFELLOW I, BENGIO Y, and COURVILLE A. Deep Learning[M]. Cambridge: MIT press, 2016: 1–800.

[13] WANG Xiaying, CAVIGELLI L, EGGIMANN M, *et al*. Hr-SAR-NET: A deep neural network for urban scene segmentation from high-resolution SAR data[C]. 2020 IEEE Sensors Applications Symposium (SAS), Kuala Lumpur, Malaysia, 2020: 1–6.

[14] 杜康宁, 邓云凯, 王宇, 等. 基于多层神经网络的中分辨SAR图像时间序列建筑区域提取[J]. 雷达学报, 2016, 5(4): 410–418. doi: 10.12000/JR16060.
DU Kangning, DENG Yunkai, WANG Yu, *et al*. Medium resolution SAR image time-series built-up area extraction based on multilayer neural network[J]. *Journal of Radars*, 2016, 5(4): 410–418. doi: 10.12000/JR16060.

[15] SHERMEYER J, HOGAN D, BROWN J, *et al*. SpaceNet 6: Multi-sensor all weather mapping dataset[C]. The 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, USA, 2020: 768–777.

[16] SHAHZAD M, MAURER M, FRAUNDORFER F, *et al*. Buildings detection in VHR SAR images using fully convolution neural networks[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2019, 57(2): 1100–1116. doi: 10.1109/TGRS.2018.2864716.

[17] JING Hao, SUN Xian, WANG Zhirui, *et al*. Fine building segmentation in high-resolution SAR images via selective pyramid dilated network[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2021, 14: 6608–6623. doi: 10.1109/JSTARS.2021.3076085.

[18] CHEN Jiankun, QIU Xiaolan, DING Chibiao, *et al*. CVCMFF Net: Complex-valued convolutional and multifeature fusion network for building semantic segmentation of InSAR images[J]. *IEEE Transactions on Geoscience and Remote Sensing*, in press. doi: 10.1109/TGRS.2021.3068124.

[19] SUN Yao, HUA Yuansheng, MOU Lichao, *et al*. CG-Net: Conditional GIS-Aware network for individual building segmentation in VHR SAR images[J]. *IEEE Transactions on Geoscience and Remote Sensing*, in press. doi: 10.1109/TGRS.2020.3043089.

[20] CHEN L, ZHU Yukun, PAPANDREOU G, *et al*. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]. The European Conference on Computer Vision, Munich, Germany, 2018: 833–851.

[21] CHEN Ting, KORNBLITH S, NOROUZI M, *et al*. A simple framework for contrastive learning of visual representations[C]. The 37th International Conference on Machine Learning, Virtual Event, 2020: 1597–1607.

[22] HE Kaiming, FAN Haoqi, WU Yuxin, *et al*. Momentum contrast for unsupervised visual representation learning[C]. The 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, USA, 2020: 9726–9735.

[23] KHOSLA P, TETERWAK P, WANG Chen, *et al*. Supervised contrastive learning[C]. Advances in Neural Information Processing Systems, Virtual, 2020: 18661–18673.

[24] LIU Shikun, ZHI Shuaifeng, JOHNS E, *et al*. Bootstrapping semantic segmentation with regional contrast[J]. arXiv:

2104.04465, 2021.

[25] LIN T Y, GOYAL P, GIRSHICK R, *et al*. Focal loss for dense object detection[C]. The 2017 IEEE International Conference on Computer Vision, Venice, Italy, 2017: 2999–3007.

[26] RONNEBERGER O, FISCHER P, and BROX T. U-Net: Convolutional networks for biomedical image segmentation[C]. 18th International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 2015: 234–241.

[27] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, *et al*. Deep residual learning for image recognition[C]. The 2016 IEEE Conference on Computer Vision and Pattern Recognition, Caesars Palace, Las Vegas, USA, 2016: 770–778.

作 者 简 介

康　健(1991–)，男，2019年在慕尼黑工业大学获得博士学位，现任苏州大学电子信息学院副教授，硕士生导师，IEEE会员，雷达学报客座编辑。主要研究方向为遥感图像智能解译。

王智睿(1990–)，男，2018年在清华大学获得博士学位，现任中国科学院空天信息创新研究院助理研究员。主要研究方向为SAR图像智能解译。

祝若鑫(1991–)，男，德国工学博士(Dr.-Ing.)，西安测绘研究所助理研究员。主要研究方向为社会感知和时空数据挖掘。

孙　显(1981–)，男，中国科学院空天信息创新研究院研究员，博士生导师，IEEE高级会员，雷达学报青年编委。主要研究方向为计算机视觉与遥感图像理解。