

中图法分类号: TP301.6 文献标识码: A 文章编号: 1006-8961(2021)12-3015-12

论文引用格式: Li X W, Li Y S and Zhang Y J. 2021. Weakly supervised deep semantic segmentation network for water body extraction based on multi-source remote sensing imagery. Journal of Image and Graphics, 26(12): 3015-3026 (李鑫伟, 李彦胜, 张永军. 2021. 弱监督深度语义分割网络的多源遥感影像水体检测. 中国图象图形学报, 26(12): 3015-3026) [DOI: 10.11834/jig.200192]

弱监督深度语义分割网络的多源遥感影像水体检测

李鑫伟, 李彦胜*, 张永军

武汉大学遥感信息工程学院, 武汉 430079

摘要: **目的** 深度语义分割网络的优良性能高度依赖于大规模和高质量的像素级标签数据。在现实任务中, 收集大规模、高质量的像素级水体标签数据将耗费大量人力物力。为了减少标注工作量, 本文提出使用已有的公开水体覆盖产品来创建遥感影像对应的水体标签, 然而已有的公开水体覆盖产品的空间分辨率低且存在一定错误。对此, 提出采用弱监督深度学习训练深度语义分割网络。**方法** 在训练阶段, 将原始数据集划分为多个互不重叠的子数据集, 分别训练深度语义分割网络, 并将训练得到的多个深度语义分割网络协同更新标签, 然后利用更新后的标签重复前述过程, 重新训练深度语义分割网络, 多次迭代后可以获得好的深度语义分割网络。在测试阶段, 多源遥感影像经多个代表不同视角的深度语义分割网络分别预测, 然后投票产生最后的水体检测结果。**结果** 为了验证本文方法的有效性, 基于原始多源遥感影像数据创建了一个面向水体检测的多源遥感影像数据集, 并与基于传统的水体指数阈值分割法和基于低质量水体标签直接学习的深度语义分割网络进行比较, 交并比 (intersection-over-union, IoU) 分别提升了 5.5% 和 7.2%。**结论** 实验结果表明, 本文方法具有收敛性, 并且光学影像和合成孔径雷达 (synthetic aperture radar, SAR) 影像的融合有助于提高水体检测性能。在使用分辨率低、噪声多的水体标签进行训练的情况下, 训练所得多视角模型的水体检测精度明显优于基于传统的水体指数阈值分割法和基于低质量水体标签直接学习的深度语义分割网络。

关键词: 水体检测; 多源遥感影像; 低分辨率噪声标签; 弱监督深度语义分割网络

Weakly supervised deep semantic segmentation network for water body extraction based on multi-source remote sensing imagery

Li Xinwei, Li Yansheng*, Zhang Yongjun

School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China

Abstract: **Objective** Water body detection has shown important applications in flood disaster assessment, water resource value estimation and ecological environment protection based on remote sensing imagery. Deep semantic segmentation network has achieved great success in the pixel-level remote sensing image classification. Water body detection performance can be reasonably expected based on the deep semantic segmentation network. However, the excellent performance of deep semantic segmentation network is highly dependent on the large-scale and high-quality pixel-level labels. This research paper has intended to leverage the existing open water cover products to create water labels corresponding to remote sensing images in order to reduce the workload of labeling and meantime maintain the fair detection accuracy. The existing open water cover products have a low spatial resolution and contain a certain degree of errors. The noisy low-resolution water

收稿日期: 2020-06-05; 修回日期: 2020-10-28; 预印本日期: 2020-11-04

* 通信作者: 李彦胜 yansheng.li@whu.edu.cn

基金项目: 国家自然科学基金项目 (42030102, 41971284)

Supported by: National Natural Science Foundation of China (42030102, 41971284)

labels have inevitably affected the training of deep semantic segmentation network for water body detection. A weakly supervised deep learning method to train deep semantic segmentation network have been taken into consideration to resolve the difficulties. The optimization method to train deep semantic segmentation network using the noisy low-resolution labels for the high accuracy of water detection has been presented based on minimizing the manual annotation cost. **Method** In the training stage, the original dataset has been divided into several non-overlapped sub-datasets. The deep semantic segmentation network has been trained on each sub-dataset. The trained deep semantic segmentation networks with different sub-datasets have updated the labels simultaneously. As the non-overlapped sub-datasets generally have different data distributions, the detection performance of different networks with different sub-datasets is also complementary. The prediction of the same region by different networks is different, so the multi-perspective deep semantic segmentation network can realize the collaborative update of labels. The updated labels have been used to repeat the above process to re-train new deep semantic segmentation networks. Following each step of iteration, the output of the network has been used as the new labels. The noisy labels have been removed with the iteration process. The range of truth value of the water has also be expanded continuously along with the iteration process. Several good deep semantic segmentation networks can be obtained after a few iterations. In the test stage, the multi-source remote sensing images have been predicted by several deep semantic segmentation networks representing different perspectives and producing the final water detection voting results. **Result** The multi-source remote sensing image training dataset, validation dataset and testing dataset have been built up for verification. The multi-source remote sensing imagery has composed of Sentinel-1 SAR (synthetic aperture radar) images and Sentinel-2 optical images. The training dataset has contained 150 000 multi-source remote sensing samples with the size of 256×256 pixels. The labels of the training dataset have been intercepted with the public MODIS (moderate-resolution imaging spectroradiometer) water coverage products in geographic scale. The spatial resolution of the training dataset is low and contains massive noise. The validation dataset has contained 100 samples with the size of 256×256 pixels and the testing dataset have contained 400 samples with the size of 256×256 pixels, and the labels from the validation and testing datasets have accurately annotated with the aid of domain experts. The training, validation and testing datasets have not been overlapped each and the dataset can geographically cover in global scale. Experimental results have shown that the proposed method is convergent, and the accuracy tends to be stable based on four iterations. The fusion of optical and SAR images can improve the accuracy of water body detection. The IoU (intersection over union) has increased by 5.5% compared with the traditional water index segmentation method. The IoU has increases by 7.2% compared with the deep semantic segmentation network directly using the noisy low-resolution water labels. **Conclusion** The experimental results have shown that the current method can converge fast, and the fusion of optical and SAR images can improve the detection results. On the premise of the usage of the noisy low-resolution water labels, the water body detection accuracy of the trained multi-perspective model is obviously better than the traditional water index segmentation method and the deep semantic segmentation network based on the direct learning of the noisy low-resolution water labels. The accuracy of the traditional deep semantic segmentation method is slightly lower than that of the traditional water index method, which indicates that the effectiveness of deep learning highly depends on the quality of the training data labels. The noisy low-resolution water labels have reduced the effect of deep learning. The effect of the proposed method on small rivers and lakes has been analyzed. The accuracy on small rivers and lakes has decreased slightly. The result has still higher than the traditional water index method and the deep learning method with the direct training of the noisy low-resolution water labels.

