

doi: 10.6046/zrzyyg.2021421

引用格式: 孙盛, 蒙芝敏, 胡忠文, 等. 多尺度轻量化 CNN 在 SAR 图像地物分类中的应用[J]. 自然资源遥感, 2023, 35(1): 27-34. (Sun S, Meng Z M, Hu Z W, et al. Application of multi-scale and lightweight CNN in SAR image-based surface feature classification[J]. Remote Sensing for Natural Resources, 2023, 35(1): 27-34.)

# 多尺度轻量化 CNN 在 SAR 图像地物分类中的应用

孙盛<sup>1</sup>, 蒙芝敏<sup>1</sup>, 胡忠文<sup>2</sup>, 余旭<sup>3</sup>

(1. 广东工业大学计算机学院, 广州 510006; 2. 深圳大学自然资源部大湾区地理环境监测重点实验室, 深圳 518000; 3. 广东工业大学土木与交通工程学院, 广州 510006)

**摘要:** 结合粤港澳大湾区的亚热带气候特点, 采用 TerraSAR-X 雷达遥感卫星对实验区域进行了图像采集; 针对雷达卫星观测场景中地物目标尺度变化不一的问题, 提出了一个应用于地物分类的卷积神经网络模型 (ENet convolution spatial pyramid pooling, ENet-CSPP)。利用了普通卷积比空洞卷积更好保持领域信息的特点, 提出了多尺度特征融合模块——卷积空域金字塔池化模块; 针对 SAR 遥感图像数据集训练样本偏少的问题, 提出了将多尺度特征融合模块和轻量化卷积神经网络结合起来的方法; ENet-CSPP 网络的编码器部分由改进后的 ENet 网络和卷积空域金字塔池化模块构成, 解码器部分实现深、浅层特征的融合后输出地物分类图像。在 GDUT-Nansha 数据集上进行了定量对比实验, ENet-CSPP 模型在像素精度、平均像素精度和平均交并比 3 个性能指标上都要优于其他模型, 表明多尺度轻量化的模型有效提高了地物分类的精度。

**关键词:** 合成孔径雷达; 地物分类; 卷积神经网络; 轻量化网络

**中图法分类号:** P 236 **文献标志码:** A **文章编号:** 2097-035X(2023)01-0027-08

## 0 引言

遥感图像地物分类是指把遥感图像中的每个像元划归为某一个语义类别的过程, 其主要原理是根据地物电磁波辐射在遥感图像上的特征, 判读识别地表物体的类属<sup>[1]</sup>, 而在计算机视觉领域中, 一般称之为语义分割, 其结果可为城市规划和土地使用状况监测等提供基础数据支撑。本文以广州市南沙区和东莞市虎门镇作为研究区域, 该区域位于粤港澳大湾区, 是中国最具经济活力和发展最快的区域之一, 在过去的 20 a 里, 该区域城市建设开发速度不断加快, 居民及建设用地的面积占比从 2000 年的 8% 增加为 2020 年的 17%, 同时林地、耕地和草地等土地面积占比都有一定程度的减少<sup>[2]</sup>。该区域仍然处于快速的城市开发进程中, 对于这个区域进行连续的土地使用状况监测可为土地规划部门提供有价值的参考数据。研究区域位于广东省中南部, 属于亚热带气候, 气温较高, 降水充沛, 多云多雾, 使用光学遥感卫星技术手段采集地物信息将受到较大

的限制。合成孔径雷达 (synthetic aperture Radar, SAR) 系统<sup>[3]</sup> 具有不受天气、光线以及云层影响的特点, 能够实现全天时、全天候的成像, 比较适合于该区域的遥感信息采集。

有较多学者已使用基于卷积神经网络的经典语义分割模型在 SAR 图像上开展了地物分类的研究, 并取得了若干成果。Yao 等<sup>[4]</sup> 以德国 TerraSAR-X 雷达卫星图像和 Google Earth 光学图像构建数据集, 使用全卷积神经网络 (fully convolutional networks, FCN)<sup>[5]</sup> 在数据集上实现了地物分类, 得到的平均交并比 (mean intersection over union, mIoU) 值为 0.3; Henry 等<sup>[6]</sup> 在 TerraSAR-X 上构建了关于道路 SAR 图像数据集, 并使用 FCN, U-Net 和 DeepLabv3+ 这 3 种网络进行地物分类, 其中像素精度 (pixel accuracy, PA) 最高达到了 0.71, 交并比 (intersection over union, IoU) 值最高达到了 0.45; Wu 等<sup>[7]</sup> 以航空机载雷达得到的 SAR 图像构建了一个包含建筑物、水体、植被、道路和其他地物共 5 类目标的数据集, 并使用 FCN 和 U-Net 对其进行地物分类, 得到的 PA 值都为 0.84, 得到的 mIoU 值分别

收稿日期: 2021-12-03; 修订日期: 2022-05-05

基金项目: 国家自然科学基金项目“纠缠态的超级纠缠目击者特性及其推广研究”(编号: 61672007)、自然资源部大湾区地理环境监测重点实验室开放基金项目“基于多极化星载合成孔径雷达图像的粤港澳大湾区海岸线动态监测”(编号: 2019002)和广东省国际合作领域项目(编号: 2019A050509009)共同资助。

