

基于 U-Net 和胶囊网络的合成孔径雷达图像语义分割

敬绍迪¹, 喻玲娟^{1*}, 胡跃虹², 杨泽洲¹, 卢忠亮¹, 谢晓春³¹江西理工大学信息工程学院, 江西 赣州 341000;²广州卫富科技开发有限公司, 广东 广州 510200;³赣南师范大学物理与电子信息学院, 江西 赣州 341000

摘要 图像语义分割作为一种像素级分类技术,已应用于合成孔径雷达(SAR)图像的解译领域中。U-Net 是一种端到端的图像语义分割网络,具有典型的编码-解码结构。其中,编码部分主要由卷积层和池化层组成,可以有效提取图像中的目标特征,但难以获取目标的位置和方向等信息。胶囊网络是一种能够获取目标姿态(位置、大小、方向)等信息的神经网络,因此,提出了一种基于 U-Net 和胶囊网络的 SAR 图像语义分割方法。此外,考虑到 SAR 图像数据集较小的特点,将 U-Net 的编码部分设计成视觉几何组(VGG16)结构,将预训练的 VGG16 模型直接迁移至编码部分。为了验证本方法的有效性,在两个极化 SAR 图像数据集上开展了建筑物目标的分割实验。结果表明,相比 U-Net,本方法的精确率、召回率、F1 分数和交并比更高,且能减少网络模型的训练时间。

关键词 图像处理; 合成孔径雷达; 图像语义分割; U-Net; 胶囊网络; 迁移学习

中图分类号 TP753

文献标志码 A

doi: 10.3788/LOP202158.2010009

Semantic Segmentation of Synthetic Aperture Radar Images Based on U-Net and Capsule Network

Jing Shaodi¹, Yu Lingjuan^{1*}, Hu Yuehong², Yang Zezhou¹, Lu Zhongliang¹, Xie Xiaochun³¹ School of Information Engineering, Jiangxi University of Science and Technology, Ganzhou, Jiangxi 341000, China;² Guangzhou Wayful Technology Development Co., Ltd., Guangzhou, Guangdong 510200, China;³ School of Physics and Electronic Information, Gannan Normal University, Ganzhou, Jiangxi 341000, China

Abstract As a pixel-level classification technique, image semantic segmentation has been employed in the field of synthetic aperture radar (SAR) image interpretations. U-Net is an end-to-end image semantic segmentation network with a typical encoder-decoder architecture. Among them, the coding part mainly comprises a convolutional layer and a pooling layer, which can effectively extract the features of a target image; however, extracting information such as the target position and direction is difficult. Capsule network is a type of neural network that can obtain the target pose (position, size, and direction) and other information. Therefore, this study proposes an SAR image semantic segmentation method based on the U-Net and capsule network. Moreover, considering the small data set of SAR images, the U-Net encoder is designed to be identical to the visual geometry group (VGG16) to allow the trained VGG16 model to be directly transferred to the encoder. The effectiveness of the method is verified by conducting a segmentation experiment of building targets on two polarimetric SAR image data sets. Results show that the method can achieve improved precision, recall, F1-score, and intersection over union as well as reduce the training time of the network model when compared with the U-Net.

Key words image processing; synthetic aperture radar; image semantic segmentation; U-Net; capsule network; transfer learning

OCIS codes 100.2960; 200.4260; 280.6730

收稿日期: 2020-11-20; 修回日期: 2020-12-14; 录用日期: 2021-01-02

基金项目: 国家自然科学基金(61501210, 61701203)、江西省教育厅科技项目(GJJ190459)

通信作者: *lingjuanyusmile@163.com

1 引言

合成孔径雷达(SAR)图像的分割一直是 SAR 图像解译领域^[1-4]的研究热点,最早通过图像分割能获得 SAR 图像不同的区域,但不能获取区域中目标的类别信息,因此,图像分割主要作为 SAR 图像目标检测和识别的预处理流程。常用的 SAR 图像分割方法有基于阈值的分割方法、基于边缘和区域的水平集分割方法、基于模糊聚类的分割方法以及基于 Markov 随机场模型的分割方法等^[4-5]。为了获得更好的分割结果,人们提出了基于聚集、匀质和结构区域 3 个高级语义空间的 SAR 图像分割方法^[6],但这类方法仍然无法获取区域中目标类别的语义信息。近年来,随着深度学习的发展,SAR 图像目标识别已取得了显著的成绩。基于深度置信网络^[7-9]、深度堆叠网络^[10]、实数及复数卷积神经网络(CNN)^[11-12]和其他深度学习网络^[13-14]的方法被用于极化 SAR 图像的逐像素分类,这些方法通过对所有像素的分类实现了图像的语义分割,但均不属于端到端的分割方法。

在光学图像语义分割领域,端到端的语义分割模型已得到了人们的深入研究,包括基于对称语义分割模型(全卷积神经网络(FCN)^[15],U-Net^[16],分割网络(SegNet)^[17]、扩张卷积语义分割模型(DeepLab^[18]系列网络)、残差网络语义分割模型^[19]及生成对抗网络(GAN)语义分割模型^[20]。黄刚等^[21]采用 DeepLabv3+网络实现了极化 SAR 图像

的语义分割。陈健堃等^[22]提出了一种复数 SegNet 方法,实现了 SAR 图像中建筑物目标的分割。张金松等^[23]在 U-Net 编解码架构的启发下,提出了一种基于密集深度分离卷积的分割方法,并实现了高分三号 SAR 图像的水域分割。