Key words: water body detection; multi-source remote sensing image; noisy low-resolution labels; weakly supervised deep semantic segmentation network

0 引言

遥感影像的水体检测在洪涝灾害评估、水资源价值估算和生态环境保护等方面都具有重要意义。

现代社会对环境日益重视,人们对水体信息提取的要求越来越严格。而遥感影像成像周期短,成像精度日益提高,如何提高高分辨率影像的水体检测精度,成为关注的重点。

传统的遥感影像水体检测方法,主要分为基于

影像光谱特征的提取方法和分类器方法。基于影像光谱特征的方法对多光谱的波段进行组合,是对单个像素的不同波段值进行组合计算,包括单波段法和多波段法,其中多波段法又包括谱间关系法和水体指数法。单波段法主要利用水体在近红外或中红外波段强烈吸收的原理识别水体信息(杨莹和阮仁宗,2010;Lu等,2011),仅使用单一波段进行阈值分割,简单易行,但是精度受限。多波段法是综合遥感影像不同的波段,利用水体波段间的关系特征识别水体的一种水体检测方法。其中,谱间关系法是依据各个特征波段绘制而成的光谱特征曲线对水体进行识别和检测。更为常见的是水体指数法。Mcfeeters(1996)首先提出了归一化差异水体指数(normalized difference water index, NDWI)用于描述水体的特征,利用反射的近红外辐射和可见的绿光来增强这些特征,同时消除土壤和陆地植被特征的存在。徐涵秋(2005)提出了改进的归一化差异水体指数(modified normalized difference water Index, MNDWI),对水体指数使用的波段组合进行了修改,获得了比NDWI更好的效果。丁凤(2009)根据水体在近红外和中红外波段同时具有强烈吸收这一典型特征,提出了一种增强型的水体指数(enhanced water index, EWI)。莫伟华等人(2007)将植被指数与近红外波段进一步组合,提出了新的水体指数模型(combined index of ndvi and nir for water body identification, CIWI),实现了水体信息的有效提取。但是基于光谱特征的提取方法仍集中在遥感影像各波段的波谱特征上,对影像纹理和空间等其他特征关注较少,严重制约了水体提取的精度。基于分类器的方法主要按照某种算法规则进行图像划分,包括支持向量机、决策树和面向对象等方法。段秋亚等人(2015)分别采用归一化差异水体指数、支持向量机和面向对象方法进行水体信息提取实验,结果发现支持向量机精度最高。殷亚秋等人(2015)根据高分辨率遥感影像的特点,利用面向对象的方法对高分辨率遥感影像进行了水体提取。陈静波等人(2013)构建了城市水体提取知识规则及知识决策树模型。传统的遥感影像水体检测方法大多只能在特定的影像条件或者局部区域得到好的分类结果,当面对多源影像、复杂多样的水体光谱特性和几何形态时,采用传统的遥感影像水体检测方法难以取得较好的检测结果(廖安平 等,2014)。

深度学习技术的巨大成功极大改善了遥感影像场景分类(Li等,2016)、遥感影像检索(Li等,2018a,b)和遥感影像目标识别(Tan等,2018;Li等,2018c)等相关技术的性能。因此,合理预期水体检测同样可以基于深度学习实现水体精准的检测。何海清等人(2017)利用传统的归一化差异水体指数(NDWI)与深度学习进行融合的方法进行水体检测,利用归一化差异水体指数将图像分割成多个子区域作为水体检测的基本单元,深度卷积网络迭代地提取水体信息。相比于传统的水体指数阈值法和分类器法,精度有较为明显的优势。王雪等人(2018)提出了使用全卷积的神经网络进行水体检测的方法,验证了相比于传统的阈值方法和图方法,全卷积网络可以获得更好的泛化能力和实验精度。陈前等人(2019)使用深度学习的方法,在高分辨率遥感影像中验证了深度网络可以有效去除阴影和建筑物对水体检测的影响,提高了水体检测的精度。

基于深度学习的水体检测方法取得理想精度的前提是具备高质量和大规模的水体检测数据集。但是,由于标注的成本较高,很难获取符合质量要求的水体标签。在遥感领域,基于统一的地理坐标基准,可以借助公开的水体覆盖产品为遥感影像创建水体标签,然而公开的水体覆盖产品往往分辨率较低且含有一定的噪声,如波士顿大学利用MODIS(moderate-resolution imaging spectroradiometer)提取的1 km分辨率的水体栅格数据(Friedl等,2002),马里兰大学2005年生成的全球250 m陆表水体栅格数据(Carroll等,2009)。将分辨率低或者错误较多的水体标签直接用于深度网络的训练,会对水体检测的精度造成一定程度上的影响(Li等,2021)。

弱监督深度学习的发展为遥感影像水体检测提供了新的解决方案。当前,弱监督深度学习技术是一项重要研究课题(Guo等,2017)。相比于全监督的深度学习,它使用的监督信息并不完全,即不需要所有样本都具备标签或者标签并不一定准确。语义分割中,弱监督信息包括图像级标签监督信息(Pinheiro和Collobert,2015;Zhou等,2016)、点标签监督信息(Bearman等,2016)、包围框监督信息(Dai等,2015;Khoreva等,2017)、涂鸦标签监督信息(Lin等,2016)和半监督信息(Hung等,2018;Souly等,2017),并且分辨率低或噪声多(Reed等,2015;

Sukhbaatar 等,2015) 的标签也可以视为弱监督信息。针对遥感领域缺乏高质量、高分辨率的水体数据集问题,可以将弱监督深度学习技术应用于水体检测,在保证精度的前提下,极大地节约水体标签标注的成本,缩短训练数据生成的周期。近年来,在计算机视觉领域提出了大量弱监督深度学习的思路,有使用鲁棒的损失函数(Ghosh 等,2017)、通过迭代训练提高精度和修改网络结构等方法,在自然影像数据集上都取得了较为理想的结果。

但是,遥感影像不同于自然影像,具有目标尺度多变、场景复杂和拍摄视角灵活多变等特点,计算机视觉领域的弱监督深度学习方法往往无法在遥感领域取得预期的效果。针对上述问题,也有一些遥感领域的弱监督方法(Li 等,2020)被提出。

本文提出了一种基于弱监督深度语义分割网络的多源遥感影像水体检测方法。在训练阶段采取弱监督训练策略,将数据集分割为多个互不重叠的子数据集,在子数据集上训练代表多个视角的不同深度语义分割网络模型,利用训练得到的多个深度语义分割网络模型协同优化标签,经过多个深度语义分割网络协同优化后的水体标签噪声大幅度减少。尽管优化后的水体标签噪声减少,但是单次优化的结果无法满足高精度水体检测的要求,因此,用优化后的标签训练多个新的深度语义分割网络模型,多次迭代,以期获得最优的深度语义分割网络模型。在测试阶段,一幅多源遥感影像经多个代表不同视角的深度语义分割网络分别预测,然后进行投票产生最后的水体检测结果。

为了验证本文方法的有效性,使用一个多源遥感影像水体训练及测试数据集进行实验。多源遥感影像 Sentinel-1 号合成孔径雷达(synthetic aperture radar, SAR)影像和 Sentinel-2 号光学影像构成。结果表明,在水体标签分辨率低、噪声较多的情况下,本文提出的水体检测方法的精度优于传统的水体指数阈值法。并且相比于直接用分辨率低、噪声较多的水体标签训练深度语义分割网络的水体检测方法,水体检测精度有较大提升。