第一作者: 孙盛(1980-), 男, 博士, 副教授, 研究方向为遥感图像处理 and 计算机视觉。Email: sunsheng@gdut.edu.cn。

为 0.44 和 0.5。

但是,真实观测场景中存在尺度大小不一的观测目标,通过遥感技术获取的图像中,对应的目标尺度存在一定差异。如果分类模型中卷积核的感受野只能覆盖大尺度目标的一部分,则无法感知目标完整的特征;反之,如果感受野比目标大得多,背景像素中含有的特征会在训练过程中被模型学习到。因此,根据单一尺度的特征来预测所有像素的类别,会造成地物分类精度的降低。为了解决尺度不一的问题,可采用多个具有不同感受野的卷积核提取多尺度的图像特征,然后将提取到的不同尺度的特征融合,以提高地物分类的鲁棒性和精度。2017 年,Zhao 等<sup>[8]</sup>在 PSPNet 中提出了金字塔池化模块(pyramid pooling module, PPM),这个模块包括多个池化核,分别覆盖图像的整体、一半和小部分,提取多尺度信息并融合,在 PASCAL VOC 2012 数据集上得到的 mIoU 值为 0.854,比未采用多尺度特征融合的 DeepLabV1<sup>[9]</sup>模型提高了 13.8%;Chen 等<sup>[10]</sup>在 2017 年提出了空洞卷积空间金字塔池化(atrous spatial pyramid pooling, ASPP)模块,该模块采用多个不同膨胀率的卷积核并行提取多尺度特征后融合,结合该模块的 DeepLabV2 模型在 PASCAL VOC 2012 数据集上得到的 mIoU 值为 0.797,比 DeepLabV1 模型提高了 8.1%。相比于单一尺度模型,使用了多尺度特征融合策略的语义分割网络模型对应的分类精度都有了一定提高。

基于多卷积核并行提取多尺度特征的策略,本文提出一种应用于 SAR 图像地物分类的多尺度轻量化卷积神经网络模型 ENet - CSPP。模型使用了改进后的 ENet 网络和卷积空域金字塔池化(convolution spatial pyramid pooling, CSPP)模块作为编码器,使用了深、浅层特征融合的解码器,更加有效地提取了尺度变化较大目标的特征,在一定程度上提高了 SAR 图像地物分类的精度。

## 1 多尺度特征融合的 SAR 图像地物分类模型

### 1.1 模型构建策略

上述多尺度特征提取和融合的研究都是基于自然图像,ASPP 模块采用了多个不同膨胀率的空洞卷积核并行提取图像特征,空洞卷积有利于扩大感受野,充分捕捉图像特征,但由于卷积核中含有空洞,并不是每个输入的像素都参与计算,相邻像素的特征会丢失。SAR 图像中相邻像素的类别可能变化巨大,如果丢失邻域特征,将导致地物类型预测性能

的降低<sup>[11]</sup>。另一方面,当图像中物体较小而空洞卷积的膨胀率较大时,3×3 空洞卷积就会退化成 1×1 的卷积<sup>[12]</sup>,感受野减小了。因此,需要将 ASPP 模块中的空洞卷积改为普通卷积,称为 CSPP 模块。CSPP 模块可以多尺度的方式提取图像的特征,然后进行融合并输出,同时也保留了图像的邻域特征。

较多基于卷积神经网络的经典语义分割模型都采用了编码器 - 解码器的结构,其用于提取特征的编码器部分大都是基于 VGG<sup>[13]</sup>网络或者 ResNet<sup>[14]</sup>等深层网络,深层网络需要训练大量的网络参数<sup>[15]</sup>,而对于 SAR 图像而言,目前学术界还没有一种包含数十万或者更多样本的数据集。如果使用大型网络进行训练,容易发生过拟合,导致预测性能大幅度降低。因此,可采用轻量化网络 ENet 的一部分作为模型的特征提取部分。ENet 作为一种实时语义分割的网络,其性能在一些数据集上与 SegNet 相当,但其推理时间仅为 SegNet 的 1/18,对应的参数减少了 79 倍<sup>[16]</sup>。本文提出的模型采用了 ENet 网络的下采样部分,将其改进后与 CSPP 模块结合,作为 ENet - CSPP 的特征提取部分,即编码器。

解码器的作用主要是将编码器提取的深层特征图恢复成与原图分辨率相同的分类图像。SegNet 和 U - Net 等网络的编码器使用双线性插值法逐层恢复图像分辨率<sup>[17-18]</sup>,该方法计算量小,但恢复的图像中目标的边缘较为粗糙,导致分类精度有一定的损失;DeepLabV3+ 模型的编码器采用深、浅层特征融合的方法恢复图像分辨率<sup>[19]</sup>,该方法能更好地恢复目标的边缘细节,对分类精度有一定的提升。

根据上述分析和讨论,ENet - CSPP 模型也采用编码器 - 解码器结构。为了减少模型深度,将 ENet 网络的下采样部分轻量化改进后与 CSPP 结构结合,作为编码器模块;解码器模块融合深、浅层特征,然后进行上采样,将特征图恢复成原图分辨率。

