

基于自监督对比学习的寒旱区遥感图像河流识别方法

沈瑜¹ 王海龙¹ 梁栋² 牛东兴² 严源¹ 李阳阳¹

(1. 兰州交通大学电子与信息工程学院, 兰州 730070; 2. 中国中铁科学研究院有限公司, 成都 610032)

摘要: 针对遥感图像河流数据样本人工标注成本高且难以大量获取,以及网络在河流图像边缘细节提取的效果不佳问题,提出一种通过自监督对比学习方式利用大量无标签遥感河流图像数据进行编码器预训练,并使用少量标签数据对预训练后的编码器进行微调,同时在编解码结构中使用一种新的非均匀采样方式的语义分割网络。自监督对比学习可以利用大量无标签数据进行前置任务模型训练,仅需少量标签数据对下游河流提取任务模型微调即可;非均匀采样方式能够通过对高频区域密集采样、对低频区域稀疏采样的方式获得图像中不同类别之间清晰的边界信息和同类别区域中的细节信息,减少模型的冗余度。在河流数据集上的实验表明,利用360幅有标签数据对预训练后的网络进行微调,其像素准确率、交并比、召回率分别达到90.4%、68.6%和83.2%,与使用1200幅有标签数据训练的有监督AFR-LinkNet网络性能相当;在使用全部数据标签进行微调后,网络的像素准确率、交并比、召回率分别达到93.7%、73.2%和88.5%,相比AFR-LinkNet、DeepLabv3+、LinkNet、ResNet50和UNet网络,像素准确率分别提高3.1、7.6、12.3、14.9、19.8个百分点,交并比分别提高3.5、8.7、10.5、16.9、24.0个百分点,召回率分别提高2.1、4.8、6.7、9.4、12.9个百分点,验证了模型在河流图像上精准提取河流的有效性。该算法模型对于解决缺少大量有标签数据和分析我国寒旱区河流分布、水灾害预警、水资源合理利用以及农业灌溉发展等具有重要意义。

关键词: 遥感图像; 河流提取; 自监督对比学习; 非均匀采样; 语义分割; AFR-LinkNet

中图分类号: TP391.4 **文献标识码:** A **文章编号:** 1000-1298(2023)06-0125-11 **OSID:** 

River Extraction Method from Remote Sensing Images of Cold and Arid Regions Based on Self-supervised Comparative Learning

SHEN Yu¹ WANG Hailong¹ LIANG Dong² NIU Dongxing² YAN Yuan¹ LI Yangyang¹

(1. School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China

2. China Railway Academy Co., Ltd., Chengdu 610032, China)

Abstract: Aiming at the problems of high cost of manual labeling of river data samples in remote sensing images and difficulty in obtaining a large number of them, as well as poor effect of network extraction of river image edge details, a self-supervised comparative learning method was proposed to use a large number of unlabeled remote sensing river image data for encoder pre-training, and a small amount of label data was used to fine-tune the encoder after pre-training. Meanwhile, a semantic segmentation network based on non-uniform sampling was used in the codec structure. Self-supervised comparative learning can use a large number of unlabeled data for pre-task model training, and only a small amount of label data was needed to fine-tune the downstream river extraction task model. The non-uniform sampling method can obtain clear boundary information between different categories in the image and details in the same category by intensive sampling in the high frequency region and sparse sampling in the low frequency region, thus reducing the redundancy of the model. Experiments on river data sets showed that the pixel accuracy, intersection over union and recall rate of the pre-trained network can reach 90.4%, 68.6% and 83.2%, respectively, when 360 labeled datas was used to fine-tune the network, which was

收稿日期: 2022-10-05 修回日期: 2022-11-20

基金项目: 国家自然科学基金项目(61861025、62241106)、2021年陇原青年创新创业人才(团队)项目、甘肃省高等学校青年博士基金项目(2021QB-49)、甘肃省高校大学生就业创业能力提升工程项目(2021-35)、智能化隧道监理机器人研究项目(中铁科研院(科研)字2020-KJ016-Z016-A2)和四电BIM工程与智能应用铁路行业重点实验室开放项目(BIMKF-2021-04)

作者简介: 沈瑜(1982—),女,教授,博士生导师,主要从事数字图像处理与人工智能研究,E-mail: 18609311366@163.com

comparable to the performance of the supervised AFR – LinkNet network trained with 1 200 labeled datas. After fine-tuning with all data labels, the pixel accuracy, intersection ratio and recall rate of the network reached 93.7% , 73.2% and 88.5% , respectively. Compared with AFR – LinkNet, DeepLabv3 + , LinkNet, ResNet50 and UNet networks, the pixel accuracy was increased by 3.1 , 7.6 , 12.3 , 14.9 , 19.8 percentage points, the intersection over union was increased by 3.5 , 8.7 , 10.5 , 16.9 , 24.0 percentage points, and the recall rate was increased by 2.1 , 4.8 , 6.7 , 9.4 , 12.9 percentage points, respectively. The effectiveness of the model to accurately extract rivers from river images was verified. This algorithm model was of great significance for solving the lack of a large number of labeled data and analyzing the distribution of rivers in cold and arid regions, water disaster warning, rational utilization of water resources and agricultural irrigation development.

Key words: remote sensing image; river extraction; self-supervised comparative learning; non-uniform sampling; semantic segmentation; AFR – LinkNet

0 引言

河流作为生态系统的重要组成部分,对自然环境以及人们的生产生活有着至关重要的作用,尤其对我国寒旱区的河流进行识别,在农业灌溉、水利水电调控、生态监测和环境改善等方面有着重大意义^[1-3]。

近年来,使用深度学习方法从高分辨率遥感图像中对河流进行识别成为主要方法之一^[4-5]。文献[6]构建并训练了卷积神经网络和 DeepLabv3 两种水体识别模型,其识别精度分别达到 95.09% 和 92.14% ,均高于支持向量机、面向对象以及水体指数等方法。文献[7]通过在 DenseNet 网络中增加上采样和全卷积网络内的跳层连接,缓解梯度消失和网络退化问题,其识别精度优于其他深度神经网络,精度高达 96% 以上。文献[8]通过改进 U – Net 网络,并引入条件随机场进行后处理,精细化了分割结果,实验结果表明,该网络对小目标水体能够准确识别。文献[9]用自适应简单线性聚类算法(ASLIC)将遥感图像分割成高质量的超像素,再利用新的池化卷积神经网络提取水体高级特征并进行二分类标记,该方法实验平均总准确率为 99.14% ,高于传统方法。文献[10]采用集成预测,同时优化有标签样本上的标准监督分类损失及无标签数据上的非监督一致性损失,来训练端到端的语义分割网络。文献[11]利用半监督学习中的生成对抗性网络,用 softmax 替代最后的输出层,结合自动化分类诊断进行实验。

以上研究均依靠大量标签数据,由于有标签数据的制作难度和成本较高,利用无标签数据进行网络训练的自监督学习(Self-supervised learning, SSL)方法逐渐引起重视^[12-13]。自监督学习属于无监督学习,无需大量标签数据,只需通过构建前置任务(Pretext task)从无标签数据中学习自身的监督信