1 基于公开水体覆盖产品的多源遥感影像水体检测数据集

基于原始多源遥感影像数据(Schmitt 等,

2020),创建了一个面向水体检测的多源遥感影像数据集。其中,训练集含有 15 万个大小为 256×256 像素的多源遥感影像样本。训练集的标签借助公开的 MODIS 水体覆盖产品通过地理范围对应截取,原始分辨率为 500 m,将其重采样至 10 m 分辨率。由于训练集的标签是由低分辨率采样至高分辨率,该标签不够精细,噪声较多。标签包含水体和非水体两类。验证集包含 100 个大小为 256×256 像素的多源遥感影像样本,测试集包含 400 个大小为 256×256 像素的多源遥感影像样本,其标签均通过人工精确标注获得,标签较为精细,可以满足模型精度评估的需要。每一个多源遥感影像样本中,影像的分辨率为 10 m,包含 Sentinel-1 号 SAR 影像两个极化波段(VV 和 VH 两种极化方式);Sentinel-2 号光学影像包括 13 个波段,每个波段都重采样至 10 m 分辨率。训练集的原始数据包含 216 596 组样本,原始数据的影像分辨率、波段参数与训练集一致。按比例随机将原始数据划分为训练集、验证集和测试集。由于划分是随机的,所以在每一个采样区域,训练样本、验证样本和测试样本同时存在,三者的覆盖区域几乎相同。图 1 为原始数据地理分布示意图。可以看出,该数据集地理分布较为全面,可以覆盖所有的大洲。图 2 为训练集、验证集及测试的标签示意图,可以看到相比于人工精确标注的标签,训练集的标签分辨率低并且存在较多错误,而人工标注的标签可以作为精度评定的参考。另外,为了对本文关注的水体类型进行约束,对测试集及验证集中最细河流的宽度、最小水体的面积进行了测算,最细河流宽度为 60 m,最小水体的面积为 $9\ 700\text{ m}^2$ 。

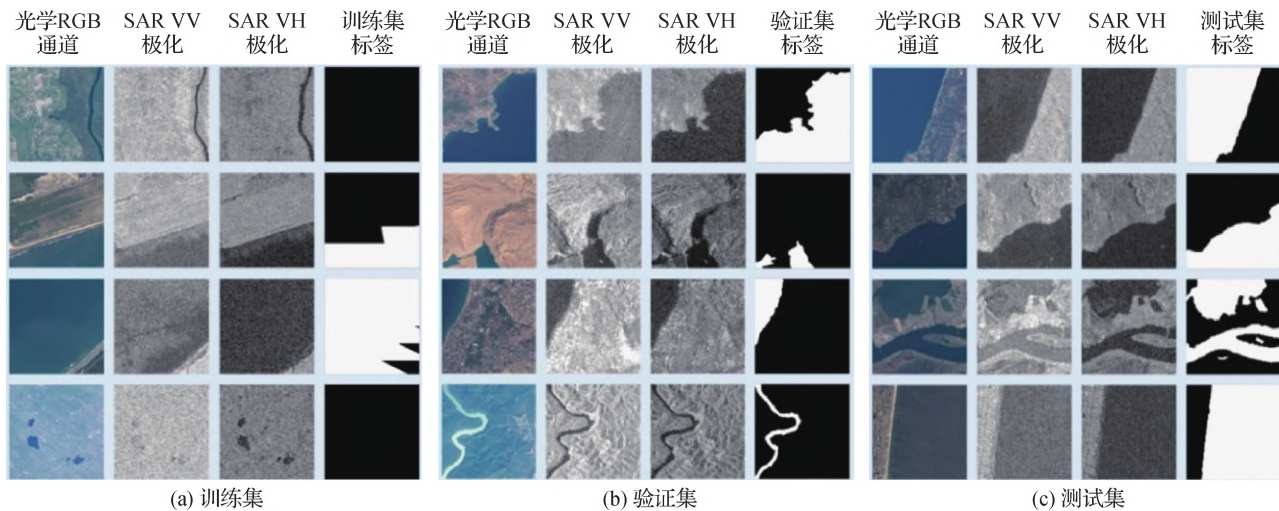
2 基于弱监督深度语义分割网络的多源遥感影像水体检测

提出的基于弱监督深度语义分割网络的多源遥感影像水体检测方法包括训练阶段和测试阶段。训练阶段将数据集划分为互不重叠的多个子数据集,在此基础上训练代表不同视角的深度语义分割网络,迭代优化水体标签,得到多个深度语义分割网络模型。测试阶段用训练得到的多个深度语义分割网络模型投票预测,最终得到水体检测结果。



图1 数据地理分布图

Fig. 1 Geographical distribution of data



(a) 训练集

(b) 验证集

(c) 测试集

图2 数据集标签示意图

Fig. 2 Visual labels of datasets((a) training set;(b) validation set;(c) testing set)

2.1 深度语义分割网络的交叉迭代学习

训练阶段流程图如图3所示。原数据集 $S = \{(I^k, L^k) | k = 1, 2, \dots, K\}$, 其中, K 表示样本总个数, k 表示样本的序号, I^k 为多源遥感影像, L^k 为多源影像的标签。随机将原数据集 S 划分为 N 个互不重叠的子数据集 $S_n(t)$, 其中, n 表示在同一次迭代中子数据集的序号 ($n = 1, \dots, N$), 并且 $S_1(t) \cap \dots \cap S_N(t) = \emptyset$, $S_1(t) \cup \dots \cup S_N(t) = S$, t 表示迭代次数。在每个子数据集 $S_n(t)$ 上, 训练一个深度语义分割网络模型 $C_n(t)$ 。这里网络模型的具体结构并没有限制, 因此该方法可以适用于多种网络结构。由于互不重叠的子数据集可能具有不同

的数据分布, 因此, 训练得到的网络映射关系也不尽相同。同一区域, 不同模型对该区域的预测不相同, 从而可以实现多视角深度语义分割网络互相纠错、协同优化标签。用训练得到的网络模型 $C_n(t)$ 分别对 N 个子数据集进行预测, 得到预测图 $P_{ij}(t)$ (i 表示模型序号, j 表示子数据集序号), 即 $i \times j$ 组输出图。预测结果 $W_{ij}(t)$ 由网络的输出图 $P_{ij}(t)$ 进行阈值分割后得到, 其中, 阈值的选取由验证集评价获得。至此, 对于每一幅影像, 都有 n 个模型的 n 幅二值预测图。对阈值分割后的预测图 $W_{ij}(t)$ 进行合并, 可得数据集新的标签 $L_n(t)$, 具体为

$$L_n(t) = \frac{[W_{1n}(t) + W_{2n}(t) + \dots + W_{Nn}(t)]}{N} \quad (1)$$

式中, n 表示子数据集的序号, N 为总的模型数, $L_n(t)$ 为第 n 个子数据集合并后的结果。对每一幅影像的任一像素, N 个模型预测的标签相同时, 则该像素为确定的像素, 即对这个像素的预测大概率是正确的, 从而保留模型的预测作为该像素新的标签; 而 N 个模型预测不相同, 则该像素为不确定像素, 无法确定模型的预测是否准确, 则将 n 个模型预测之和与模型数的比例作为新的标签值。然后根据模型的预测结果 $W_{ij}(t)$, 计算掩膜 $M_n(t)$ 。对新的标签中的确定的像素, 掩膜相应位置赋予 1; 非确定像素, 相应位置赋予 0。具体为

$$M_n(t) = W_{1n}(t) \times W_{2n}(t) \times \dots \times W_{Nn}(t) + \{ [1 - W_{1n}(t)] \times [1 - W_{2n}(t)] \times \dots \times [1 - W_{Nn}(t)] \} \quad (2)$$