SAR 图像数据集通常较小,U-Net 的编码部分主要包括卷积层和池化层,虽然能在小样本集下实现快速、有效的图像语义分割,但与 CNN 一样,U-Net 不能有效提取目标的姿态(位置、大小、方向)信息。胶囊网络^[24]是一种由胶囊作为神经元的网络,可以有效解决 CNN 在目标分类中因样本数量少导致的过拟合以及目标姿态变化带来的错误分类问题。目前,该网络已应用于图像分类^[25]、分割^[26]和目标检测^[27]等领域。因此,本文提出了一种基于 U-Net 和胶囊网络的 SAR 图像语义分割方法。首先,为了进一步提高网络的分割效果及训练速度,将 U-Net 的编码部分设计成视觉几何组(VGG16)结构。然后,将在光学图像数据集上训练好的 VGG16 模型迁移至 U-Net 的编码部分。最后,针对 San Francisco Bay 和 Oberpfaffenhofen 极化 SAR 数据集中的建筑物目标,分别开展了图像语义分割实验。

2 U-Net 和胶囊网络的基本理论

2.1 U-Net

文献[16]提出的 U-Net 具有典型的编解码对称结构,如图 1 所示。U-Net 的编码部分也被称为收缩路径,主要包括 4 个模块(每个模块有 2 个卷积

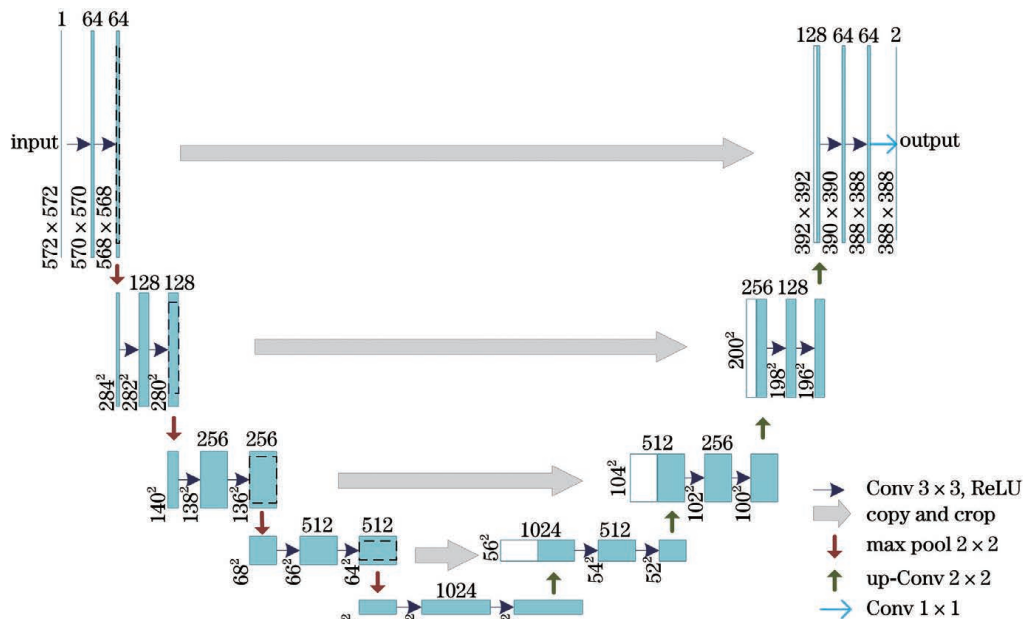


图 1 U-Net 的结构及参数

Fig. 1 Structure and parameters of the U-Net

层和 1 个最大池化层),从第 2 个模块开始,每个模块的第 1 个卷积层会使特征图的通道数加倍,最大池化层会使特征图的尺寸减半。解码部分又被称为扩张路径,其结构与收缩路径对称,也包括 4 个模块(每个模块有 1 个上采样-卷积层和 2 个卷积层)。每个模块的上采样-卷积层能使特征图尺寸加倍且通道数减半,每个上采样-卷积层后,裁剪相应收缩路径获取的特征图,并与上采样-卷积层的结果进行拼接,使特征图的通道数加倍。之后,每个模块的第 1 个卷积层使特征图的通道数减半。其中,卷积核的尺寸为 3×3 ,步长为 1;每个卷积后采用线性整流单元(ReLU)函数激活;池化窗口的尺寸为 2×2 ,步长为 2;上采样-卷积层的上采样倍数为 2×2 ,卷积核尺寸为 2×2 。

在 U-Net 扩张路径最后 1 个卷积层后采用 1 个尺寸为 1×1 的卷积层获取通道数与目标类别数相同的特征图,并采用 Softmax 函数进行逐像素分类。Softmax 函数可表示为

$$p_k(x) = \frac{\exp[a_k(x)]}{\sum_{k'} \exp[a_{k'}(x)]}, \quad (1)$$

式中, $a_k(x)$ 为第 $k(k=1,2)$ 个通道特征图中像素 $x(x=1, \dots, 388 \times 388)$ 处的取值, $k'(k'=1,2)$ 为特

征图的通道编号, $p_k(x)$ 为像素 x 属于第 k 类的概率。U-Net 的训练过程中用交叉熵作为损失函数,可表示为

$$E = \sum_x w(x) \log[p_{l(x)}(x)], \quad (2)$$

式中, $l(x)$ 为像素 x 预测为真实标签 $l(l=1,2)$, $p_{l(x)}(x)$ 为像素 x 预测为真实标签 l 时的概率, $w(x)$ 为不同像素选用的权重值。