息,然后将训练好的模型迁移到下游任务中,使用少许标签数据微调后训练,最后进行目标任务预测^[14-16]。自监督学习的主流方法包括生成式、对比式、生成式对比自监督学习(对抗学习)方法,其中对比式自监督学习方法应用最广泛^[17]。文献[18]提出了一个全局风格和局部匹配的对比学习网络,在 Postdam 数据集上 Kappa 系数提高 6%。文献[19]设计了 3 个不同的前置任务和 3 个一组的孪生网络进行训练,实验结果表明,只需 10% ~ 50% 的标记数据即可达到有监督网络相同的性能。文献[20]提出一种自监督学习算法(Inpainting based self-supervised learning, IBS)解决了电力线分割任务缺乏大规模数据的问题,实验表明其精度和速度都超越了已有的方法。文献[21]提出一种视觉表示的自监督对比学习框架(Simple framework for contrastive learning of visual representations, SimCLR),该结构不需要专门的架构,也不需要特殊的存储库,因此具有通用性且网络性能更优。自监督学习方法训练的网络模型可以学习遥感图像潜在的特征,获得更好的拟合起点,并取得优秀的分割结果,将自监督学习获得的学习参数用作预训练模型,利用迁移学习对后续遥感图像任务进行微调,可证明网络的泛化能力^[22-23]。以上基于自监督学习范式的研究虽然已经取得大量成果,但其大多数针对分类问题进行研究,并且网络收敛速度较慢,编码器常采用 ResNet50 网络,识别精度有限,在具有较多干扰的寒旱区遥感图像河流提取方面鲜有报道。

本文针对大量有标签的寒旱区遥感图像河流数据很难获得,以及河流边缘细节难以识别的问题,采用自监督对比学习方式并对其进行改进,充分利用无标签数据对遥感图像河流提取网络进行预训练,同时在下游任务训练好的编解码网络中使用一种新的非均匀采样方式,提取河流边缘细节,通过实验证明其有效性和泛化性能。

1 网络框架

网络总体框架如图 1 所示,分为基于自监督对比学习的预训练阶段和下游目标任务提取阶段两部分,以 AFR - LinkNet^[5] 网络为基础网络构建模型。首先将大量无标签遥感图像河流数据进行数据增强,获得大量正负样本对,然后输入基于改进的 SimCLR 框架的自监督网络模型进行训练,其编码器使用下游目标任务中 AFR - LinkNet 网络的编码器进行替换,并在映射头的 Dense 层和 ReLU 激活函数之间添加 BN (Batch normalization) 层。编码器与映射头结合的两次非线性映射能够大幅度提升对比学习性能,编码器后的图像特征会有更多包含图像增强信息在内的细节特征,即编码更多与任务无关

的通用细节信息,在高层网络映射头中编码更多与对比学习任务相关的信息。网络将正样本对拉近,将负样本对拉远,不断进行对比,直到正样本对比损失最小,否则不断梯度回传更新编码器参数。最后将训练好的编码器模型参数作为预训练的编码器,利用迁移学习方式迁移到下游目标任务的语义分割网络的编码器中。其次,将少量带有标签的遥感图像河流数据输入 AFR - LinkNet 语义分割网络,微调后进行训练。将编码器输出的粗粒度语义图和解码器输出的细粒度语义图,通过对高频区域多采样、低频区域少采样的非均匀采样策略后进行特征融合,最后将融合后输出的高精度河流分割语义图与语义标签图进行损失函数校验,如果损失最小,则输出分割结果,否则进行模型参数更新。

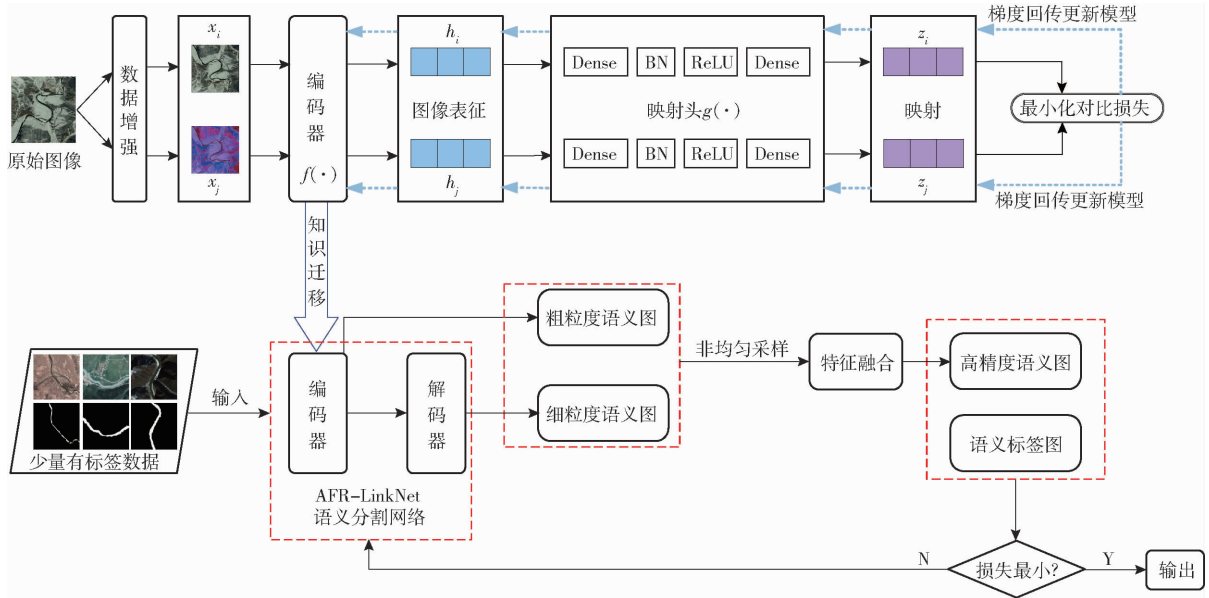


图 1 网络框架

Fig. 1 Network framework

2 相关原理

2.1 自监督学习

机器学习一般分为有监督学习、半监督学习、无监督学习和强化学习,而自监督学习是一种新的无监督机器学习方法,能够挖掘出数据自身的监督信息关系作为标注样本进行训练,通过这种方式可以获得更多的语义特征,解决了有监督学习中需要大量标注数据的问题,其核心是在无标签的预训练数据上进行自监督任务的训练,通过最小化损失函数更新模型参数,得到一个可以对数据产生良好特征表征的编码器模型^[24]。

自监督学习原理如图 2 所示,为了更好地迁移模型参数,自监督代理任务和下游目标任务使用相同的编码器模型。在自监督代理任务阶段,首先将

大量无标签数据输入编码器,得到含有自身监督信息的图像特征,再通过自监督训练任务模型(如生成型、对比型等)得到预测图像,然后进行损失对比,直到损失最小时停止梯度回传,将预训练好的编码器模型参数迁移到下游目标任务。在下游目标任务阶段,将少量有标签数据输入到经过预训练的编码器模型中,然后微调参数后开始训练,将提取到的图像特征通过下游目标任务模型(如图像分类、语义分割等)获得预测图像,最后进行损失函数校验,符合损失函数最小条件后输出最终的预测图像。

自监督对比学习是目前主流研究方法之一,并且取得了较好的效果,其核心思想是通过一个辅助任务,构建正负样本对,使网络通过比较正负样本的距离差学习到将相似样本(正样本对)拉近,不相似样本(负样本对)拉远的能力,从而可以获得特征可