在生成了新的标签 $L_n(t)$ 和掩膜 $M_n(t)$ 后, 用

新的标签替代原有的标签。在此基础上, 重新在 N 个子数据集上训练模型。训练过程中, 利用掩膜 $M_n(t)$ 屏蔽不确定的像素, 参加训练的仅为确定的像素。重复上述过程, 不断地训练模型、优化标签、重新训练, 直到迭代次数 t 达到最大迭代次数 T 结束。最终得到 N 个训练好的深度语义分割网络模型。具体的损失函数计算为

$$Loss^{t+1} = - \sum_i \sum_j M_{ij}(t) \times \{ L_{ij}(t) \times \log P_{ij}(t+1) + [1 - L_{ij}(t)] \times \log [1 - P_{ij}(t+1)] \} \quad (3)$$

式中, t 表示迭代次数, i, j 表示图像的行列数, $L_{ij}(t)$ 、 $P_{ij}(t+1)$ 、 $M_{ij}(t)$ 分别表示标签、模型预测、掩膜在第 i 行第 j 列的像素值。

值得注意的是, 水体标签并非一直不变, 而是在每次迭代后, 用网络的输出作为新的标签, 标签中的噪声随着迭代次数的增加不断去除, 标签中确定的像素数会随着迭代次数的增加而不断增加, 水体真值的范围也不断扩大。