2.2 胶囊网络

与 CNN 不同,胶囊网络的神经元是向量。向量中的值能表示目标的姿态(位置、大小、方向)、形变、速度等信息。文献[24]提出的胶囊网络主要包括 2 个卷积层和 1 个全连接层,如图 2 所示。第 1 个卷积层的卷积核尺寸为 9×9 ,数量为 256,步长为 1,卷积后采用 ReLU 函数激活。第 2 个卷积层又被称为初级胶囊(Primary caps)层,卷积核尺寸为 9×9 ,步长为 2。该层输出 32 个尺寸为 6×6 的初级胶囊,每个胶囊为 1×8 维的矢量。全连接层又被称为数字胶囊(Digitcaps)层,每个数字胶囊由前一层所有的初始胶囊动态路由得到。该层输出 10 个数字胶囊,每个胶囊为 1×16 维的矢量。数字胶囊层之后,根据 10 个数字胶囊的 L_2 范数最大值,获取图像所属的类别。

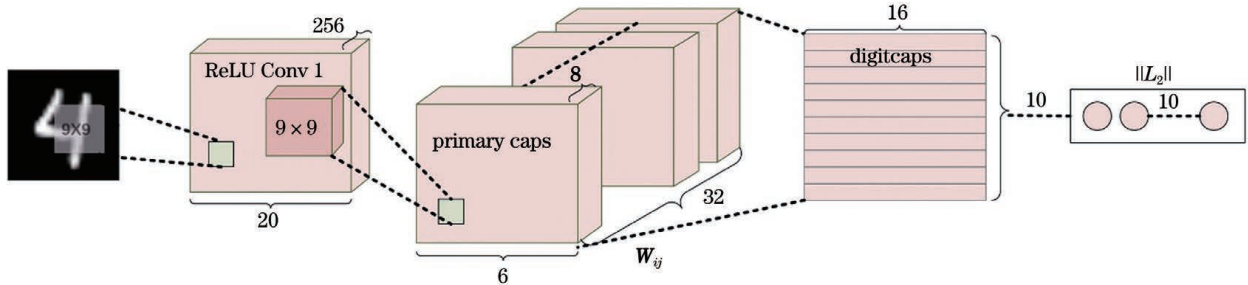


图 2 胶囊网络的结构及参数

Fig. 2 Structure and parameters of the capsule network

胶囊网络中,初级胶囊动态路由到数字胶囊是一个关键步骤。图 2 中,记第 $i(i=1, \dots, 32 \times 6 \times 6)$ 个初级胶囊为 u_i ,第 $j(j=1, \dots, 10)$ 个数字胶囊为 v_j , u_i 经过变换矩阵 W_{ij} (维度为 8×16) 得到预测向量 $\hat{u}_{j|i}$

$$\hat{u}_{j|i} = W_{ij} u_i. \quad (3)$$

令 b_{ij} 为预测向量 $\hat{u}_{j|i}$ 与输出向量 v_j 的相似度,通过 Softmax 函数计算得到的耦合系数 c_{ij} 为

$$c_{ij} = \frac{\exp(b_{ij})}{\sum_d \exp(b_{ij})}, \quad (4)$$

式中, $d(d=1, \dots, 10)$ 为目标的类别。对所有的初级胶囊,根据 c_{ij} 和 $\hat{u}_{j|i}$ 计算加权和 a_j

$$a_j = \sum_i c_{ij} \hat{u}_{j|i}, \quad (5)$$

利用 Squash 函数对加权和进行挤压,得到输出胶囊 v_j

$$v_j = \frac{\|a_j\|^2}{1 + \|a_j\|^2} \cdot \frac{a_j}{\|a_j\|}, \quad (6)$$

利用 v_j 和 $\hat{u}_{j|i}$ 更新 b_{ij} ,可表示为

$$b_{ij} \leftarrow b_{ij} + \hat{u}_{j|i} \cdot v_j. \quad (7)$$

动态路由迭代过程中,重复计算(4)式~(7)式,可依次更新 c_{ij} 、 a_j 、 v_j 及 b_{ij} 。但(3)式中的 W_{ij} 需要根据损失函数进行更新。在胶囊网络的训练过程中,用 Margin 作为损失函数,可表示为

$$L_d = T_d \max(0, m^+ - \|\mathbf{v}_d\|)^2 + \lambda(1 - T_d) \max(0, \|\mathbf{v}_d\| - m^-)^2, \quad (8)$$

式中, T_d 、 m^+ 、 m^- 、 λ 为超参数。当样本中存在类别 d 时, $T_d = 1$; 否则, $T_d = 0$ 。文献[24]中, $m^+ = 0.9$, $m^- = 0.1$, $\lambda = 0.5$ 。

3 基于 U-Net 和胶囊网络的 SAR 图像语义分割

基于 U-Net 和胶囊网络的 SAR 图像语义分割方法采用的网络结构如图 3 所示, 主要包括编码、胶囊网络和解码三部分。其中, 编码部分的结构与 VGG16 的前 14 层相同, 共包含 10 个卷积层和 4 个池化层, 可分成 4 个模块。前 2 个模块均包含 2 个卷积层和 1 个最大池化层; 后 2 个模块均包含 3 个卷积层和 1 个最大池化层。卷积层的卷积核尺寸为 3×3 , 步长为 1; 卷积后采用 ReLU 函数激活, 并进行批归一化(BN)处理; 最大池化层的池化窗口尺寸为 2×2 , 步长为 2。胶囊网络主要包括 3 个卷积层、1 个初级胶囊层和 1 个分割胶囊(Segcaps)层。其