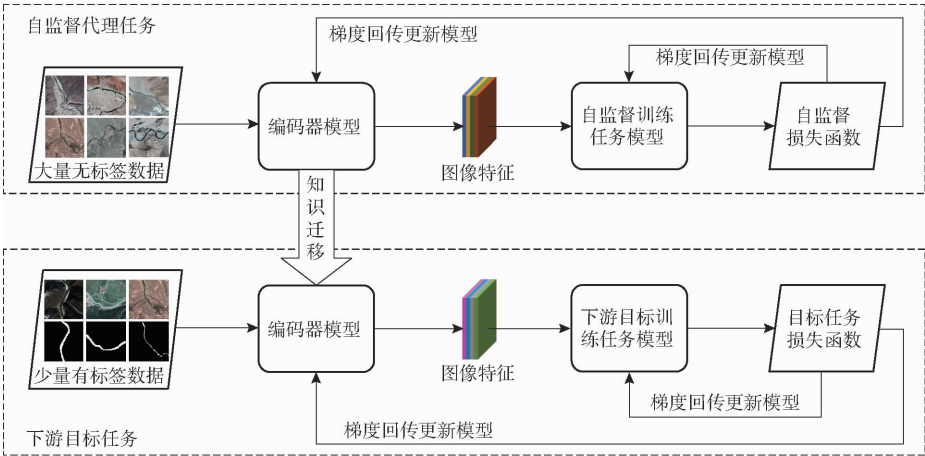


图 2 自监督学习原理图

Fig.2 Schematic of self-supervised learning

区分性表达的目标,即

$$\text{score}(f(x),f(x^+))\gg\text{score}(f(x),f(x^-)) \quad (1)$$

式中 $\text{score}(\cdot)$ ——样本相似度比较函数

$f(x)$ ——原始图像预测图

$f(x^+)$ ——正样本预测图

$f(x^-)$ ——负样本预测图

其实现步骤如图 3 所示。

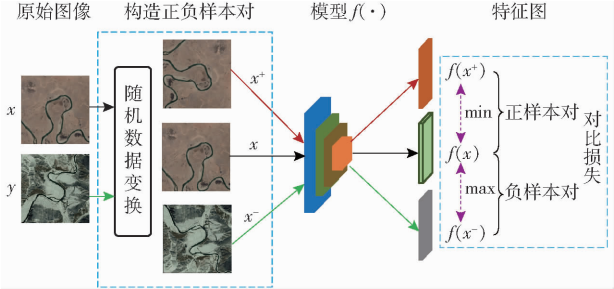


图 3 自监督对比学习原理图

Fig.3 Schematic of self-supervised comparative learning

其中,合理构建正样本对,并包含足够多且足够难的负样本对能够有效提升对比自监督学习的学习性能。文献[25]提出使用动量对比学习的方法对负样本编码器进行更新,并且保持该负样本队列足够大,巧妙地解耦模型批次大小和一个学习批次可容纳的负样本数量。

目前,常见的主流自监督对比学习框架包括 MoCo、SimCLR、SwAV、SimSiam 等。其中, SimCLR 模型兼具框架简单与网络性能高的特点,但其收敛速度较慢、预训练编码器提取精度有限。因此,本文将对 SimCLR 框架进行改进,并作为自监督预训练阶段的网络模型。

SimCLR 模型主要包含 3 部分:数据随机增强、深度神经网络和损失对比函数,如图 4 所示,其核心原理是利用同一样本数据增强后具有相似特征、不同样本之间数据增强后具有差异特征构造损失函数

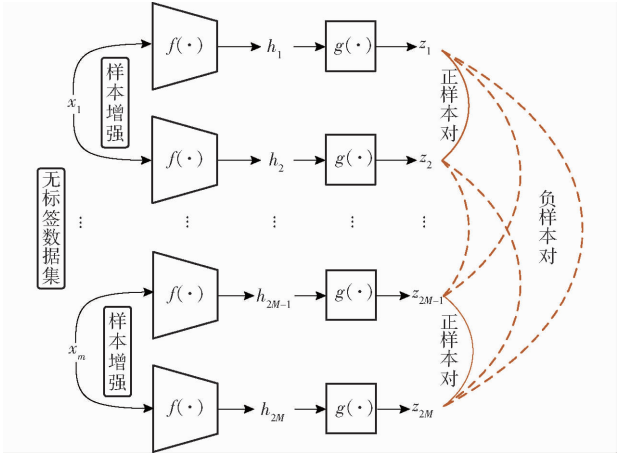


图 4 SimCLR 原理图

Fig.4 SimCLR principle framework

进行特征学习。

(1)数据随机增强

自监督学习利用样本自身的信息进行学习,不需要样本标签。自监督对比学习只需通过原始数据构造出正负样本对即可进行学习。如图 4 所示,从训练集中随机选取 M 幅河流图像为一个批次进行样本增强,将每幅河流图像经过两次随机裁剪获得两个增强数据,使得其中任意一幅图像保持不变,另一幅图像进行随机色度变换、旋转、高斯模糊、随机颜色失真等。同一图像经数据增强后的两个样本为一个正样本对,剩余的 $2(M-1)$ 个增强的样本数据全部为该图像的负样本。因为正样本来源于同一图像,具有内在信息联系,负样本来源于不同图像,具有内在信息差异,因此可以进行图像信息学习。

(2)深度神经网络

如图 4 所示, SimCLR 网络由特征提取模块 $f(\cdot)$ 和特征映射模块 $g(\cdot)$ 构成。特征提取模块一般使用 ResNet 等经典网络实现,本文使用 AFR-LinkNet 网络中的编码器作为特征提取模块,获得输

入图像样本的有效特征。若 x_i 为原始图像数据样本, $\hat{x}_i$ 为增强后的图像数据, 则经过编码器后的特征表示为 $h_i = f(\hat{x}_i)$ 。特征映射模块使用包含非线性激活函数的多层全连接网络实现, 是为了将 h_i 映射到损失函数所在的空间进行对比损失计算, 本文在非线性激活函数前面增加 BN 层进行改进, 提高网络训练和收敛速度。SimCLR 网络中, $z_i = g(h_i)$, 为特征映射模块的输出, 单层全连接网络为 $g(h_i) = \mathbf{W}^{(2)} \sigma(\mathbf{W}^{(1)} h_i)$, 其中 σ 为激活函数 ReLU, $\mathbf{W}^{(1)}$ 、 $\mathbf{W}^{(2)}$ 为权重参数矩阵 $\mathbf{W}$ 的一阶和二阶导数。

(3) 损失对比函数

数据增强后的样本通过特征提取和特征映射模块, 最后需要使用归一化加权交叉熵损失函数 (Normalized temperature-scaled cross entropy loss, NT-Xent) 将正样本对之间的相似度最大化, 负样本对之间的相似度最小化, 最终达到能够学习遥感河流图像中的通用特征表示。NT-Xent 损失函数定义为

$$L = \frac{1}{2M} \sum_{i=1}^M (\ell(2i-1, 2i) + \ell(2i, 2i-1)) \quad (2)$$

式中 M ——批次大小

其中 $\ell(j, k)$ 为正样本对的损失对比函数, 定义为

$$\ell(j, k) = -\lg \frac{\exp \frac{\text{sim}(z_j, z_k)}{\tau}}{\sum_{m=1}^{2M} f(m, j) \exp \frac{\text{sim}(z_j, z_m)}{\tau}} \quad (3)$$

其中

$$f(m, j) = \begin{cases} 1 & (m \neq j) \\ 0 & (m = j) \end{cases} \quad (4)$$

$$\text{sim}(z_j, z_k) = \frac{z_j^T z_k}{\|z_j\| \|z_k\|} \quad (5)$$

式中 z ——特征映射网络输出

τ ——权重, 取 0.1

m ——数据增强后正样本

j ——用于数据增强样本

$f(m, j)$ ——指示函数

$\text{sim}(\cdot)$ ——数据相似度函数

2.2 非均匀采样

图像采样可分为均匀采样和非均匀采样, 河流与非河流边界处具有不规则性, 且阴影等干扰因素较多, 均匀采样针对河流不规则边界处区分度低、识别精度较差。因此, 要对河流边缘高精度提取, 只能通过对高频区域密集采样、对低频区域稀疏采样的非均匀采样方式获得图像中不同类别之间清晰的边界信息和同类别区域中的细节信息, 同时减少模型的冗余度。