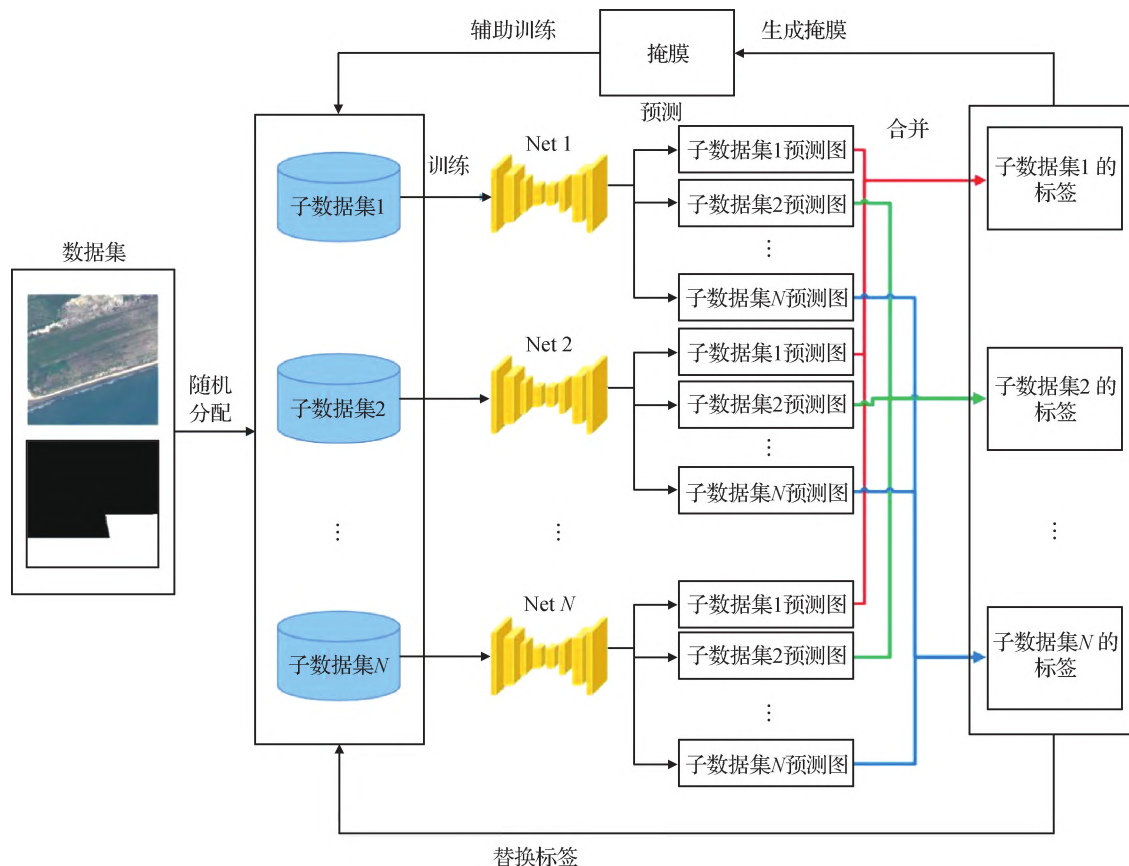


图3 训练阶段流程图

Fig. 3 Workflow of the training stage

2.2 多源遥感影像水体检测

测试阶段流程图如图 4 所示。首先用训练阶

段获得的 N 个模型分别对测试样本 I 进行预测, 得到输出图 P_i , i 表示模型的序号。对 N 个模型的

输出 P_i 进行平均, 获得平均后的预测图 P_{mean} , 具体为

$$P_{\text{mean}} = (P_1 + P_2 + \dots + P_N) / N \quad (4)$$

在验证集上测试出最好的阈值, 用该阈值对预测图 P_{mean} 进行阈值分割, 得到 W 作为图像 I 的水体检测结果。

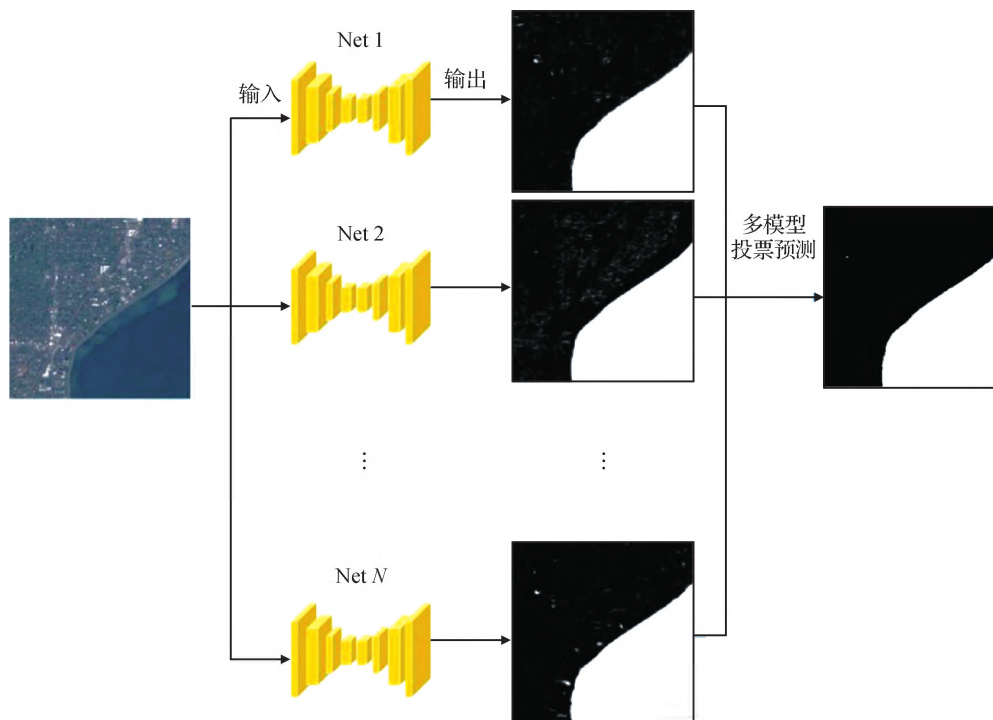


图4 测试阶段流程图

Fig. 4 Workflow of the test stage

3 实验及结果分析

3.1 实验设置及评估指标

在实验中, 利用每个波段的均值和标准差将数据范围归一化至相近的范围。深度语义分割网络采用 U-Net (Ronneberger 等, 2015) 网络结构, 每一次卷积通过填充保持特征图尺寸不变, 在网络的最后, 用一个 1×1 卷积核进行卷积, 代替全连接层, 可以不受图像尺寸限制, 然后用 sigmoid 激活函数将输出值限制在 $[0, 1]$ 范围内。最终得到一幅网络输出图, 其上每一个像素的值表示每一个像素属于水体的概率。网络训练过程中, 选择 Adam (Kingma 和 Ba, 2015) 为优化算法, 学习率设置为 0.000 1, 每批样本数量为 1, 训练过程共迭代 5 次, 记录每次迭代后模型在验证集上的精度并选择精度最高的模型作为最终模型。本文实验将子数据集的数量设为 3。采用准确率 (precision, P)、召回率 (recall, R)、F1 值 (F1-measure, F1)、像素精度 (pixel accuracy, PA) 和交并

比 (intersection over union, IoU) 作为精度评估指标, 具体公式为

$$P = \frac{P_{00}}{P_{00} + P_{10}} \quad (5)$$

$$R = \frac{P_{00}}{P_{00} + P_{01}} \quad (6)$$

$$F1 = \frac{2 \times P \times R}{P + R} \quad (7)$$

$$PA = \frac{P_{00} + P_{11}}{P_{00} + P_{01} + P_{10} + P_{11}} \quad (8)$$

$$IoU = \frac{P_{00}}{p_{00} + p_{01} + p_{10}} \quad (9)$$

式中, p_{00} 表示原本为水体预测为水体的像素数, p_{01} 表示原本为水体预测为非水体的像素数, p_{10} 表示原本为非水体预测为水体的像素数, p_{11} 表示原本为非水体预测为非水体的像素数。

3.2 关键参数分析

为准确评估本文方法的性能, 对迭代次数 T 和数据源组合两种参数对水体检测效果的影响分析。

首先, 设置输入波段为 Sentinel-1 号 SAR 影像

的 2 个波段和 Sentinel-2 号光学影像的 10 个波段。由于 Sentinel-2 号光学影像的 B1、B9、B10 波段为大气相关波段,原始分辨率为 60 m,较其他波段原始分辨率较低,故在实验中舍弃。迭代次数与水体检测精度的关系如图 5 所示。可以看到,随着迭代次数的增加,精度均在稳定上升,并在第 4 次迭代之后趋向收敛,此后的迭代精度不再上升,维持在一个较高的精度上。说明该方法的精度提升有上限,在抵达精度上限后就无法继续用该方法提升精度。最后一次迭代与直接用原始低分辨率标签进行训练的第一次迭代相比,IoU 提升了 4%,有较明显的提升。

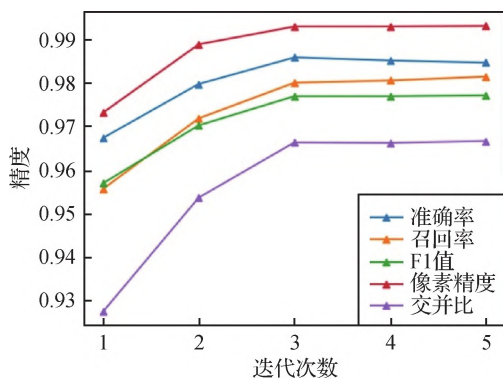


图 5 迭代次数与精度关系曲线

Fig. 5 Performance variation along with different iterations

然后,固定迭代次数 $T = 5$,分析数据源组合对水体检测效果的影响。实验对比了 3 种数据源组合的精度。1) 仅使用 Sentinel-1 号 SAR 影像的 2 个波段作为输入;2) 仅使用 Sentinel-2 号光学影像的 10 个波段作为输入,其中 B1、B9、B10 没有使用;3) 使用 Sentinel-1 号 SAR 影像的 2 个波段和 Sentinel-2 号光学影像的 10 个波段,其中 B1、B9、B10 没有使用。对比结果如图 6 所示。可以看出,3 种数据源组合中,仅使用 Sentinel-1 号 SAR 影像两个波段作为输入的精度远小于其他两种,说明只用 SAR 影像无法满足水体检测的需求。同时使用 SAR 影像和光学影像的水体检测精度略好于仅使用光学影像的水体检测精度。说明尽管单独使用 SAR 影像为输入的水体检测精度并不高,但与光学影像组合后,可以提升水体检测精度。

3.3 与已有方法的对比分析

为了验证本文方法的有效性,与归一化差异水体指数阈值法 NDWI (Mcfeeters, 1996) 和原始标签直接训练的 U-Net (Ronneberger 等, 2015) 进行实验

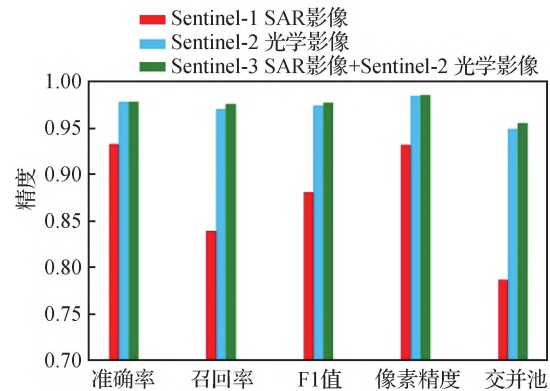


图 6 数据源组合和精度关系柱状图

Fig. 6 Performance variation along with different combinations of data sources

对比,精度评定结果和水体检测结果分别如表 1 和图 7 所示。从表 1 可以看出:1) 3 种方法中,本文方法效果最好,相比于直接用原始标签训练深度语义分割网络和传统的水体指数方法,IoU 分别提高了 7% 和 6%。说明本文方法可以有效解决水体数据集中水体标签分辨率低、噪声较多而影响水体检测精度问题。2) 直接用原始标签训练深度语义分割网络的方法精度比传统方法精度稍低,说明基于深度学习的水体检测方法并不一定比传统的水体检测方法效果更好。深度学习的效果依赖于训练数据的质量,低分辨率和噪声较多的水体标签反而会降低深度学习水体检测效果。从图 7 可以看出,水体指数阈值法和直接用原始标签训练的结果都一定程度上存在错分情况,而本文方法则可以较为精确地区分水体与非水体。