### 1.2 ENet - CSPP 模型结构

#### 1.2.1 编码器

编码器的作用主要是提取图像的深层特征,模型的编码器部分基于改进后的 ENet 网络的下采样部分和 CSPP 结构。具体而言,对 ENet 网络做出了如下改进:①去除 initial 层,减少了下采样次数,这将有助于保持特征图分辨率、减少空间信息的损失、提高模型精度;②移除了 section2,减少了网络深度,进一步减少模型的参数量和计算量,较大降低了过拟合风险;③去除了 bottleneck4.0 ~ bottleneck5.1,使用 CSPP 模块代替,而其中的上采样模块放到了解码器部分。输入图像分辨率为 512 像素×512 像素时 ENet 和 ENet - CSPP 的网络结构如表 1 所示,

表 1 中  $C$  为分类的类别数目;  $N1$  为浅层特征通道数;  $N2$  为深层特征通道数。

表 1 ENet 和 ENet - CSPP 的结构对比  
Tab. 1 Comparison of ENet and ENet - CSPP structures

ENet			ENet - CSPP		
层名	采样类型	输出大小/像素	层名	采样类型	输出大小/像素
initial		$16 \times 256 \times 256$			
bottleneck1.0	downsampling	$64 \times 128 \times 128$	bottleneck1.0	downsampling	$64 \times 256 \times 256$
4bottleneck1.x		$64 \times 128 \times 128$	4bottleneck1.x		$64 \times 256 \times 256$
bottleneck2.0	downsampling	$128 \times 64 \times 64$	bottleneck2.0	downsampling	$128 \times 128 \times 128$
8bottleneck2.x		$128 \times 64 \times 64$	8bottleneck2.x		$128 \times 128 \times 128$
repeat section 2, without bottleneck2.0			CSPP		
bottleneck4.0	upsampling	$64 \times 128 \times 128$	upsample1.0	upsampling	$N2 \times 256 \times 256$
2bottleneck4.x		$64 \times 128 \times 128$	concat		$(N1 + N2) \times 256 \times 256$
bottleneck5.0	upsampling	$16 \times 256 \times 256$	lastconv		$C \times 256 \times 256$
2bottleneck5.x		$16 \times 256 \times 256$	upsample2.0	upsampling	$C \times 512 \times 512$
fullconv		$C \times 512 \times 512$			

为了进一步提取多尺度的特征信息,编码器后半部分将使用 CSPP 模块对 ENet 网络输出的特征进行处理。主要对 ASPP 模块的改进为:移除膨胀率分别为 6,12,18 的空洞卷积层,替换成卷积核尺寸分别为  $3 \times 3, 5 \times 5, 7 \times 7$  的常规卷积层。使用普通卷积将有助于改进空洞卷积造成邻域信息丢失的缺点,提高对小目标的分类效果。另一方面,选择了  $5 \times 5$  和  $7 \times 7$  等较大尺寸的卷积核,有助于模型维持大的感受野。

改进的 ENet 网络得到的输出,输入到 CSPP 模块,该模块包含了多个不同尺寸的卷积层以及一个全图平均池化层,得到了不同分辨率的特征。将多层特征通过 concat 操作拼接融合,由于全图平均池化层的输出特征图分辨率与其他卷积层特征图分辨率不同,与其他分辨率的特征融合之前,需要对其进行上采样 2 倍的操作。拼接融合的特征最后通过  $1 \times 1$  卷积层,从而减少了通道数量;最终,输出通

道数为  $N2$ 、分辨率为  $128$  像素  $\times$   $128$  像素的特征图到解码器模块中。

### 1.2.2 解码器

解码器的作用主要是将特征图还原成与原输入图像具有相同分辨率的图像,从而恢复图像的细节。首先,将编码器的输出上采样 2 倍后,与来自编码器中提取的相同分辨率的低层特征进行拼接融合;然后,将得到的融合特征经过 2 个  $3 \times 3$  和 1 个  $1 \times 1$  卷积层处理后,缩减其通道数;最后,对特征图再进行 2 倍上采样,得到与原输入图像相同分辨率的特征图。

### 1.2.3 总体结构

总体结构如图 1 所示,编码器输出的浅层特征和深层特征经过 2 个  $1 \times 1$  卷积层处理后,得到  $N1$  通道数的浅层特征和  $N2$  通道数的深层特征,然后再输入到解码器中进行后续处理。