目前多数研究都是针对输入图像的非均匀采样, 本文针对编解码器输出的粗粒度图像和细粒度

图像分别进行非均匀采样, 粗粒度图像中包含不同类别之间详细的类别信息, 能够明确局部边界特征, 细粒度图像中蕴含同类别的细节信息, 具有更多的语义信息和上下文信息, 能够提供全局特征信息, 其原理如图 5 所示。首先, 原始遥感河流图像经过编解码网络分别输出粗粒度图像和细粒度图像, 再使用轻量级分割头对检测到的河流对象进行粗滤预测 (红色框内), 利用点选择策略选择一组点 (红点), 使用小的多层感知器 (Multilayer perceptron, MLP) 对每个点独立预测, 进行细化。使用双线性插值作用于粗粒度图像上的这些点, 获得图像低级信息, 同时使用 2 倍双线性插值作用于细粒度图像上的这些点, 获得图像高级信息, 将高级信息和低级信息特征融合后得到最终的特征向量。最后 MLP 使用在这些点上计算的插值特征 (蓝色虚线) 对特征向量进行掩膜预测, 获得每个点的预测结果。不断使用该过程迭代地细化预测掩膜的不确定区域, 实现非均匀采样, 直到获得预期分辨率的河流语义图。

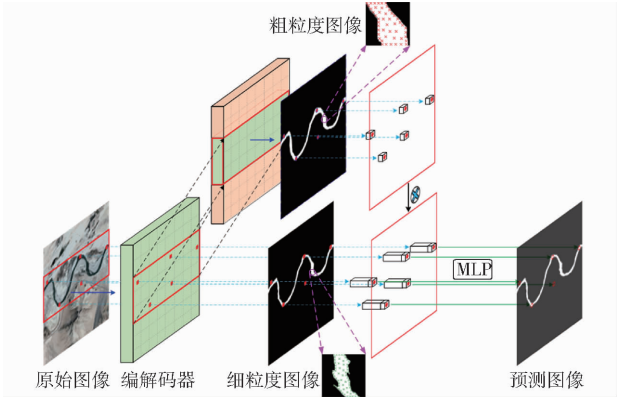


图 5 非均匀采样原理图

Fig. 5 Non-uniform sampling

2.2.1 双线性插值

原始遥感河流图像经过特征提取器后, 其图像尺寸和图像中特定点的像素与原始图像相比较已经发生较大变化, 为将其恢复到原始图像尺寸, 可通过双线性插值对其进行上采样操作。双线性插值是在两个方向分别进行一次线性插值操作, 原理是待插点像素取原图像中与其相邻的 4 个点像素的水平、垂直两个方向上的线性内插, 即根据待采样点与周围 4 个邻点的距离确定相应的权重, 从而计算出待采样点的像素。其原理和步骤如图 6、7 所示, 图中灰色板表示像素点的坐标, 蓝色、红色和绿色点分别表示初始像素点、中间像素点和输出像素点。

为得到未知函数 f 在点 $P = (x, y)$ 的值, 假设已知函数 f 在 $Q_{11} = (x_1, y_1)$ 、 $Q_{12} = (x_1, y_2)$, $Q_{21} = (x_2, y_1)$ 以及 $Q_{22} = (x_2, y_2)$ 4 个点的值。 f 就是一个像素点的像素值, 首先在 x 方向进行线性插值, 得到

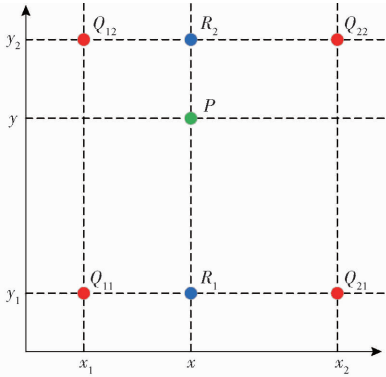


图 6 双线性插值
Fig. 6 Bilinear interpolation

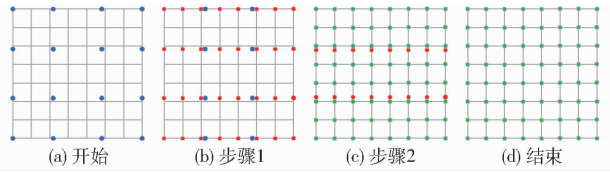


图 7 基于插值的上采样方法

Fig. 7 Interpolation based upsampling methods

$$f(x, y_1) = \frac{x_2 - x}{x_2 - x_1} f(Q_{11}) + \frac{x - x_1}{x_2 - x_1} f(Q_{21}) \quad (6)$$

$$f(x, y_2) = \frac{x_2 - x}{x_2 - x_1} f(Q_{12}) + \frac{x - x_1}{x_2 - x_1} f(Q_{22}) \quad (7)$$

然后在 y 方向进行线性插值,得到

$$f(P) = \frac{y_2 - y}{y_2 - y_1} f(x, y_1) + \frac{y - y_1}{y_2 - y_1} f(x, y_2) \quad (8)$$

综合得到双线性插值结果为

$$f(x, y) = \frac{f(Q_{11})}{(x_2 - x_1)(y_2 - y_1)} (x_2 - x)(y_2 - y) + \frac{f(Q_{21})}{(x_2 - x_1)(y_2 - y_1)} (x - x_1)(y_2 - y) + \frac{f(Q_{12})}{(x_2 - x_1)(y_2 - y_1)} (x_2 - x)(y - y_1) + \frac{f(Q_{22})}{(x_2 - x_1)(y_2 - y_1)} (x - x_1)(y - y_1) \quad (9)$$

2.2.2 点选择策略

针对遥感图像中不规则的河流边界,如果对全局点采样将增大计算量,非均匀采样方式基于点选择策略,只对预测困难点进行采样,可以大幅度降低计算量,增加采样灵活性,有效恢复图像中河流的边缘细节。

点选择策略核心思想是能够在河流图像平面上的高频区域附近(如河流边界等)灵活、自适应、密集地选择用于预测分割标签的点。选择过程中只计算值与邻近值显著不同的可能性较大的位置,实现高分辨率图像的有效分割,对于所有其他位置的预测值是通过插值已经计算的输出值(粗预测图)来获得。针对各个区域,均使用由粗到细的方式不断

地迭代输出预测图,对规则网格上的点只需要进行最粗级别预测即可。在每次迭代预测中,利用双线性插值对其之前预测的河流语义分割图实行上采样,然后在较为密集的网格上选取概率为 0.5 的 N 个点,最后计算这 N 个点的特征表征,同时预测出它们的标签,并不断重复该步骤,直到获得期望的分辨率。如图 8 所示,使用双线性插值将分辨率 4×4 上的预测进行 2 倍上采样,获得分辨率 8×8 ,然后对这 N 个点(蓝色点)进行预测,以便于在更高分辨率的图像上恢复河流边缘的细节信息。