表 1 3 种方法精度对比

Table 1 Accuracy comparison of three methods

方法	准确率	召回率	F1 值	像素精度	交并比
NDWI	0.913 0	0.984 2	0.947 3	0.967 0	0.899 8
U-Net	0.948 9	0.927 8	0.938 2	0.963 1	0.883 6
本文	0.984 6	0.981 4	0.983 0	0.993 0	0.966 6

注:加粗字体为各列最优结果。

3.4 针对细小水体的方法对比

本文方法与直接用原始标签训练在精度上的差异主要是由细小河流和小湖泊等 500 m 分辨率样本标签中不能表达的水体造成的。为了更好地说明本文弱监督方法的效果,对 3 种方法在细小河流和小湖泊上的效果进行对比,并给出定量的分析结果。

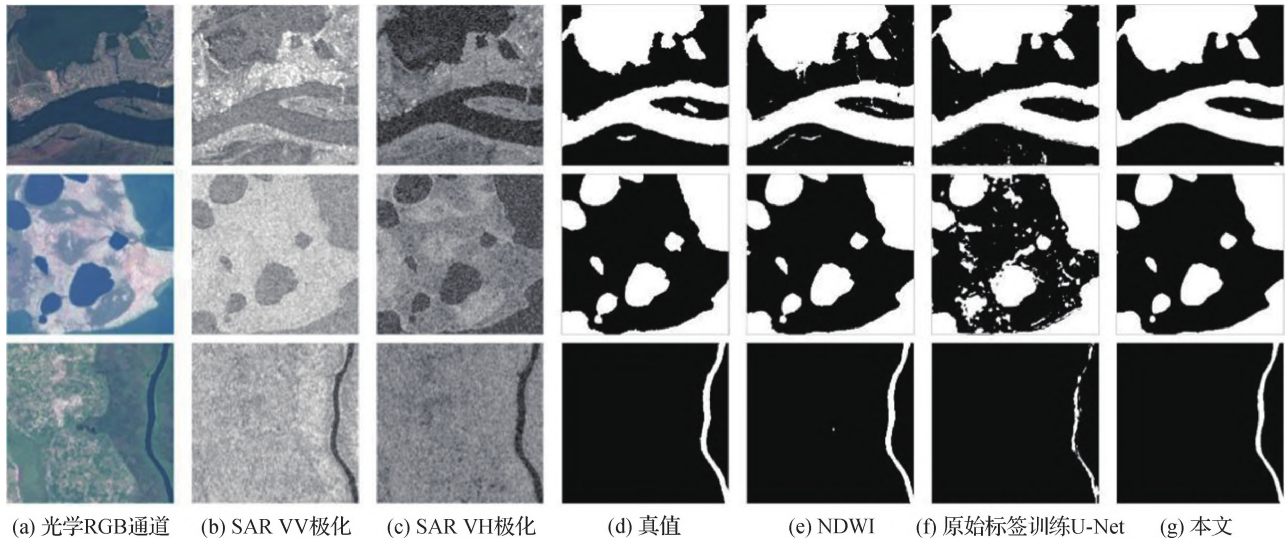


图7 不同方法水体检测结果对比

Fig. 7 Comparison of water detection results with different methods((a)optical RGB image;(b)SAR VV polarization;(c)SAR VH polarization;(d)ground truth;(e)NDWI;(f)prediction of U-Net;(g)ours)

为了针对性地分析不同方法在细小河流和小湖泊上的效果差异,从400个测试样本中,挑出了90个包含细小河流和小湖泊的样本作为新的测试样本。3种方法对细小水体的精度对比及包含细小河流和小湖泊样本的结果示意图分别如表2和图8所示。从表2可以看出,本文方法在细小河流和小湖泊上的精度略有下降,但是仍高于传统的水体指数阈值法及直

表2 3种方法对细小水体的精度对比

Table 2 Accuracy comparison of three methods for small water body

方法	准确率	召回率	F1 值	像素精度	交并比
NDWI	0.906 7	0.980 1	0.942 0	0.961 1	0.890 3
U-Net	0.914 5	0.878 8	0.896 3	0.945 0	0.812 1
本文	0.982 7	0.961 6	0.972 0	0.989 1	0.945 6

注:加粗字体为各列最优结果。

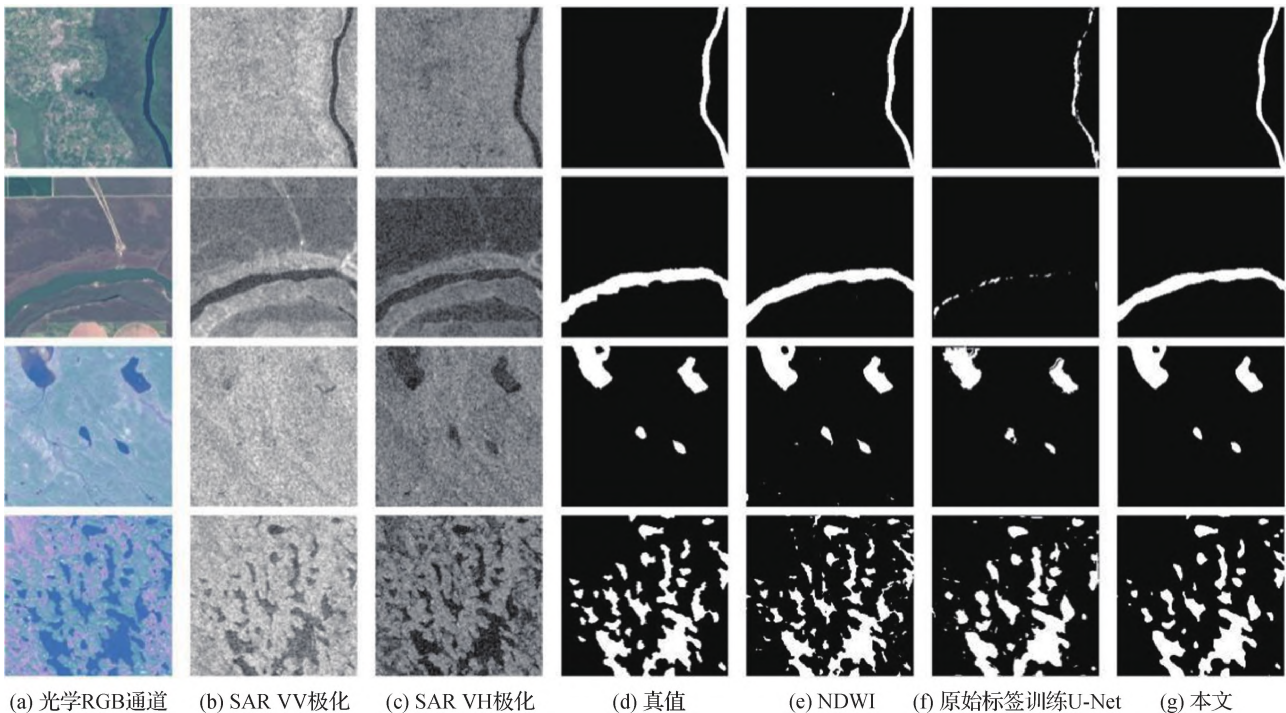


图8 不同方法对细小水体检测结果对比

Fig. 8 Comparison of detection results of fine water by different methods((a)optical RGB image;(b)SAR VV polarization;(c)SAR VH polarization;(d)ground truth;(e)NDWI;(f)prediction of U-Net;(g)ours)