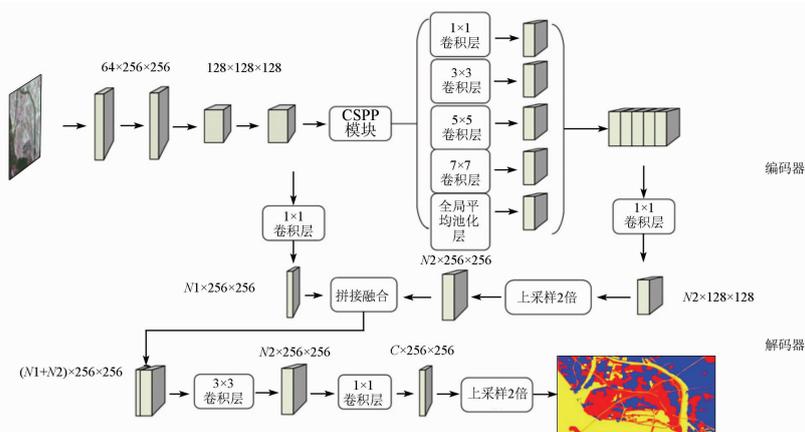


图 1 总体结构

Fig. 1 Overall architecture

### 1.3 数据集及分类参考图

数据集来源于 GDUT - Nansha<sup>[20]</sup> SAR 图像数

据集,该数据集基于 TerraSAR - X<sup>[21]</sup> 双极化数据构建(极化方式为 HH/VV),覆盖范围为广州市南沙

区整体和东莞市虎门镇一部分。该景 TerraSAR - X 双极化图像的参数为: 采集于 2019 年 5 月 14 日 22: 48: 31(UTC 时间), 图像大小为 23 642 像素 × 8 432 像素, 近入射角为 26. 065°, 远射角为 27. 746°, 卫星的运动方向为降轨, 图像的方位向像元分辨率为 6. 59 m, 距离向像元分辨率为 2. 70 m。该数据集包含 92 幅 2 048 像素 × 1 024 像素的图像, 裁剪后得到 704 幅 512 像素 × 512 像素的图像。为了扩充样本数量, 使用旋转 90°, 180°, 270° 以及镜像变换等方法对数据进行增广, 同时随机分配 15%

样本作为验证集, 最终得到含有 4 788 幅图像的训练集, 含有 844 幅图像的验证集, 部分训练集数据如图 2 所示。地物类型设定为 4 个类型, 分别为植被、建筑物、水体和道路。分类参考图对应的区域为东莞市虎门镇, 来源于第三方测绘公司航拍采的光学图像, 该图像的航测时间为 2019 年 1—2 月, 与 TerraSAR - X 卫星数据采集时间基本同步。将图像进行内业数据处理后, 结合人工实地调查核对, 得到与测试集数据具有相同地理区域的分类参考图。

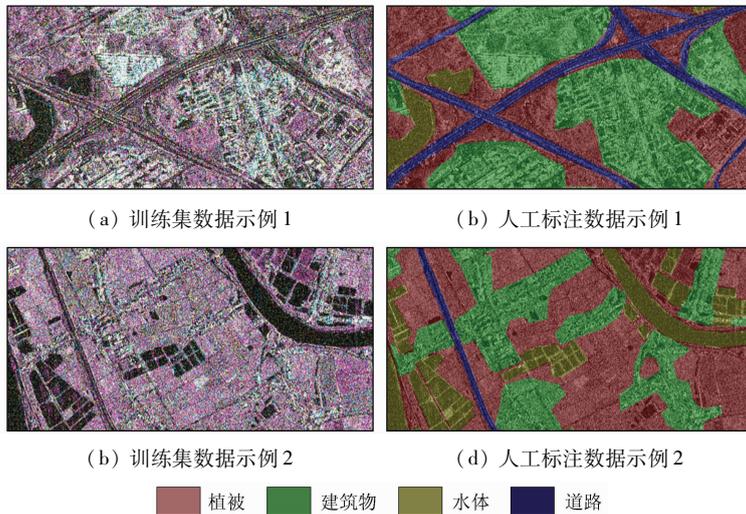


图 2 部分训练集数据

Fig. 2 Partial training set data

测试集数据裁剪于另一景 TerraSAR - X 数据, 裁剪区域位于东莞市虎门镇的一部分, 此区域图像没有包含在训练集中, 该区域具体位置如图 3 所示。该景 TerraSAR - X 双极化图像的参数为: 采集于 2019 年 5 月 15 日 22: 31: 26(UTC 时间), 图像大小为 26 015 像素 × 8 496 像素, 近入射角为 24. 899°, 远入射角为 26. 680°; 卫星的运动方向、图像的方位向和距离向像元分辨率与训练集数据相同。

基本核心频率为 1 320 MHz; 使用交叉熵函数作为损失函数, 损失函数值  $loss$  的公式为:

$$loss = - \frac{1}{MN} \sum_{m=1}^M \sum_{n=1}^N \sum_{c=1}^C y_{ic} \ln(p_{ic}) \quad (1)$$

式中:  $M$  为 BatchSize 的大小, 即每次迭代送入网络进行训练的样本数目, 本文设置为 8;  $N$  为输入图像的像素数量, 本文输入图像分辨率为 512 像素 × 512 像素,  $N$  为 262 144;  $C$  为类别的数目, 本文设置为 4 类;  $y_{ic}$  为符号函数, 如果样本  $i$  的真实类别等于  $c$  则取 1, 否则取 0;  $p_{ic}$  为样本  $i$  预测类别为  $c$  的概率。

为了防止过拟合, 模型采取了早停法和动态调整学习率相结合的训练策略。如果在连续 7 个轮次之内, 验证集的损失函数值没有降低, 则自动停止训练; 如果在连续 4 个轮次之内, 验证集的损失函数值没有降低, 则自动调整学习率为上一轮学习率的 0.1 倍。