中, 第 1 个卷积层的卷积核尺寸为 5×5 , 最后 2 个卷积层的卷积核尺寸为 1×1 , 步长均为 1; 3 个卷积后均采用 ReLU 函数激活; 初级胶囊层采用的卷积核尺寸为 5×5 , 步长为 1; 分割胶囊层是数字胶囊层的改进, 其与初级胶囊之间为局部连接的动态路由^[26], 卷积核尺寸为 1×1 , 步长为 1。解码部分与编码部分对称, 主要包含 10 个卷积层和 4 个上采样-卷积层, 也可分成 4 个模块。前 2 个模块均包含 1 个上采样-卷积层和 3 个卷积层; 后 2 个模块均包含 1 个上采样-卷积层和 2 个卷积层。每个上采样-卷积层中的上采样倍数为 2×2 , 卷积核尺寸也为 2×2 ; 上采样-卷积运算后, 将得到的特征图与编码部分获取的特征图进行拼接。卷积层的卷积核大小为 3×3 , 步长为 1; 每个卷积后采用 ReLU 函数激活, 并进行 BN 处理。此外, 解码部分的末端还接有 1 个 1×1 卷积层, 并用 Sigmoid 函数激活。

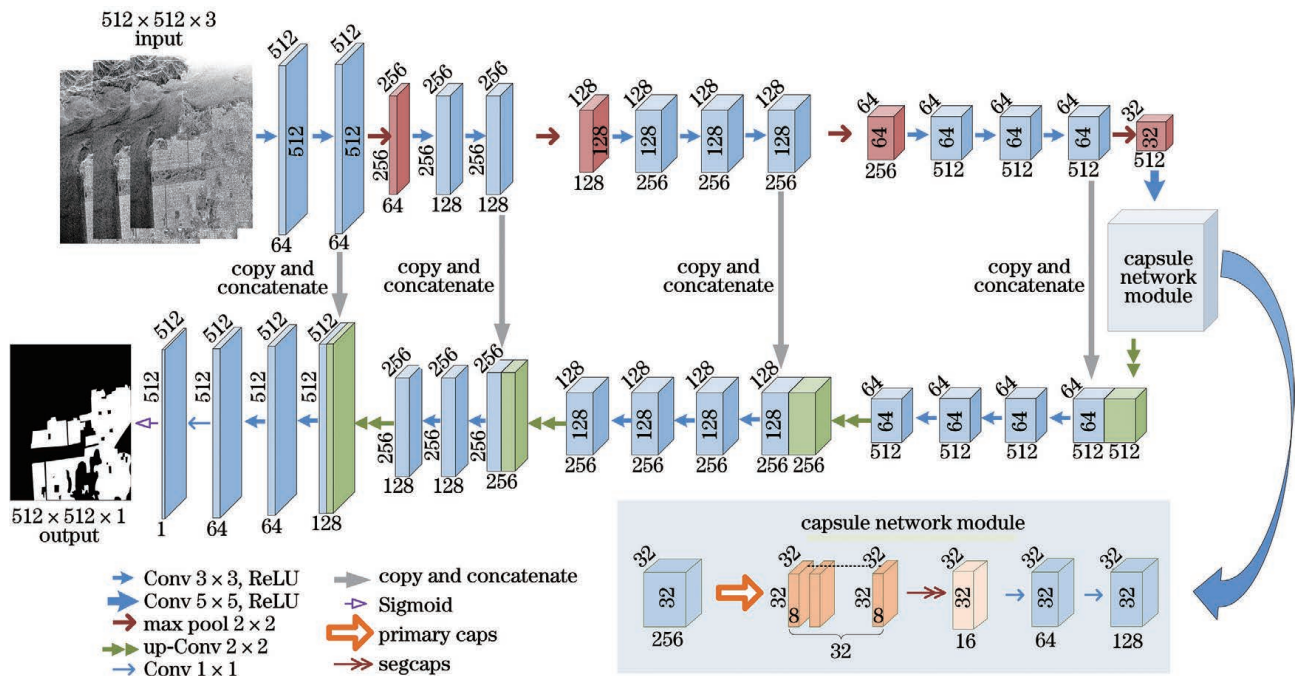


图 3 基于 U-Net 和胶囊网络的 SAR 图像语义分割网络

Fig. 3 Network based on U-Net and capsule network for semantic segmentation of SAR image

基于 U-Net 和胶囊网络的 SAR 图像语义分割步骤: 1) 对 SAR 数据集进行预处理, 获取训练和测试样本; 2) 将光学图像数据集上训练好的 VGG16 模型迁移至编码部分; 3) 用训练样本对网络模型进行训练; 4) 用训练好的网络模型对测试样本进行分割, 得到分割结果。

4 实验结果与分析

4.1 数据集

4.1.1 San Francisco Bay 数据集

实验使用的第一个数据集是 AIRSAR 系统采集的 San Francisco Bay 极化 SAR 数据, 对该数据

进行 Pauli 分解后,可得到尺寸为 900 pixel × 1024 pixel 的 RGB(Red, Green, Blue)三通道图像,如图 4(a)所示。该数据集主要包含水体、植被、建筑物等类别,将建筑物作为分割对象。采用 Labelme 标注工具制作标签,结果如图 4(b)所示。其中,白色区域表示建筑物目标,黑色区域表示背景。由于原始 RGB 图像的尺寸较小,若直接对图像进行裁剪,难以获取足够的训练样本数。因此,对原始图像旋转 90°、180°、270°以及左右和垂直翻转,得到 5 张尺寸为 900 pixel × 1024 pixel 的 RGB 图像。然后,对这 5 张图像进行裁剪,窗口尺寸为 512 pixel × 512 pixel,滑动步长为 100 pixel。裁剪后得到 120 张尺寸为 512 pixel × 512 pixel 的图像,