接用原始标签训练的方法。从图 8 可以看出,直接用原始标签训练的方法错分现象非常明显,精度与水体指数阈值法和本文方法都有较大差异。而本文方法在细小河流和小湖泊上仍能有较好的效果。

4 结 论

针对公开水体覆盖产品中水体标签分辨率低和噪声较多导致深度语义分割网络性能下降问题,提出一种基于弱监督深度语义分割网络的多源遥感影像水体检测方法。在训练阶段,训练多视角深度语义分割网络,协同更新初始水体标签,并利用更新后标签训练新的多视角深度语义分割网络模型。在测试阶段,集成多个深度语义分割网络模型投票预测,得到最终的水体检测结果。

为了验证本文方法的有效性,创建了一个多源遥感影像水体检测数据集,训练集标签借助公开的 MODIS 水体覆盖产品生成,测试集和验证集标签均通过人工精确标注获得。实验表明,本文方法具有收敛性,在迭代 4 次后精度趋于稳定。并且光学影像、SAR 影像的融合对水体检测结果有一定提高。另外,在使用分辨率低、噪声多的水体标签进行训练的前提下,训练所得多视角模型的水体检测精度明显优于基于传统的水体指数阈值分割法和基于低质量水体标签直接学习的深度语义分割网络。

客观来说,训练和测试过程中,集成多个深度语义分割网络模型的方法采用的是简单的平均投票方法,仍有提升空间。未来考虑采用相对多数投票法、加权投票法,甚至学习法(通过另一个学习器进行集成),继续改善当前的模型集成方法,以提高水体检测精度。

参考文献 (References)

- Bearman A, Russakovsky O, Ferrari V and Li F F. 2016. What's the point: semantic segmentation with point supervision//Proceedings of the 14th European Conference on Computer Vision. Amsterdam, the Netherlands: Springer: 549-565 [DOI: 10.1007/978-3-319-46478-7_34]
- Carroll M L, Townshend J R, DiMiceli C M, Noojipady P and Sohlberg R A. 2009. A new global raster water mask at 250 m resolution. *International Journal of Digital Earth*, 2(4): 291-308 [DOI: 10.1080/17538940902951401]
- Chen J B, Liu S X, Wang C Y, You S C and Wang Z W. 2013. Research on urban water body extraction using knowledge-based decision Tree. *Remote Sensing Information*, 28(1): 29-33, 37 (陈静波, 刘顺喜, 汪承义, 尤淑撑, 王忠武. 2013. 基于知识决策树的城市水体提取方法研究. *遥感信息*, 28(1): 29-33, 37) [DOI: 10.3969/j.issn.1000-3177.2013.01.007]
- Chen Q, Zheng L J, Li X J, Xu C B, Wu Y, Xie D H and Liu L. 2019. Water body extraction from high-resolution satellite remote sensing images based on deep learning. *Geography and Geo-Information Science*, 35(4): 43-49 (陈前, 郑利娟, 李小娟, 徐崇斌, 吴侯, 谢东海, 刘亮. 2019. 基于深度学习的高分辨率遥感影像水体提取模型研究. *地理与地理信息科学*, 35(4): 43-49) [DOI: 10.3969/j.issn.1672-0504.2019.04.007]
- Dai J F, He K M and Sun J. 2015. BoxSup: exploiting bounding boxes to supervise convolutional networks for semantic segmentation//Proceedings of 2015 IEEE International Conference on Computer Vision (ICCV). Santiago, Chile: IEEE: 1635-1643 [DOI: 10.1109/iccv.2015.191]
- Ding F. 2009. A new method for fast information extraction of water bodies using remotely sensed data. *Remote Sensing Technology and Application*, 24(2): 167-171 (丁凤. 2009. 一种基于遥感数据快速提取水体信息的新方法. *遥感技术与应用*, 24(2): 167-171) [DOI: 10.11873/j.issn.1004-0323.2009.2.167]
- Duan Q Y, Meng L K, Fan Z W, Hu W G and Xie W J. 2015. Applicability of the water information extraction method based on GF-1 image. *Remote Sensing for Land & Resources*, 27(4): 79-84 (段秋亚, 孟令奎, 樊志伟, 胡卫国, 谢文君. 2015. GF-1 卫星影像水体信息提取方法的适用性研究. *国土资源遥感*, 27(4): 79-84) [DOI: 10.6046/gtzyyg.2015.04.13]
- Friedl M A, Meiver D K, Hodges J C F, Zhang X Y, Muchoney D, Strahler A H, Woodcock C E, Gopal S, Schneider A, Cooper A, Baccini A, Gao F and Schaaf C. 2002. Global land cover mapping from MODIS: algorithms and early results. *Remote Sensing of Environment*, 83(1/2): 287-302 [DOI: 10.1016/S0034-4257(02)00078-0]
- Ghosh A, Kumar H and Sastry P S. 2017. Robust loss functions under label noise for deep neural networks//Proceedings of the 31st AAAI Conference on Artificial Intelligence. San Francisco, USA: AAAI: 1919-1925
- Guo Y M, Liu Y, Georgiou T and Lew M S. 2017. A review of semantic segmentation using deep neural networks. *International Journal of Multimedia Information Retrieval*, 7(2): 87-93 [DOI: 10.1007/s13735-017-0141-z]
- He H Q, Du J, Chen T and Chen X Y. 2017. Remote sensing image water body extraction combing NDWI with convolutional neural network. *Remote Sensing Information*, 32(5): 82-86 (何海清, 杜敬, 陈婷, 陈晓勇. 2017. 结合水体指数与卷积神经网络的遥感水体提取. *遥感信息*, 32(5): 82-86) [DOI: 10.3969/j.issn.1000-3177.2017.05.013]