图 3 测试集地理区域

Fig. 3 Testing set area

### 1.4 训练策略

模型的训练使用 Adam 优化器, 初始学习率为 0.01; 训练使用的 NVIDIA GeForce RTX 3060 显卡, 其显存容量为 12 GB, CUDA 核心数量为 3 584 个,

## 2 模型预测与实验结果分析

### 2.1 模型性能评价指标

本文主要采用 PA、平均像素精度 (mean pixel

accuracy, mPA)、IoU 和 mIoU 这 4 个指标来评估模型。PA 为分类正确的像素点数和所有的像素点数的比值; mPA 为类别像素精度的平均值; IoU 为预测值和真实值的交集与预测值和真实值的并集之间的比值。这个指标用来表达真实值和预测值之间的相关程度,该值越高,相关度越高。mIoU 为类别 IoU 的平均值。其公式分别为:

$$PA = \frac{\sum_{i=0}^C P_{ii}}{\sum_{i=0}^C \sum_{j=0}^C P_{ij}}, \quad (2)$$

$$mPA = \frac{1}{C+1} \sum_{i=0}^C \frac{P_{ii}}{\sum_{j=0}^C P_{ij}}, \quad (3)$$

$$IoU = \sum_{i=0}^C \frac{P_{ii}}{\sum_{j=0}^C P_{ij} + \sum_{j=0}^C P_{ji} - P_{ii}}, \quad (4)$$

$$mIoU = \frac{1}{C+1} \sum_{i=0}^C \frac{P_{ii}}{\sum_{j=0}^C P_{ij} + \sum_{j=0}^C P_{ji} - P_{ii}}, \quad (5)$$

式中:  $p_{ii}$  为第  $i$  类别正确分类的像素点个数;  $p_{ij}$  为真实值为  $i$  类别的像素预测为  $j$  类别的像素点个数;  $p_{ji}$  为真实值为  $j$  类别的像素预测为  $i$  类别的像素点个数。

## 2.2 交叉熵损失函数权重对模型预测性能的影响

为了减少类别不平衡对模型性能的影响,使用了带权重的交叉熵损失函数。为了验证不同类别权重对模型精度的影响,使用 PA, mPA 和 mIoU 作为性能评价指标开展实验。实验结果如表 2 所示,权重参数一列的数值依次为植被、建筑物、水体、道路的权重数值,该数值  $Weight$  的计算公式为:

$$Weight = 1 - \frac{P_c}{P}, \quad (6)$$

式中:  $P_c$  为样本类别为  $c$  的像素数目;  $P$  为总样本的像素数目。

表 2 交叉熵损失函数类别权重对模型性能的影响

权重参数	PA	mPA	mIoU
无权重	88.8	78.8	70.1
0.64, 0.73, 0.65, 0.96	88.6	81.5	71.0

训练集中各个类别地物对应的像素数量是不均衡的,其中道路类别对应像素数量远少于其他类别。式(6)根据样本类别的比例来计算权重,减小了植被、建筑物和水体的权重,3种地物目标的语义分割

性能指标有轻微下降,PA 相对于无权重参数的设置下降了 0.2 百分点;因为加大了道路的权重,道路地物目标的语义分割精度得到提高,相对于无权重参数的设置下,其 mPA 上升了 2.7 百分点, mIoU 也上升了 0.9 百分点,上述实验验证了带权重的交叉熵损失函数对于解决样本不均衡问题具有较好的效果。

## 2.3 融合特征通道的数量对模型预测性能的影响

在表 1 ENet - CSPP 结构中提到的参数  $N1$  和  $N2$  分别为浅层特征通道数和深层特征通道数。这 2 个参数决定浅层特征和深层特征对模型预测性能的影响。以 PA, mPA 和 mIoU 作为指标,验证不同浅层、深层特征通道数融合后对实验结果的影响。实验结果如表 3 所示。 $N2$  的取值范围为  $(0, 128]$ ,  $N1$  的取值范围为  $[0, 64]$ ,  $N2$  和  $N1$  的取值为离散取值,取最大取值的  $1/2, 1/4, 1/8$  以及最大值本身;深层特征作为网络的主要特征输出,因此  $N2$  不能取 0,而浅层特征作为分支特征,可以通过取 0 验证特征融合的必要性。

表 3 浅层特征和深层特征通道数对模型性能的影响

Tab. 3 Effect of shallow and deep feature channel numbers on the model (%)

$N2 + N1$	PA	mPA	mIoU
64 + 0	88.6	76.9	68.6
64 + 8	88.4	79.2	69.9
64 + 16	88.4	78.7	69.6
64 + 32	88.2	79.0	69.6
64 + 64	88.3	78.9	69.6
16 + 8	88.6	81.5	71.0
32 + 8	88.6	77.6	69.2
128 + 8	88.7	79.3	70.2