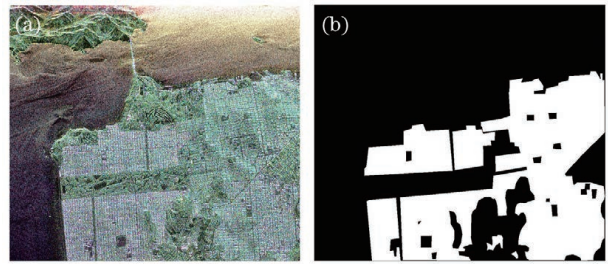


图 4 San Francisco Bay 数据。(a)RGB 图像;(b)标签
Fig. 4 San Francisco Bay data. (a) RGB image; (b) label
将其作为训练样本,裁剪得到的部分训练样本如图 5 所示。用相同方法对原始图像进行裁剪,可得到 24 张尺寸为 512 pixel × 512 pixel 的图像,将其作为测试样本。



图 5 San Francisco Bay 数据集中的部分训练样本

Fig. 5 Part of training samples in the San Francisco Bay data set

4.1.2 Oberpfaffenhofen 数据集

实验使用的第二个数据集是 ESAR 系统采集的 Oberpfaffenhofen 极化 SAR 数据集,对该数据进行 Pauli 分解后可得到尺寸为 1200 pixel × 1300 pixel 的 RGB 图像,如图 6(a)所示。该数据集中主要包含建筑物、植被等类别,同样将其中的建筑物作为分割对象,采用 Labelme 工具制作的标签如图 6(b)所示。与第一个数据集进行相同的预处理,最终得到 280 张训练样本和 56 张测试样本,部分裁剪样本如图 7 所示。

4.2 分割性能评价指标

为了定量评价建筑物目标的分割效果,采用精确率 (Precision)、召回率 (Recall)、F1 分数 (F1-score) 及交并比 (IoU) 作为评价指标,可表示为

$$X_{\text{Precision}} = \frac{T_P}{T_P + F_P}, \quad (9)$$

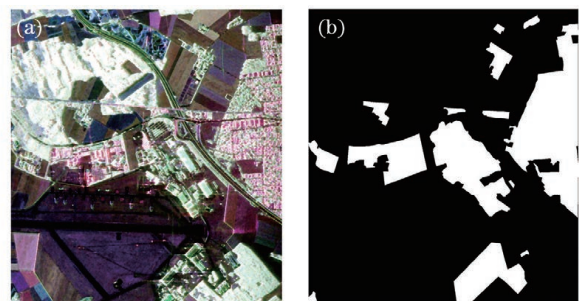


图 6 Oberpfaffenhofen 数据。(a)RGB 图像;(b)标签
Fig. 6 Oberpfaffenhofen data. (a) RGB image; (b) label

$$X_{\text{Recall}} = \frac{T_P}{T_P + F_N}, \quad (10)$$

$$X_{\text{F1-score}} = \frac{2 \times X_{\text{Precision}} \times X_{\text{Recall}}}{X_{\text{Precision}} + X_{\text{Recall}}}, \quad (11)$$

$$X_{\text{IoU}} = \frac{T_P}{T_P + F_P + F_N}, \quad (12)$$

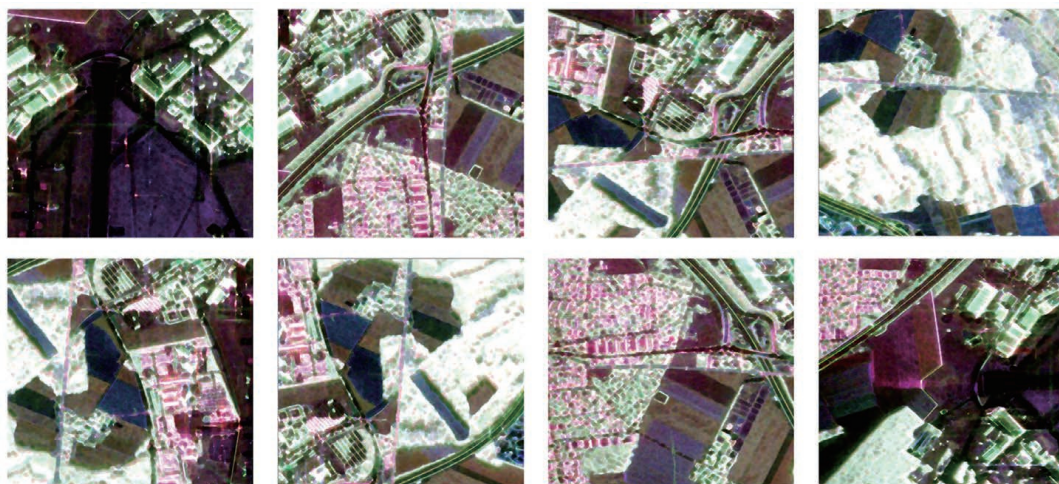


图 7 Oberpfaffenhofen 数据集中的部分训练样本

Fig. 7 Part of training samples in the Oberpfaffenhofen data set

式中, T_P 为标签和分类结果均为正类(建筑物)的像素数, F_P 为标签为负类(背景)但分类结果为正类的像素数, F_N 为标签为正类但分类结果为负类的像素数, T_N 为标签和分类结果均为负类的像素数。