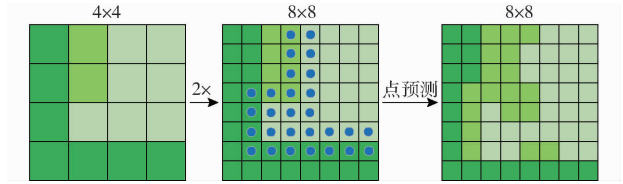


图 8 分辨率 4×4 细化到分辨率 8×8 示意图

Fig. 8 Resolution 4×4 refined to resolution 8×8 schematic

假设初始分辨率为 $S_0 \times S_0$,期望能够获得 $S \times S$ 分辨率,点选策略的预测点数要求不能超过 $Nlb \frac{S}{S_0}$ 个点的预测,这样可以更有效地进行高分辨率预测。由于点选择策略引入了连续步骤,影响神经网络的反向传播,因此在训练中需要使用基于随机采样的非迭代策略。首先在特征图上选择 N 个点训练,并遵循一定规则使得这些点的选择倾向于概率为 0.5 的区域,同时能够保留一定程度的均匀覆盖。其遵循的规则为:①通过随机采样 kN 个点作为候选点($k > 1$),保证覆盖面均匀。②重点关注进行插值计算后的 kN 个候选点中不确定的点,并且对不确定性进行估计,从 kN 个候选点中选择最不确定的 αN 个点, $\alpha \in [0, 1]$ 。③其余的 $(1 - \alpha)N$ 个点从均匀分布中进行随机采样。

为了平衡网络数据计算量和均匀覆盖问题,本文 k 取 4, α 取 0.75(适度偏置能够使得训练更有效),在进行训练时,只计算 N 个采样点上的预测值和损失函数值,使计算效率更高。

3 数据集与预处理

实验数据集从高分二号卫星图像中截取 600 幅尺寸为 1024 像素 $\times$ 1024 像素的寒旱区遥感河流图像,将其中 300 幅图像进行二值化人工标注,河流标注为 1,其他的均标注为 0。通过数据扩增将这 300 幅原始图像和对应的语义图像按照相同的数据扩增方式均拓展为 1500 幅,最后将剩余未标注的 300 幅图像通过数据扩增拓展为 1500 幅无标签数据,并单独截取 100 幅相同规格的河流图像进行语

义标注,作为网络的测试集。

因此,遥感图像河流数据集共 3 100 幅河流图像,其中训练集、验证集、测试集比例为 24:6:1。训练集中 1 200 幅无标签数据用于训练自监督对比学习网络,1 200 幅有标签数据用于训练有监督网络和对自监督对比学习网络进行微调,图 9 为数据扩增示意图。

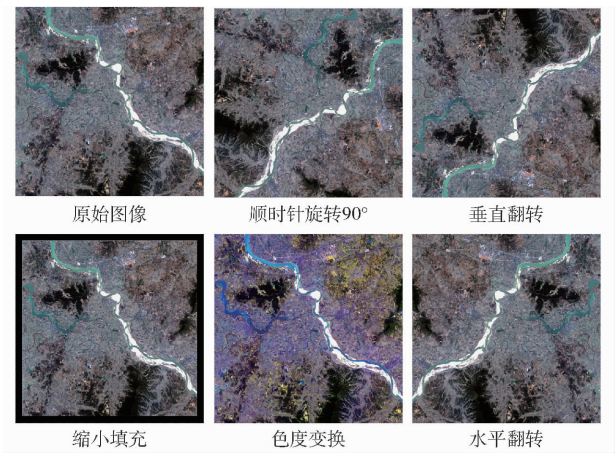


图 9 数据扩增示意图

Fig. 9 Schematic of data amplification

使用最大最小标准化对原始图像进行归一化预处理,可以将所有数据映射到[0,1]区间内,减少模型计算量以及加快模型的收敛速度。

4 实验结果与分析

4.1 实验设置

实验平台为 Windows 10 操作系统、CPU i9 - 11980HK、GPU RTX3080 (16 GB 显存)、内存 32 GB,使用 PyTorch 框架搭建网络模型,编程语言为 Python。自监督学习时的批次大小 M 设置为 10,同批次生成增强样本时按概率 0.25 进行随机裁剪、旋转、翻转和颜色失真,由此构成正负样本对。自监督学习的初始学习率设置为 2.5×10^{-4} ,迭代轮次为 400,优化器均选择 Adam 优化器,在少量带标签数据上微调模型时使用相同的学习率和优化器,迭代训练至网络收敛,网络全部使用文献[26]的初始化权重。

4.2 评价指标

采用像素准确率 (Accuracy, ACC)、召回率 (Recall) 和交并比 (Intersection over union, IoU) 作为评价指标。

4.3 实验结果

通过对比和消融实验,证明本文提出的基于自监督对比学习的遥感图像河流识别网络在只有少量标签数据情况下,只需对用无标签数据训练好的自监督网络进行微调,各项评价指标便超过了使用大

量标签数据训练的有监督网络模型。实验结果表明,网络模型的像素准确率、交并比和召回率分别达到 93.7%、73.2% 和 88.5%,超过有监督网络 AFR - LinkNet、DeepLabv3 +、LinkNet、ResNet50 和 UNet,河流图像提取结果在边缘细节上有了较高提升,优于其他有监督网络;使用 360 幅有标签数据微调网络时,其像素准确率达到 90.4%,与有监督 AFR - LinkNet 网络提取精度相当。

4.3.1 对比实验

将使用 1 200 幅有标签数据进行微调的 SimCLR + AFR - LinkNet + 非均匀采样网络(自监督对比学习网络)与使用相同数量有标签数据进行有监督训练的网络 AFR - LinkNet、DeepLabv3 +、LinkNet、ResNet50 和 UNet 进行对比,图 10 为验证集上各网络的像素准确率变化曲线,图 11 为网络在训练集和验证集上的损失值变化曲线,表 1 为各网络在测试集上的评价指标结果。

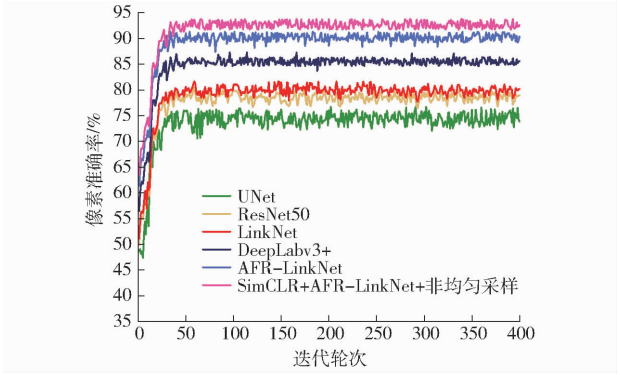


图 10 各网络像素准确率变化曲线

Fig. 10 Change curves of pixel accuracy of each network

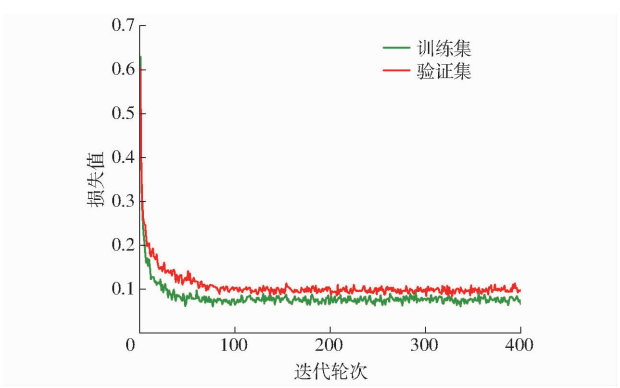


图 11 训练集与验证集上网络损失值变化曲线

Fig. 11 Change curves of network loss function on training set and test set

由图 10、11 可以看出,自监督网络在 50 个迭代轮次时已经趋于稳定,像素准确率高其他有监督网络,收敛速度较快,训练集和验证集上的损失都较小,在验证集上的损失稍大于训练集上的损失,在前 50 个迭代轮次时损失下降较快,验证了自监督网络的各种超参数设置较为合理且网络对河流提取准确