从表 3 中前 5 行数据可看出,模型在保持  $N2 = 64$  不变时,在  $N1 = 8$  时,PA, mPA 和 mIoU 这 3 项评价精度相对最优;而当  $N1 = 0$ ,即不融合低层特征时,模型的 mPA 和 mIoU 相对于  $N1 = 8$  时的模型性能都分别下降了 2.3 百分点和 1.3 百分点,这说明了融合低层特征的必要性。后 3 行数据是在维持  $N1 = 8$  时,不断增大  $N2$  得到的实验结果。结果表明,当  $N2 = 16$  时,模型的 3 项评价精度取得最优,相对于  $N2 = 32$  时所得到的评价精度,前者在 mPA 和 mIoU 上高了 3.9 百分点和 1.8 百分点。

实验结果表明,对于 ENet - CSPP 模型,可以通过减少深层、浅层特征通道数来获取较好的预测性能。SAR 图像存在尺度不同的地物目标,对于大的地物目标,像素占比较多,样本数量较多,其特征信息在融合特征中占比较多;小的地物目标,像素占比较少,样本数量较少,其特征信息融合特征中占比

较少。特征通道数代表了当前层特征信息数量的大小,特征信息数量越多,其中包含的冗余信息越多。减少特征通道数,一方面可以减少冗余信息,有利于提高 ENet - CSPP 模型的整体分类精度;另一方面可以减少融合特征中大地物目标的特征信息占比,更利于下一层网络对小地物目标特征的提取,从而提高模型对小地物目标的分类精度。但随着特征通道数的下降,所包含的特征信息可能不足以让下一层网络提取地物目标类别的特征,导致整体分类精度下降。因此,具体特征通道数的数量还需要根据实验结果决定。综上所述,网络模型在  $N2 = 16$  和  $N1 = 8$  时,特征融合层保持了较小的特征输出,提高了较小地物目标的分类精度,使模型预测达到了相对的最优性能。

## 2.4 模型性能对比分析

将 FCN, SegNet, U - Net, DeepLabV3 +, ENet 和 ENet - CSPP 等模型在 GDUT - Nansha 数据集上训练,在同一测试集上预测,各个模型在测试集图像上的地物分类效果如图 4 所示。其中,图 4(a) 测试集图像是 TerraSAR - X 图像对应的伪彩色图像。以  $PA$ ,  $mPA$ ,  $mIoU$  以及每个类别的  $IoU$  作为性能评价指标进行评测(表 4)。其中 ENet - CSPP 模型是在  $N2 = 16$  和  $N1 = 8$  时得到的实验结果;除 ENet 和 ENet - CSPP 模型外,其余 4 个模型都使用了基于 VGG - 16 的编码器和预训练权重;所有模型的损失函数都采用了带权重的交叉熵损失函数,其中权重设置是基于式(6)计算得到,4 种地物目标对应的权重分别如表 2 中所示。

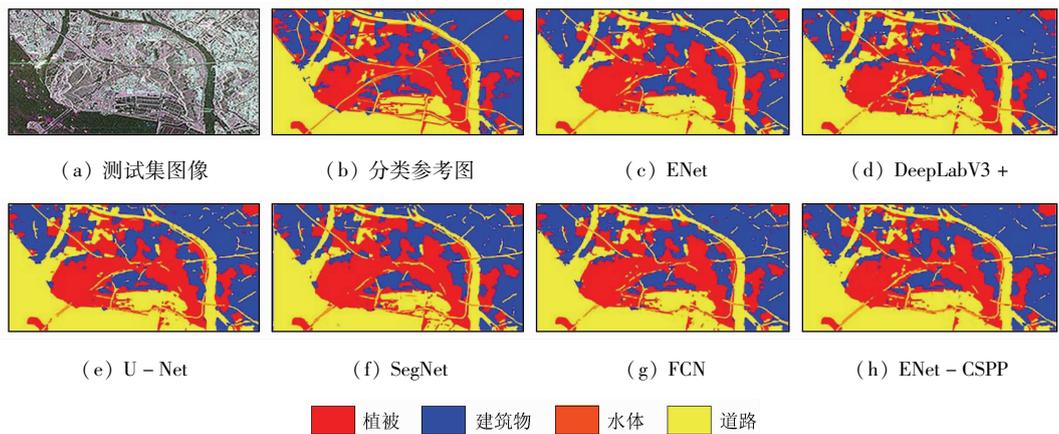


图 4 不同模型在测试集遥感图像上的地物分类结果

Fig. 4 Classification images of the different models on the test image

表 4 不同模型在 GDUT - Nansha 数据集上的对比

Tab. 4 Comparison of different models on GDUT - Nansha data set (%)

模型	$PA$	$mPA$	$IoU$				$mIoU$
			植被	建筑物	水体	道路	
FCN	87.5	78.3	72.8	80.1	89.7	30.3	68.2
SegNet	87.1	79.8	72.9	79.5	88.8	32.9	68.5
U - Net	88.0	80.0	73.3	80.5	90.5	34.1	69.6
DeepLabV3 +	87.1	80.2	70.5	79.7	89.1	37.0	69.1
ENet	85.7	77.7	69.3	76.7	88.8	29.4	66.1
ENet - CSPP	88.6	81.5	74.4	81.8	90.3	37.3	71.0