4.3 实验结果及分析

为验证 U-Net 中引入胶囊网络并采用迁移学习的有效性,分别用基于 U-Net 的方法(方法 1)、基于 U-Net 和迁移学习的方法(方法 2)、基于 U-Net 和胶囊网络的方法(方法 3)以及基于 U-Net 和胶囊网络、迁移学习的方法(方法 4)进行对比实验。硬件配置:CPU 为 Intel Xeon Gold 6152;内存为 256 G;GPU 为 NVIDIA 2080Ti。软件环境:操作系统为 Ubuntu;编程语言为 Python;深度学习平台

为 Tensorflow。网络模型训练前,批处理块大小设定为 4,卷积核权重采用正态分布初始化,初始学习率为 0.001。训练过程中,用 Adam 优化器进行参数的更新。

4.3.1 San Francisco Bay 数据集的实验结果

将所有测试样本的分割结果进行拼接,得到不同方法对 San Francisco Bay 数据集中完整建筑物的分割结果如图 8 所示。可以发现,采用迁移学习后,将背景错分为建筑物目标的像素数(F_P)明显减少, F_P 的减少会提高网络的精确率和交并比,并进一步提高 F1 分数;采用胶囊网络后,不仅 F_P 有明显减少,建筑物中被错误分为背景的像素数(F_N)也明显减少,从而提高了网络的精确率、召回率、交并比和 F1 分数。

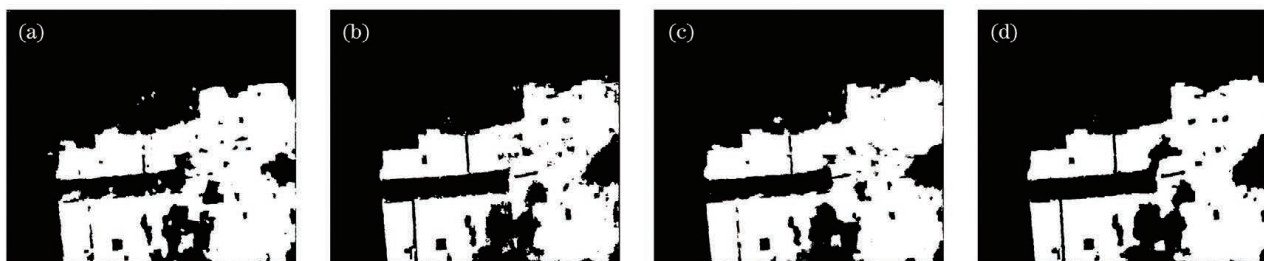


图 8 不同方法在 San Francisco Bay 数据集上的分割结果。(a)方法 1;(b)方法 2;(c)方法 3;(d)方法 4

Fig. 8 Segmentation results of different methods on the San Francisco Bay data set. (a) Method 1; (b) method 2; (c) method 3; (d) method 4

统计不同方法对建筑物和背景的正确、错误分类像素数,得到 4 种方法的分割性能如表 1 所示。可以发现,相比其他 3 种方法,方法 4 具有更高的精确率、召回率、F1 分数以及交并比。表 2 为只使用胶囊网络或迁移学习对网络性能的提升效果,可以发现,相比方法 1,方法 2 的精确率、F1 分

数、交并比分别提高了 6.19%、2.82%、5.31%,召回率下降了 0.64%;相比方法 3,方法 4 的精确率、召回率、F1 分数、交并比分别提高了 5.10%、0.01%、2.58%、4.92%。这表明 U-Net 中采用迁移学习后,网络的精确率、F1 分数、交并比有明显提高,而对召回率的影响较小。相比方法 1,方

表 1 不同方法在 San Francisco Bay 数据集上的分割性能

Table 1 Segmentation performance of different methods on the San Francisco Bay data set unit: %

Method	$X_{\text{Precision}}$	X_{Recall}	$X_{\text{F1-score}}$	X_{IOU}
Method 1	87.46	95.72	91.40	84.17
Method 2	92.87	95.11	93.98	88.64
Method 3	89.31	96.30	92.67	86.35
Method 4	93.86	96.31	95.07	90.60

表 2 分割性能的提升百分比(San Francisco Bay 数据集)

Table 2 Percentage improvement of the segmentation performance (San Francisco Bay data set) unit: %

Method	$X_{\text{Precision}}$	X_{Recall}	$X_{\text{F1-score}}$	X_{IOU}	
Use transfer learning	Method 2 vs. method 1	6.19	-0.64	2.82	5.31
	Method 4 vs. method 3	5.10	0.01	2.58	4.92
Use capsule network	Method 3 vs. method 1	2.12	0.61	1.39	2.59
	Method 4 vs. method 2	1.07	1.26	1.16	2.21

法 3 的精确率、召回率、F1 分数、交并比分别提高了 2.12%、0.61%、1.39%、2.59%；相比方法 2，方法 4 的精确率、召回率、F1 分数、交并比分别提高了 1.07%、1.26%、1.16%、2.21%。这表明在 U-Net 中引入胶囊网络后，4 项性能指标均得到了提高。原因是胶囊网络能获取目标的姿态(位置、大小、方向)等信息，进而使整个网络具有更强的目标特征提取能力。

4 种方法训练过程中，一个 epoch 所需的平均时间如表 3 所示。可以发现，在 U-Net 中引入胶囊网络后，训练时间增加了 1 s；在 U-Net 中采用迁移学习后，训练时间减少了 3 s；在 U-Net 中同时引入