- Hung W C, Tsai Y, Liou Y, Lin Y Y and Yang M H. 2018. Adversarial learning for semi-supervised semantic segmentation [EB/OL]. [2020-06-05]. <https://arxiv.org/pdf/1802.07934.pdf>
- Kingma D P and Ba J L. 2015. Adam: a method for stochastic optimization [EB/OL]. [2020-06-05]. <https://arxiv.org/pdf/1412.6980.pdf>
- Khoreva A, Benenson R, Hosang J, Hein M and Schiele B. 2017. Simple does it: weakly supervised instance and semantic segmentation// Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA: IEEE: 1665-1674 [DOI: 10.1109/CVPR.2017.181]
- Li Y S, Zhang Y J, Huang X, Zhu H and Ma J Y. 2018a. Large-scale remote sensing image retrieval by deep hashing neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 56(2): 950-965 [DOI: 10.1109/TGRS.2017.2756911]
- Li Y S, Tao C, Tan Y H, Shang K and Tian J W. 2016. Unsupervised multilayer feature learning for satellite image scene classification. *IEEE Geoscience and Remote Sensing Letters*, 13(2): 157-161 [DOI: 10.1109/LGRS.2015.2503142]
- Li Y S, Zhang Y J, Huang X and Ma J Y. 2018b. Learning source-invariant deep hashing convolutional neural networks for cross-source remote sensing image retrieval. *IEEE Transactions on Geoscience and Remote Sensing*, 56(11): 6521-6536 [DOI: 10.1109/TGRS.2018.2839705]
- Li Y S, Zhang Y J, Huang X and Yuille A L. 2018c. Deep networks under scene-level supervision for multi-class geospatial object detection from remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 146: 182-196 [DOI: 10.1016/j.isprsjprs.2018.09.014]
- Li Y S, Zhang Y J and Zhu Z H. 2021. Error-tolerant deep learning for remote sensing image scene classification. *IEEE Transactions on Cybernetics*, 51(4): 1756-1768 [DOI: 10.1109/TCYB.2020.2989241]
- Li Y S, Chen W, Zhang Y J, Tao C, Xiao R and Tan Y H. 2020. Accurate cloud detection in high-resolution remote sensing imagery by weakly supervised deep learning. *Remote Sensing of Environment*, 250: #112045 [DOI: 10.1016/j.rse.2020.112045]
- Liao A P, Chen L J, Chen J, He C Y, Cao X, Chen J, Peng S, Sun F D and Gong P. 2014. High-resolution remote sensing mapping of global land water. *Science China Earth Sciences*, 57(10): 2305-2316 (廖安平, 陈利军, 陈军, 何超英, 曹鑫, 陈晋, 彭舒, 孙芳蒂, 宫鹏. 2014. 全球陆表水体高分辨率遥感制图. *中国科学: 地球科学*, 44(8): 1634-1645) [DOI: 10.1007/s11430-014-4918-0]
- Lin D, Dai J F, Jia J Y, He K M and Sun J. 2016. ScribbleSup: scribble-supervised convolutional networks for semantic segmentation// Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, USA: IEEE: 3159-3167 [DOI: 10.1109/cvpr.2016.344]
- Lu S L, Wu B F, Yan N N and Wang H. 2011. Water body mapping method with HJ-1A/B satellite imagery. *International Journal of Applied Earth Observation and Geoinformation*, 13(3): 428-434 [DOI: 10.1016/j.jag.2010.09.006]
- Mcfeeters S K. 1996. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing*, 17(7): 1425-1432 [DOI: 10.1080/01431169608948714]
- Mo W H, Sun H, Zhong S Q, Huang Y L and He L. 2007. Research on the CIWI model and its application. *Remote Sensing Information*, (5): 16-21 (莫伟华, 孙涵, 钟仕全, 黄永璘, 何立. 2007. MODIS 水体指数模型(CIWI)研究及其应用. *遥感信息*, (5): 16-21) [DOI: 10.3969/j.issn.1000-3177.2007.05.004]
- Pinheiro P O and Collobert R. 2015. From image-level to pixel-level labeling with convolutional networks// Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, USA: IEEE: 1713-1721 [DOI: 10.1109/CVPR.2015.7298780]
- Reed S, Lee H, Anguelov D, Szegedy C, Erhan D and Rabinovich A. 2015. Training deep neural networks on noisy labels with bootstrapping [EB/OL]. [2020-06-05]. <https://arxiv.org/pdf/1412.6596v1.pdf>
- Ronneberger O, Fischer P and Brox T. 2015. U-Net: convolutional networks for biomedical image segmentation// Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention. Munich, Germany: Springer: 234-241 [DOI: 10.1007/978-3-319-24574-4_28]
- Schmitt M, Prexl J, Ebel P, Liebel L and Zhu X X. 2020. Weakly supervised semantic segmentation of satellite images for land cover mapping-challenges and opportunities. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences*, 5(3): 795-802 [DOI: 10.5194/isprs-annals-V-3-2020-795-2020]
- Souly N, Spampinato C and Shah M. 2017. Semi supervised semantic segmentation using generative adversarial network// Proceedings of 2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE: 5689-5697 [DOI: 10.1109/ICCV.2017.606]
- Sukhbaatar S, Bruna J, Paluri M, Bourdev L and Fergus R. 2015. Training convolutional networks with noisy labels [EB/OL]. [2020-06-05]. <https://arxiv.org/pdf/1406.2080v4.pdf>
- Tan Y H, Xiong S Z and Li Y S. 2018. Automatic extraction of built-up areas from panchromatic and multispectral remote sensing images using double-stream deep convolutional neural networks. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(11): 3988-4004 [DOI: 10.1109/JSTARS.2018.2871046]
- Wang X, Sui L C, Zhong M Q, Li D M and Dang L L. 2018. Fully convolution neural networks for water extraction of remote sensing images. *Bulletin of Surveying and Mapping*, (6): 41-45 (王雪, 隋立

春, 钟棉卿, 李顶萌, 党丽丽. 2018. 全卷积神经网络用于遥感影像水体提取. 测绘通报, (6): 41-45 [DOI: 10.13474/j.cnki.11-2246.2018.0173]

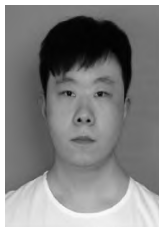
Xu H Q. 2005. A study on information extraction of water body with the modified normalized difference water index (MNDWI). Journal of Remote Sensing, 9(5): 589-595 (徐涵秋. 2005. 利用改进的归一化差异水体指数(MNDWI)提取水体信息的研究. 遥感学报, 9(5): 589-595) [DOI: 10.11834/jrs.20050586]

Yang Y and Ruan R Z. 2010. Extraction of plain lake water body based on TM imagery. Remote Sensing Information, (3): 60-64 (杨莹, 阮仁宗. 2010. 基于TM影像的平原湖泊水体信息提取的研究. 遥感信息, (3): 60-64) [DOI: 10.3969/j.issn.1000-3177.2010.03.013]

Yin Y Q, Li J G, Yu T, Yang H Y and Zhang Y H. 2015. The study of object-oriented water body extraction method based on high resolution RS image. Bulletin of Surveying and Mapping, (1): 81-85 (殷亚秋, 李家国, 余涛, 杨红艳, 张永红. 2015. 基于高分辨率遥感影像的面向对象水体提取方法研究. 测绘通报, (1): 81-85) [DOI: 10.13474/j.cnki.11-2246.2015.0016]

Zhou B L, Khosla A, Lapedriza A, Oliva A and Torralba A. 2016. Learning deep features for discriminative localization//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, USA: IEEE: 2921-2929 [DOI: 10.1109/CVPR.2016.319]

作者简介



李鑫伟, 1998年生, 男, 硕士研究生, 主要研究方向为基于深度学习的遥感图像处理。

E-mail: xinwei.lee@outlook.com



李彦胜, 通信作者, 男, 副教授, 主要研究方向为人工智能驱动的遥感大数据处理与知识挖掘。

E-mail: yansheng.li@whu.edu.cn

张永军, 男, 教授, 主要研究方向为航空航天摄影测量和智能化遥感影像解译。E-mail: zhangyj@whu.edu.cn