如表 4 所示,为各个模型在 GDUT - Nansha 数据集上的对比结果。纵向对比,ENet - CSPP 模型比 ENet 模型,在  $PA$ ,  $mPA$  和  $mIoU$  上分别高出了 2.9 百分点、3.8 百分点和 4.9 百分点;关于 4 种地物的  $IoU$  上,ENet - CSPP 模型都有较大程度地提高,特别对于道路的分类,其  $IoU$  比 ENet 高出了 7.9 百分点,这表明 CSPP 模块融合多尺度特征的模式,有效提高了小尺度地物目标的分类精度。横向对比,ENet - CSPP 模型比其余 4 个经典模型中性能最

优的 U - Net 模型,在  $PA$ ,  $mPA$  和  $mIoU$  上分别高出了 0.6 百分点、1.5 百分点和 1.4 百分点;在 4 个类别的  $IoU$  上,ENet - CSPP 模型对水体的分类精度有所降低,但对于其他 3 个地物目标,尤其是在道路地物的分类上,比 U - Net 模型高出 3.2 百分点,表明了 ENet - CSPP 模型的有效性。值得注意的是,DeepLabV3 + 模型对于道路地物的分类,其  $IoU$  也达到了 37.0%,相比于 ENet - CSPP 模型仅降低 0.3 百分点,但对植被地物目标,其  $IoU$  相比于 ENet - CSPP 模型降低了 3.9 百分点;这是因为该模型使用的 ASPP 模块也为多尺度特征融合模块,有效提高了模型对道路地物目标的分类精度;但另一方面,ASPP 中使用空洞卷积作为特征提取手段,会丢失相邻像素的特征,DeepLabV3 + 模型丢失了植被地物目标的重要特征,降低了植被地物目标的分类精度。CSPP 模块将 ASPP 模块中的空洞卷积改为了普通卷积,保留了图像的邻域特征,通过 ENet - CSPP 与 DeepLabV3 + 预测性能的对比,可以看出 CSPP 模块的有效性。

### 3 结论

针对华南地区亚热带气候多云多雾的特点,采用 TerraSAR - X 雷达遥感图像来完成目标区域遥感信息的获取。针对 SAR 图像中地物目标尺度变化较大的问题,基于多卷积核并行提取多尺度特征的策略,提出了 ENet - CSPP 模型。模型的编码器使用了轻量化 ENet 网络作为特征提取网络,同时结合了基于 ASPP 模块改进的 CSPP 模块;解码器部分融合了来自模型浅层网络输出的特征以及深层网络输出的特征,最终输出地物分类图。模型训练中使用了具有合理权重的交叉熵函数作为损失函数,完成了深层、浅层特征通道数对模型性能影响的实验,并利用多个卷积神经网络模型在 GDUT - NanSha 数据集上进行了对比。实验结果表明 ENet - CSPP 模型可以得到更高的分类精度,更有利于监测土地的覆盖情况。通过理论分析和实验验证,初步得到如下结论:

1) 在卷积神经网络中结合多尺度特征融合模块的方法,有利于 SAR 图像地物分类任务,模型分类精度有明显提高。

2) 训练中使用具有合理权重的交叉熵函数作为损失函数,能有效减少样本不平衡问题对模型分类性能的影响,提高模型分类性能。

3) 对于 ENet - CSPP 模型,可以通过减少深层、浅层特征通道数来提高模型对小地物目标类别的分类精度,以获取较好的总体预测性能,但具体特征通道数的数量需要通过实验决定。

后续将考虑研究弱标注样本的生成方法,以扩充数据集的样本,在提高模型性能的同时减少人力标注的投入。

### 参考文献 (References):

- [1] 吕 启, 窦 勇, 牛 新, 等. 基于 DBN 模型的遥感图像分类[J]. 计算机研究与发展, 2014, 51(9): 1911 - 1918.  
Lyu Q, Dou Y, Niu X, et al. Remote sensing image classification based on DBN model[J]. Journal of Computer Research and Development, 2014, 51(9): 1911 - 1918.
- [2] 高照忠, 魏海霞, 黄铁兰. 粤港澳大湾区土地覆盖及景观格局时空变化分析[J]. 测绘通报, 2021(5): 25 - 29.  
Gao Z Z, Wei H X, Huang T L. Analysis of spatio-temporal changes of land cover and landscape pattern in Guangdong - Hong Kong - Macao Greater Bay Area[J]. Bulletin of Surveying and Mapping, 2021(5): 25 - 29.
- [3] van Zyl J J, Kim Y. Synthetic aperture Radar polarimetry[M]. Hoboken: Wiley, 2011: 3.
- [4] Yao W, Marmanis D, Datcu M. Semantic segmentation using deep

neural networks for SAR and optical image pairs[C]//Big Data From Space, 2017: 1 - 4.