率较高、收敛性较好。

表 1 各网络评价指标

Tab. 1 Evaluation indicators of each network %			
网络模型	像素准确率	交并比	召回率
UNet	73.9	49.2	75.6
ResNet50	78.8	56.3	79.1
LinkNet	81.4	62.7	81.8
DeepLabv3 +	86.1	64.5	83.7
AFR - LinkNet	90.6	69.7	86.4
SimCLR + AFR - LinkNet + 非均匀	93.7	73.2	88.5
采样			

由表 1 可知,自监督网络像素准确率、交并比和召回率分别达到 93.7%、73.2% 和 88.5%,各项评价指标均高于有监督网络 AFR - LinkNet、DeepLabv3 +、LinkNet、ResNet50 和 UNet,像素准确率分别提高 3.1、7.6、12.3、14.9、19.8 个百分点,交并比分别提高 3.5、8.7、10.5、16.9、24.0 个百分点,召回率分别提高 2.1、4.8、6.7、9.4、12.9 个百分点。

图 12 为 3 幅不同网络的遥感河流图像的语义分割效果图,图中红色标记为不同网络提取效果的主要区别点。从图 12 可以看出,UNet 网络的识别效果最差,ResNet50、LinkNet 和 DeepLabv3 + 网络错误地将道路、冰雪以及山体阴影识别为河流,在细小河流处出现较多断续,AFR - LinkNet 网络与 SimCLR + AFR - LinkNet + 非均匀采样网络在细节提取上更精细、更准确,但是因为自监督对比学习模型以 AFR - LinkNet 网络的编码器为基础进行无标签数据训练,具有较强的抗干扰性,并且学习到了有监督模型无法学习到的图像内在蕴含的一些细节信息,因此在河流的边缘细节提取效果上表现更好,在特别细小的河流处基本没有出现断流和误识别的情况,河流识别结果更加完整,但是在特别复杂的干扰因素以及更少的标签数据情况下的提取精度还有一定提升空间。

4.3.2 消融实验

为验证各个模块的有效性,从改变用于微调的标签数量、样本数据增强以及自监督对比学习与非均匀采样 3 方面进行消融实验,验证其对网络性能的影响。

(1) 微调标签数量对网络性能的影响

为了验证用于微调下游任务模型的标签数量对网络性能的影响,将标签数据按 1/10、3/10、5/10、7/10、1 的比例对下游任务模型进行微调训练,图 13 为测试集上像素准确率和交并比随着标签数量比例的不同而变化的曲线,表 2 为微调标签数量对网络

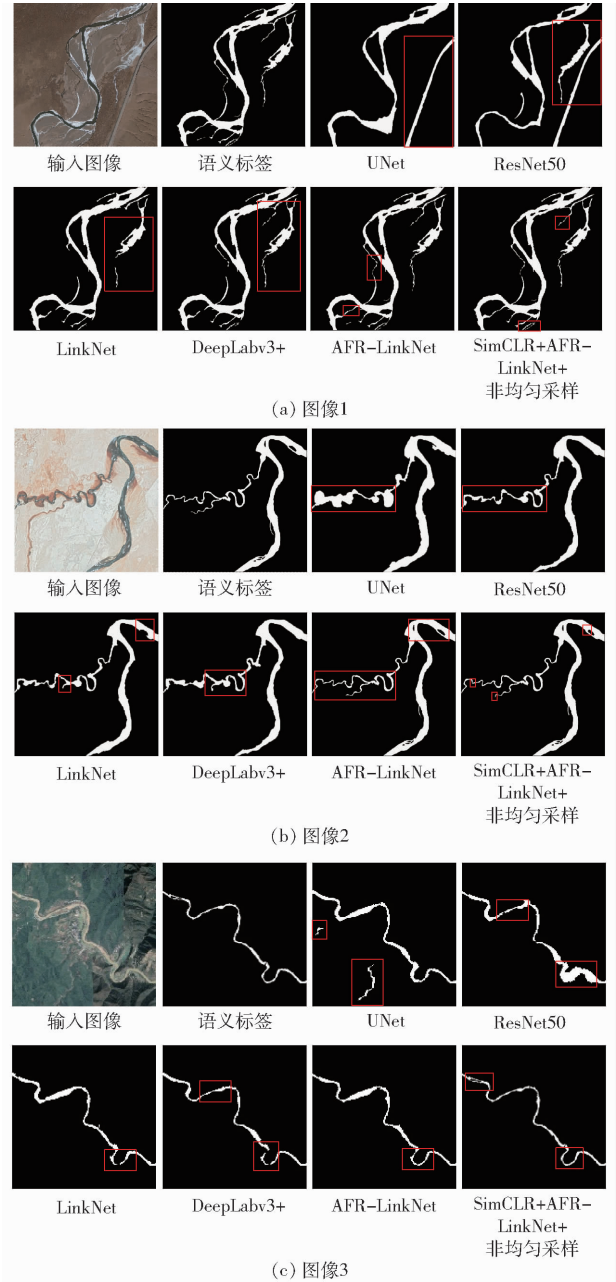


图 12 3 幅图像不同网络识别结果对比

Fig. 12 Comparisons of recognition results of three images with different networks

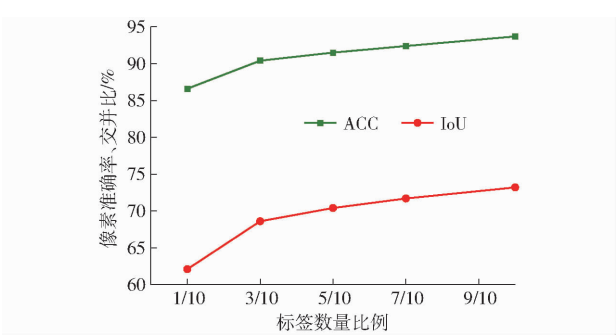


图 13 网络像素准确率与交并比变化曲线

Fig. 13 Change curves of network pixel accuracy and intersection over union

模型性能影响的各项评价指标。

表 2 不同标签数量比例下网络各项评价指标

Tab.2 Network evaluation index under different label quantity ratios

标签数量比例	像素准确率	交并比	召回率
0	72.7	51.8	70.1
1/10	86.6	62.1	78.8
3/10	90.4	68.6	83.2
5/10	91.5	70.4	85.1
7/10	92.4	71.7	86.3
1	93.7	73.2	88.5

由图 13 可知,随着微调标签数量的不断增加,网络的像素准确率和交并比均呈不断提高趋势,当标签比例由 1/10 增加为 3/10 时,提升速度最快,后面随着标签比例的增加,提升速度较慢,当所有标签数据全部用于微调模型时,像素准确率和交并比都达到最大值,此时像素准确率为 93.7%,交并比为 73.2%(表 2);当微调标签比例为 3/10 时(即使用 360 幅标签数据对模型进行微调训练),网络像素准确率、交并比和召回率分别达到 90.4%、68.6% 和 83.2%,已经和使用 1 200 幅标签数据进行有监督训练的 AFR-LinkNet 网络各项评价指标(像素准确率 90.6%、交并比 69.7%、召回率 86.4%)相当;当使用 1 200 幅全部标签数据进行微调时,其网络性能的各项评价指标相比有监督训练有了较大的提高。这验证了本文提出的自监督对比学习方式进行网络预训练,并结合非均匀采样,可以在少量有标签数据情况下使网络达到较高的河流提取效果,解决了遥感图像河流标签数据难以大量获取的问题。