Table 3 Training time of different methods on the San Francisco Bay data set

Method	Method 1	Method 2	Method 3	Method 4
Training time /s	16	13	17	14

胶囊网络和迁移学习后，训练时间减少了 2 s，这表明迁移学习能有效减少网络模型的训练时间。

4.3.2 Oberpfaffenhofen 数据集的实验结果及分析

将所有测试样本的分割结果进行拼接，得到不同方法对 Oberpfaffenhofen 数据集中完整建筑物的分割结果如图 9 所示。可以发现，采用迁移学习后，网络的 F_p 明显减少，精确率、F1 分数、交并比得到了一定程度的提高。引入胶囊网络后，网络的 F_p 和 F_N 均减少，精确率、召回率、F1 分数、交并比均得到了提高。统计不同方法对建筑物和背景的正确、错误分类像素数，得到 4 种方法的分割性能如表 4 所示。可以发现，与表 1 的结论一致，方法 4 的各项性能指标均最优。根据表 4 中的数据计算只使用胶囊网络或迁移学习对网络性能的提升效果，结果如表 5 所示。可以发现，采用迁移学习后，网络的精确率、F1 分数、交并比均有明显提高，而召回率的

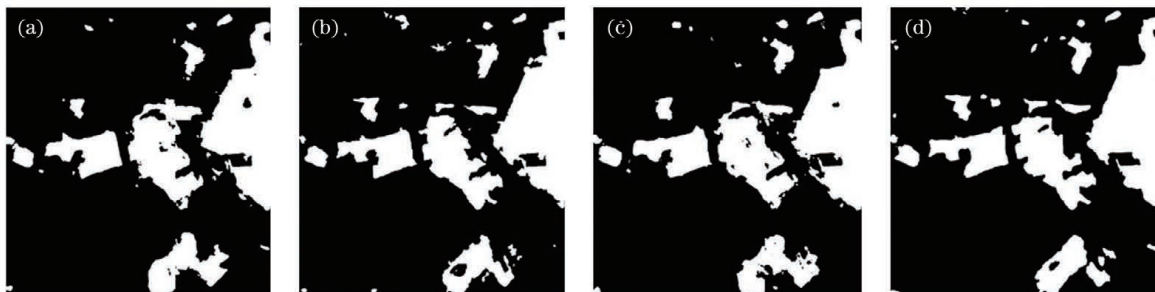


图 9 不同方法在 Oberpfaffenhofen 数据集上的分割结果。(a)方法 1；(b)方法 2；(c)方法 3；(d)方法 4

Fig. 9 Segmentation results of different methods on the Oberpfaffenhofen data set. (a) Method 1; (b) method 2; (c) method 3; (d) method 4

表 4 不同方法在 Oberpfaffenhofen 数据集上的分割性能

Table 4 Segmentation performance of different methods on the Oberpfaffenhofen data set unit: %

Method	$X_{\text{Precision}}$	X_{Recall}	$X_{\text{F1-score}}$	X_{IOU}
Method 1	87.60	91.50	89.51	81.01
Method 2	90.89	91.53	91.21	83.84
Method 3	89.16	93.07	91.07	83.61
Method 4	92.06	93.18	92.62	86.24

表 5 分割性能的提升百分比(Oberpfaffenhofen 数据集)

Table 5 Percentage improvement of the segmentation performance (Oberpfaffenhofen data set) unit: %

Method	$X_{\text{Precision}}$	X_{Recall}	$X_{\text{F1-score}}$	X_{IOU}	
Use transfer learning	Method 2 vs. method 1	3.76	0.03	1.90	3.49
	Method 4 vs. method 3	3.25	0.12	1.70	3.15
Use capsule network	Method 3 vs. method 1	1.78	1.72	1.74	3.21
	Method 4 vs. method 2	1.29	1.80	1.55	2.86

提高较小。引入胶囊网络后,网络的 4 项性能指标均得到了一定的提高。4 种方法的训练过程中,一个 epoch 所需的平均时间如表 6 所示。综上所述,基于 U-Net 和胶囊网络并采用迁移学习的方法不仅能有效减少网络模型的训练时间,且能获得更好的分割性能。

表 6 不同方法在 Oberpfaffenhofen 数据集上的训练时间

Table 6 Training time of different methods on the Oberpfaffenhofen data set

Method	Method 1	Method 2	Method 3	Method 4
Training time /s	37	29	40	32

5 结 论

首先介绍了 U-Net 和胶囊网络的基本理论。然后,提出了一种基于 U-Net 和胶囊网络的 SAR 图像语义分割方法。该方法采用的网络结构包括编码、胶囊网络和解码三部分,其中,编码部分为 VGG16 结构,可在网络训练时将已训练好的 VGG16 模型直接迁移至编码部分。针对 San Francisco Bay 和 Oberpfaffenhofen 两个极化 SAR 数据集分别开展了建筑物分割实验,结果表明,相比其他方法,本方法不仅能减少网络模型的训练时间,且能获得更高的精确率、召回率、F1 分数以及交并比。

参 考 文 献

[1] Zhao Y F, Zhang B H, Zhang Y Y, et al. Ship

detection based on SAR images using deep feature pyramid and cascade detector[J]. Laser & Optoelectronics Progress, 2020, 57(12): 121019.

赵云飞, 张宝华, 张艳月, 等. 基于深度特征金字塔和级联检测器的 SAR 图像舰船检测[J]. 激光与光电子学进展, 2020, 57(12): 121019.

[2] Liu X W, Lei J C, Wu Y P. Synthetic aperture radar target-recognition method based on bidimensional empirical mode decomposition[J]. Laser & Optoelectronics Progress, 2020, 57(4): 041004.

柳小文, 雷军程, 伍雁鹏. 基于二维经验模态分解的合成孔径雷达目标识别方法[J]. 激光与光电子学进展, 2020, 57(4): 041004.

[3] Zhang H, Zuo X L, Huang Y. Feature selection based on the correlation of sparse coefficient vectors with application to SAR target recognition[J]. Laser & Optoelectronics Progress, 2020, 57(14): 141029.