- [5] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(4): 640 - 650.
- [6] Henry C, Azimi S M, Merkle N. Road segmentation in SAR satellite images with deep fully convolutional neural networks[J]. IEEE Geoscience and Remote Sensing Letters, 2018, 15(12): 1867 - 1871.
- [7] Wu W J, Li H L, Xin W, et al. PolSAR image semantic segmentation based on deep transfer learning - realizing smooth classification with small training sets[J]. IEEE Geoscience and Remote Sensing Letters, 2019, 16(6): 977 - 981.
- [8] Zhao H S, Shi J P, Qi X J, et al. Pyramid scene parsing network[C]//IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 6230 - 6239.
- [9] Chen L C, Papandreou G, Kokkinos I, et al. Semantic image segmentation with deep convolutional nets and fully connected CRFs[J]. Computer Science, 2014(4): 357 - 361.
- [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] 张 静, 靳淇兆, 王洪振, 等. 多尺度信息融合的遥感图像语义分割模型[J]. 计算机辅助设计与图形学学报, 2019, 31(9): 1510 - 1517.  
Zhang J, Jin Q Z, Wang H Z, et al. Semantic segmentation on remote sensing images with multi-scale feature fusion[J]. Journal of Computer - Aided Design and Computer Graphics, 2019, 31(9): 1510 - 1517.
- [12] Chen L C, Papandreou G, Schroff F. Rethinking atrous convolution for semantic image segmentation[EB/OL]. (2017 - 06 - 17) (2021 - 11 - 17). <https://arxiv.org/abs/1706.05587>.
- [13] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[C]//International Conference on Learning Representations, 2015: 1 - 14.
- [14] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 770 - 778.
- [15] 刘 钊, 赵 桐, 廖斐凡, 等. 基于语义分割网络的高分遥感影像城市建成区提取方法研究与对比分析[J]. 国土资源遥感, 2021, 33(1): 45 - 53. doi: 10. 6046/gtzyyg. 2020162.  
Liu Z, Zhao T, Liao F F, et al. Research and comparative analysis of urban built-up area extraction methods from high-resolution remote sensing images based on semantic segmentation network[J]. Remote Sensing for Land and Resources, 2021, 33(1): 45 - 53. doi: 10. 6046/gtzyyg. 2020162.
- [16] Paszke A, Chaurasia A, Kim S, et al. ENet: A deep neural network architecture for real-time semantic segmentation[EB/OL]. (2016 - 06 - 07) [2021 - 11 - 17]. <https://arxiv.org/abs/1606.02147>.
- [17] 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.

- [18] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation[C]//International Conference on Medical Image Computing and Computer - Assisted Intervention. Berlin: Springer, 2015: 234 - 241.
- [19] Chen L C, Zhu Y, Papandreou G, et al. Encoder - decoder with atrous separable convolution for semantic image segmentation[C]//Computer Vision - European Conference on Computer Vision. Berlin: Springer, 2018: 833 - 851.
- [20] 水文泽, 孙盛, 余旭, 等. 轻量化卷积神经网络在 SAR 图像语义分割中的应用[J]. 计算机应用研究, 2021, 38(5): 1573 - 1580.
- Shui W Z, Sun S, Yu X, et al. Application of lightweight convolutional neural network in SAR image semantic segmentation[J]. Application Research of Computers, 2021, 38(5): 1573 - 1580.
- [21] Werninghaus R, Buckreuss S. The TerraSAR - X mission and system design[J]. IEEE Trans on Geoscience and Remote Sensing, 2009, 48(2): 606 - 614.

## Application of multi - scale and lightweight CNN in SAR image - based surface feature classification

SUN Sheng<sup>1</sup>, MENG Zhimin<sup>1</sup>, HU Zhongwen<sup>2</sup>, YU Xu<sup>3</sup>

(1. School of Computer Science and Technology, Guangdong University of Technology, Guangzhou 510006, China; 2. Key Laboratory for Geo - Environmental Monitoring of Great Bay Area, Ministry of Natural Resources, Shenzhen University, Shenzhen 518000, China; 3. School of Civil and Transportation Engineering, Guangdong University of Technology, Guangzhou 510006, China)

**Abstract:** Targeting the subtropical climate characteristics of the Guangdong - Hong Kong - Macao Greater Bay Area, this study acquired the images of the experimental area from the TerraSAR - X Radar remote sensing satellite. Given the varying scale of the surface feature targets in the Radar satellite observation scenes, this study proposed an ENet convolution spatial pyramid pooling module (ENet - CSPP) model for surface feature classification. Since ordinary convolution can more effectively maintain domain information than atrous convolution, this study proposed a multi - scale feature fusion module based on convolution spatial pyramid pooling. Since there were a few training samples in the SAR remote sensing image dataset, this study combined the multi - scale feature fusion module with the lightweight convolutional neural network. The encoder of the ENet - CSPP network consisted of an improved ENet network and the convolution spatial pyramid pooling module. The decoder output surface feature classification images after the fusion of deep and shallow features. The quantitative comparison experiments were conducted on the GDUT - Nansha dataset. The ENet - CSPP model outperformed other models in three performance indices, namely pixel accuracy, average pixel accuracy, and mean intersection over union. This result indicates that the multi - scale lightweight model effectively improved the accuracy of surface feature classification.

**Keywords:** synthetic aperture Radar (SAR); surface feature classification; convolutional neural network; lightweight network

(责任编辑: 陈理)