(2) 样本数据增强对网络性能的影响

通过数据增强可以增加训练样本的多样性,提高模型鲁棒性,有效避免网络过拟合。图 14 为经过数据增强和没有进行数据增强的网络各项评价指标的对比,从图 14 可以看出,经过数据增强后,网络的像素准确率、召回率和交并比相较于没有经过数据增强的网络,分别提高 5.3、4.4、4.8 个百分点,这表明将样本进行数据增强能更好地提升网络性能。

(3) 自监督对比学习与非均匀采样对网络性能的影响

选用 AFR-LinkNet 网络为基准网络,使用 1 200 幅有标签数据训练有监督网络 AFR-LinkNet,同时对通过自监督对比学习方式训练的网络进行微调,并将非均匀采样与两种网络分别结合,验证其对网络性能的影响。图 15 为验证集上各网络像素准确率变化曲线,表 3 为测试集上网络对应的各项评价指标。

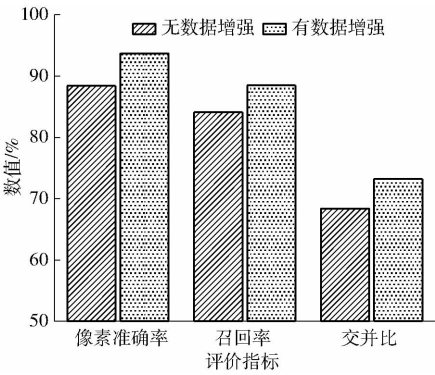


图 14 数据增强对网络评价指标的影响

Fig. 14 Influence of data enhancement on network evaluation index

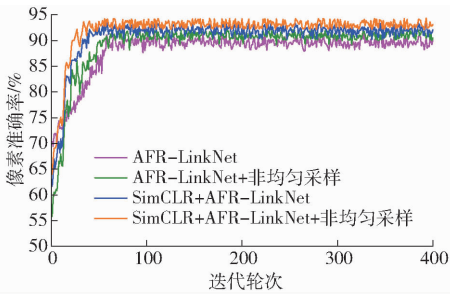


图 15 像素准确率变化曲线

Fig. 15 Change curves of pixel accuracy of each network

表 3 网络各项评价指标

Tab.3 Network evaluation indicators

网络模型	SimCLR	非均匀采样	像素准确率	交并比	召回率
AFR-LinkNet			90.6	69.7	86.4
		✓	91.5	70.8	87.3
	✓		92.1	71.9	87.8
	✓	✓	93.7	73.2	88.5

注:✓表示含有该组件。

由图 15 和表 3 可知,没有引入非均匀采样时,仅通过自监督对比学习方式预训练编码器,使得网络学习到有监督网络无法学习到的图像内在信息,网络收敛速度快于有监督的 AFR-LinkNet 网络,像素准确率达到 92.1%,提高 1.5 个百分点,交并比、召回率分别提高 2.2、1.4 个百分点;当引入非均匀采样后,有监督和自监督网络像素准确率都进一步提升,自监督网络的像素准确率达到 93.7%,相比有监督网络 AFR-LinkNet 提高 2.2 个百分点,交并比和召回率分别提高 2.4、1.2 个百分点,验证了非均匀采样的有效性和较强的泛化性能。图 16 为 3 幅相应网络的遥感图像河流分割效果,通过自监督对比方式训练并引入非均匀采样的网络,其河流提取的连续性得到提升,细小河流更加连贯,边缘细节提取也更加准确,能够有效优化河流边缘细节,在边缘容易误判处提取效

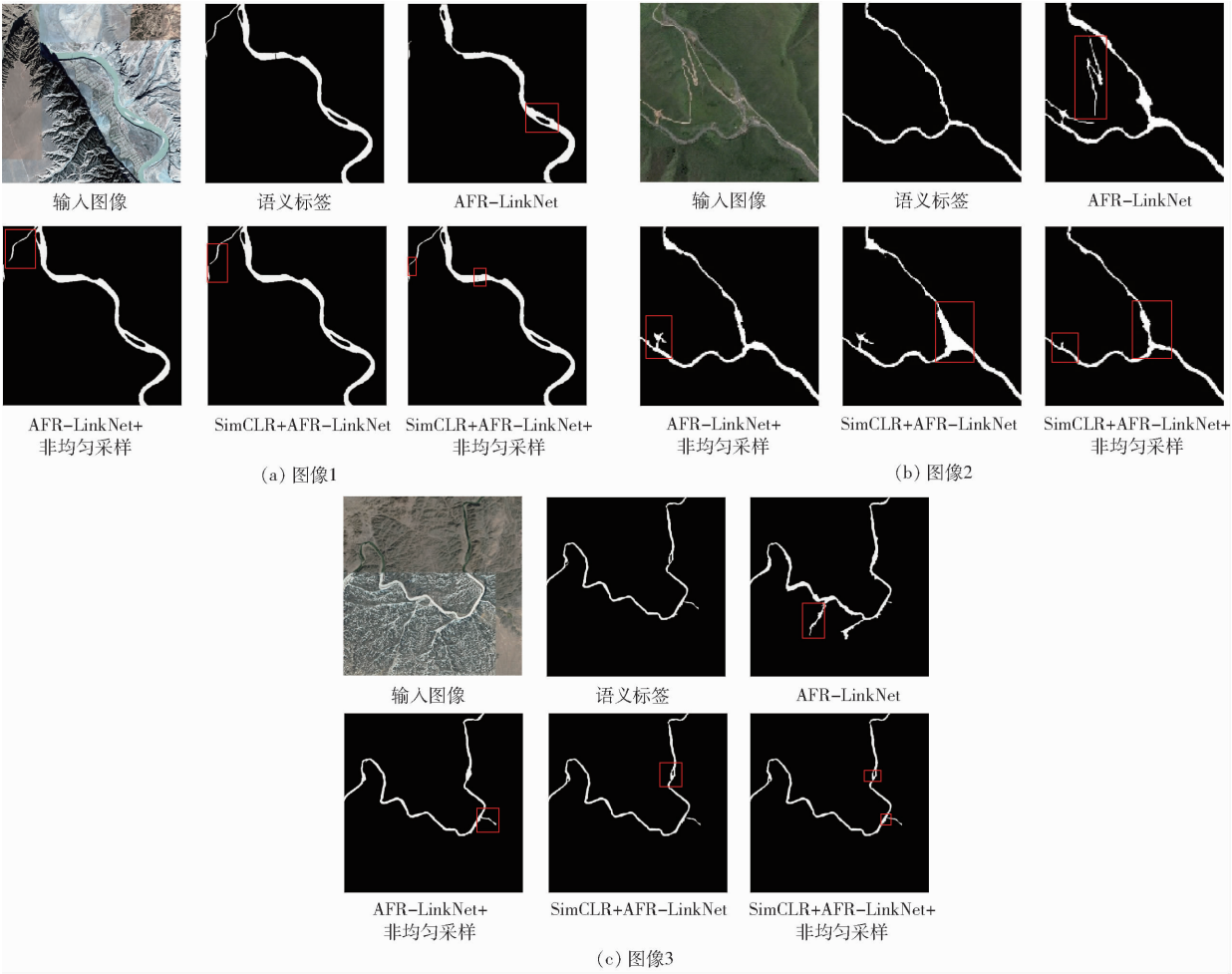


图 16 3 幅图像不同网络提取结果对比

Fig. 16 Comparisons of extraction results of three images with different networks

果也更加平滑,降低了对道路、冰雪、高山阴影的错误识别。

5 结束语

针对有监督网络训练需要大量遥感图像河流标签数据以及河流提取边缘细节分割效果不佳问题,提出一种通过自监督对比学习方式利用大量无标签数据预训练编码器,并将训练好的编码器利用迁移学习方式迁移到下游河流提取任务中,并使用少量