张虹, 左鑫兰, 黄瑶. 基于稀疏表示系数相关性的特征选择及 SAR 目标识别方法[J]. 激光与光电子学进展, 2020, 57(14): 141029.

[4] Wan L, You H J, Cheng Y B, et al. Research progress of synthetic aperture radar image segmentation[J]. Remote Sensing Technology and Application, 2018, 33(1): 10-24.

万玲, 尤红建, 程跃兵, 等. 合成孔径雷达图像分割研究进展[J]. 遥感技术与应用, 2018, 33(1): 10-24.

[5] Li N, Niu S L. 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.

- 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.
- [6] Gao M Y. SAR image segmentation based on semantic and ridgelet deconvolutional network[D]. Xi'an: Xidian University, 2016.
高梦瑶. 基于语义和脊波反卷积网络的 SAR 图像分割[D]. 西安: 西安电子科技大学, 2016.
- [7] Lü Q, Dou Y, Niu X, et al. Urban land use and land cover classification using remotely sensed SAR data through deep belief networks[J]. Journal of Sensors, 2015, 2015: 1-10.
- [8] Guo Y H, Wang S, Gao C Q, et al. Wishart RBM based DBN for polarimetric synthetic radar data classification[C]//2015 IEEE International Geoscience and Remote Sensing Symposium(IGARSS), July 26-31, 2015, Milan, Italy. New York: IEEE Press, 2015: 1841-1844.
- [9] Liu F, Jiao L C, Hou B, et al. POL-SAR image classification based on wishart DBN and local spatial information[J]. IEEE Transactions on Geoscience and Remote Sensing, 2016, 54(6): 3292-3308.
- [10] Jiao L C, Liu F. Wishart deep stacking network for fast POLSAR image classification[J]. IEEE Transactions on Image Processing, 2016, 25(7): 3273-3286.
- [11] Zhou Y, Wang H P, Xu F, et al. Polarimetric SAR image classification using deep convolutional neural networks[J]. IEEE Geoscience and Remote Sensing Letters, 2016, 13(12): 1935-1939.
- [12] Zhang Z M, Wang H P, Xu F, et al. Complex-valued convolutional neural network and its application in polarimetric SAR image classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2017, 55(12): 7177-7188.
- [13] Bi H X, Xu F, Wei Z Q, et al. An active deep learning approach for minimally supervised PolSAR image classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57(11): 9378-9395.
- [14] Hu T, Li W H, Qin X X. Semantic segmentation of polarimetric synthetic aperture radar images based on multi-layer deep feature fusion[J]. Chinese Journal of Lasers, 2019, 46(2): 0210001.
胡涛, 李卫华, 秦先祥. 基于多层深度特征融合的极化合成孔径雷达图像语义分割[J]. 中国激光, 2019, 46(2): 0210001.
- [15] 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.
- [16] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation[M]//Navab N, Hornegger J, Wells W M, et al. Medical image computing and computer-assisted intervention-MICCAI 2015. Lecture notes in computer science. Cham: Springer, 2015, 9351: 234-241.
- [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] Chen L C, Zhu Y K, Papandreou G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11211: 833-851.
- [19] Alom M Z, Yakopcic C, Hasan M, et al. Recurrent residual U-Net for medical image segmentation[J]. Journal of Medical Imaging, 2019, 6(1): 014006.
- [20] Zhu X B, Zhang X M, Zhang X Y, et al. A novel framework for semantic segmentation with generative adversarial network[J]. Journal of Visual Communication and Image Representation, 2019, 58: 532-543.
- [21] Huang G, Liu X L. Semantic segmentation of polarimetric synthetic aperture radar images based on deep learning[J]. Science of Surveying and Mapping, 2019, 44(6): 166-169, 192.
黄刚, 刘先林. 深度学习的极化合成孔径雷达影像语义分割[J]. 测绘科学, 2019, 44(6): 166-169, 192.
- [22] Chen J K, Peng L X, Qiu X L, et al. A 3D building reconstruction method for SAR images based on deep neural network[J]. Scientia Sinica (Informationis), 2019, 49(12): 1606-1625.
陈健堃, 彭凌霄, 仇晓兰, 等. 基于深度神经网络的 SAR 建筑目标三维重建方法[J]. 中国科学: 信息科学, 2019, 49(12): 1606-1625.
- [23] Zhang J S, Xing M D, Sun G C. A water segmentation algorithm for SAR image based on dense depthwise separable convolution[J]. Journal of Radars, 2019, 8(3): 400-412.
张金松, 邢孟道, 孙光才. 一种基于密集深度分离卷积的 SAR 图像水域分割算法[J]. 雷达学报, 2019,

- 8(3): 400-412.
- [24] Sabour S, Frosst N, Hinton G E. Dynamic routing between capsules [C] // Proceedings of in the 31st Conference on Neural Information Processing Systems, December 4-9, 2017, Long Beach, CA, USA. New York: Curran Associates Inc, 2017: 3859-3869.
- [25] Zhang W, Tang P, Zhao L J. Remote sensing image scene classification using CNN-CapsNet[J]. Remote Sensing, 2019, 11(5): 494.
- [26] LaLonde R, Bagci U. Capsules for object segmentation[EB/OL]. (2018-04-11)[2020-11-15]. <http://arxiv.org/abs/1804.04241v1>.
- [27] Mekhalfi M L, Bejiga M B, Soresina D, et al. Capsule networks for object detection in UAV imagery[J]. Remote Sensing, 2019, 11(14): 1694.