标签数据进行微调编码器,然后结合一种新的非均匀采样方式对寒旱区遥感河流图像进行提取的网络。实验结果表明,仅用 360 幅有标签数据就达到与使用 1 200 幅有标签数据训练的有监督网络相当的提取效果,当全部标签用于微调网络编码器时,其像素准确率、交并比和召回率分别达到 93.7%、73.2% 和 88.5%,均高于有监督网络,解决了大量标签数据难以获取的问题,并提高了河流提取精细度,证明了其有效性和泛化性。

参 考 文 献

[1] 张帆,任冲锋,蔡宴朋,等. 基于复合多目标方法的灌区水资源优化配置[J]. 农业机械学报, 2021, 52(11): 297 – 304. ZHANG Fan, REN Chongfeng, CAI Yanpeng, et al. Optimally allocating of water resources among irrigation districts based on composite multiobjective method[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(11): 297 – 304. (in Chinese)

[2] CHEN S T, LIU Y, ZHANG C. Water-body segmentation from multispectral remote sensing imagery by feature pyramid enhancement and pixel pair matching[J]. International Journal of Remote Sensing, 2021, 42(13): 5029 – 5047.

[3] 许耀文,王一涵,梁冬玲,等. 基于农业可持续发展的锦西灌区灌溉水资源配置研究[J]. 农业机械学报, 2020, 51(10): 299 – 309. XU Yaowen, WANG Yihan, LIANG Dongling, et al. Irrigation water resources allocation in Jinxi irrigation district based on agricultural sustainability[J]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(10): 299 – 309. (in Chinese)

- [4] 沈瑜,苑玉彬,彭静,等. 基于深度学习的寒旱区遥感影像河流提取[J]. 农业机械学报, 2020, 51(7): 192–201.
SHEN Yu, YUAN Yubin, PENG Jing, et al. River extraction from remote sensing images in cold and arid regions based on deep learning[J]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(7): 192–201. (in Chinese)
- [5] 沈瑜,王海龙,苑玉彬,等. 基于改进 LinkNet 的寒旱区遥感图像河流识别[J]. 农业机械学报, 2022, 53(7): 217–225.
SHEN Yu, WANG Hailong, YUAN Yubin, et al. Recognition of rivers in remote sensing images in cold and arid regions based on improved LinkNet[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(7): 217–225. (in Chinese)
- [6] 陈前,郑利娟,李小娟,等. 基于深度学习的高分遥感影像水体提取模型研究[J]. 地理与地理信息科学, 2019, 35(4): 43–49.
CHEN Qian, ZHENG Lijuan, LI Xiaojuan, et al. Water body extraction from high-resolution satellite remote sensing images based on deep learning[J]. Geography and Geo-information Science, 2019, 35(4): 43–49. (in Chinese)
- [7] WANG G, WU M, WEI X, et al. Water identification from high-resolution remote sensing images based on multidimensional densely connected convolutional neural networks[J]. Remote Sensing, 2020, 12(5): 795.
- [8] 何红术,黄晓霞,李红春,等. 基于改进 U-Net 网络的高分遥感影像水体提取[J]. 地球信息科学学报, 2020, 22(10): 2010–2022.
HE Hongshu, HUANG Xiaoxia, LI Hongga, et al. Water body extraction of high resolution remote sensing image based on improved U-Net network[J]. Geo-Information Science, 2020, 22(10): 2010–2022. (in Chinese)
- [9] YANG C, FAN R, YANG X, et al. Extraction of urban water bodies from high-resolution remote-sensing imagery using deep learning[J]. Water, 2018, 10(5): 585.
- [10] GENG Y L, TAO C, SHEN J, et al. High-resolution remote sensing image semantic segmentation based on semi-supervised full convolution network method[J]. Acta Geodaetica et Cartographica Sinica, 2020, 49(4): 499–508.
- [11] LIU K, WANG D, RONG M X. X-ray image classification algorithm based on semi-supervised generative adversarial networks[J]. Acta Optica Sinica, 2019, 39(8): 117–125.
- [12] 陶超,阴紫薇,朱庆,等. 遥感影像智能解译: 从监督学习到自监督学习[J]. 测绘学报, 2021, 50(8): 1122–1134.
TAO Chao, YIN Ziwei, ZHU Qing, et al. Remote sensing image intelligent interpretation: from supervised learning to self-supervised learning[J]. Acta Geodaetica et Cartographica Sinica, 2021, 50(8): 1122–1134. (in Chinese)
- [13] ERICSSON L, GOUK H, LOY C C, et al. Self-supervised representation learning: introduction, advances, and challenge[J]. Signal Processing Magazine, 2022, 39(3): 42–62.
- [14] SCHMARJE L, SANTAROSSA M, SCHRÖDER S M, et al. A survey on semi-, self- and unsupervised learning for image classification[J]. IEEE Access, 2021, 9: 82146–82168.
- [15] JING L, TIAN Y. Self-supervised visual feature learning with deep neural networks: a survey[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(11): 4037–4058.
- [16] LIU X, ZHANG F, HOU Z, et al. Self-supervised learning: generative or contrastive[J]. IEEE Transactions on Knowledge and Data Engineering, 2023, 35(1): 857–876.
- [17] 李巍华,何琛,陈祝云,等. 基于对称式对比学习的齿轮箱无监督故障诊断方法[J]. 仪器仪表学报, 2022, 43(3): 121–131.
LI Weihua, HE Chen, CHEN Zhuyun, et al. Unsupervised fault diagnosis method for gearbox based on symmetric contrast learning[J]. Chinese Journal of Scientific Instrument, 2022, 43(3): 121–131. (in Chinese)
- [18] LI H, LI Y, ZHANG G, et al. Global and local contrastive self-supervised learning for semantic segmentation of HR remote sensing images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 1–14.
- [19] LI W, CHEN H, SHI Z. Semantic segmentation of remote sensing images with self-supervised multitask representation learning[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2021, 14: 6438–6450.
- [20] 陈梅林,王逸舟,戴彦,等. SaSnet: 基于自监督学习的电力线实时分割网络[J]. 中国电机工程学报, 2022, 42(4): 1365–1375.
CHEN Meilin, WANG Yizhou, DAI Yan, et al. SaSnet: power line segmentation network in real time based on self-supervised learning[J]. Proceedings of the CSEE, 2022, 42(4): 1365–1375. (in Chinese)
- [21] CHEN T, KORNBLITH S, NOROUZI M, et al. A simple framework for contrastive learning of visual representations[C] // Proc. of the 37th IEEE International Conference on Machine Learning, 2020: 1597–1607.
- [22] JUNG H, OH Y, JEONG S, et al. Contrastive self-supervised learning with smoothed representation for remote sensing[J]. IEEE Geoscience and Remote Sensing Letters, 2022, 19: 1–5.
- [23] WALTER K, GIBSON M J, SOWMYA A. Self-supervised remote sensing image retrieval[C] // 2020 IEEE International Geoscience and Remote Sensing Symposium, 2020: 1683–1686.
- [24] JAISWAL A, BABU A R, ZADEH M Z, et al. A survey on contrastive self-supervised learning[J]. Technologies, 2021, 9(1): 1–22.
- [25] HE K, FAN H, WU Y, et al. Momentum contrast for unsupervised visual representation learning[C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 9729–9738.
- [26] HE K, ZHANG X, REN S, et al. Delving deep into rectifiers: surpassing human-level performance on imagenet classification[C] // Proceedings of the IEEE International Conference on Computer Vision, 2015: 1026–1034.