

<https://doi.org/10.3799/dqkx.2021.176>



基于多模态特征融合的地质体识别方法

付 偲^{1,3}, 李超岭², 张海燕^{1,3*}, 刘 畅², 李丰丹²

1. 北京林业大学信息学院, 北京 100083
2. 中国地质调查局发展研究中心, 北京 100037
3. 国家林业草原林业智能信息处理工程技术研究中心, 北京 100083

摘要: 将深度学习技术应用于地质填图, 挖掘不同模态数据深层次信息, 可以实现更为准确的地质填图. 提出了一种基于多模态特征融合的地质体识别方法, 综合考虑地球物理、地球化学数据和遥感影像数据, 该方法先利用深度神经网络和卷积神经网络分别提取这两种不同模态数据的特征, 然后进行特征拼接得到多模态特征, 最后利用全连接神经网络进行特征融合完成地质体分类. 交叉验证结果表明, 提出的多模态特征融合方法比仅使用地球物理地球化学数据或遥感影像数据的深度学习方法相比有明显的优势, 在分类准确率上分别提升了 14.08% 和 2.79%, 证明了该方法可以实现更为精准的地质体识别, 进而更好地辅助地质填图.

关键词: 地质体识别; 地质填图; 多模态特征融合; 深度学习; 矿产勘察.

中图分类号: P623

文章编号: 1000-2383(2023)10-3743-10

收稿日期: 2021-07-02

Geological Body Recognition Based on Multi-Modal Feature Fusion

Fu Si^{1,3}, Li Chaoling², Zhang Haiyan^{1,3*}, Liu Chang², Li Fengdan²

1. School of Information Science & Technology, Beijing Forestry University, Beijing 100083, China
2. Development Research Center, China Geological Survey, Beijing 100037, China
3. Engineering Research Center for Forestry-Oriented Intelligent Information Processing of National Forestry and Grassland Administration, Beijing 100083, China

Abstract: Applying deep learning technology to geological mapping to mine the deep-level information of different modal data, so as to achieve more accurate geological mapping. Considering the geophysical and geochemical data and remote sensing image data, in this paper it proposes a geological body recognition method based on multi-modal feature fusion. Firstly, Using deep neural network and convolution neural network to extract the features of the two different modal data and then performs feature splicing to obtain multi-modal features, finally, the fully connected neural network is used for feature fusion to complete the geological body classification. The cross-validation results show that the proposed multi-modal feature fusion method has obvious advantages compared with the deep learning methods using geophysical and geochemical data or remote sensing image data along, and the classification accuracy rate is increased by 14.08% and 2.79%. This result proves that this method can realize more accurate geological body identification, and then better assist geological mapping.

Key words: geological body recognition; geological mapping; multi-modal feature fusion; deep learning; mineral exploration.

基金项目: 中国地质调查局地质调查项目(No.DD20190416).

作者简介: 付偲(1995—), 女, 硕士研究生, 主要从事人工智能方面的研究. ORCID:0000-0001-7760-9752. E-mail: fusi1995@bjfu.edu.cn

* **通讯作者:** 张海燕, ORCID:0000-0002-8629-2838. E-mail: zhyzml@bjfu.edu.cn

引用格式: 付偲, 李超岭, 张海燕, 刘畅, 李丰丹, 2023. 基于多模态特征融合的地质体识别方法. 地球科学, 48(10):3743-3752.

Citation: Fu Si, Li Chaoling, Zhang Haiyan, Liu Chang, Li Fengdan, 2023. Geological Body Recognition Based on Multi-Modal Feature Fusion. *Earth Science*, 48(10):3743-3752.

1 研究现状

地质填图是指在野外实地观察研究的基础上,按一定比例尺将各种地质体和地质现象填绘在地理底图上而构成地质图的工作过程.传统的地质填图方法是通过野外观察来完成的,由于地质路线所穿越的地质体不可能对地质体的边界全部控制,地质体的大部分边界是通过一定的规则预测连接的,其连接的准确性、合理性和精度,受地质人员的主观和专业水平影响比较大.虽然随着计算机技术的发展,数字地质填图技术应运而生,但在地质对象识别技术方法上,除人工地质路线外,仍然保持在航空像片和遥感影像地质解译技术方法上,无论是航空像片还是遥感影像,都只能反映地质对象所在区域的光谱及地质构造信息,由于地质信息单一,其应用能力容易受到各方面的影响和约束.

近年来,深度学习在各个领域都展现出独特的优势(Chen *et al.*, 2016; Cheng and Guo, 2017; Zhang *et al.*, 2018; Ran *et al.*, 2019; Su *et al.*, 2019; Wang *et al.*, 2019; Zheng *et al.*, 2019; Zhou *et al.*, 2020).在地质领域,一些深度学习方法开始用于处理地质数据及遥感影像数据,应用神经网络进一步挖掘提取数据信息,在成矿预测、岩相判别及地质填图(徐永洋等, 2020; Anderson *et al.*, 2020; Pires de Lima *et al.*, 2020; 高松, 2020; 左仁广, 2021; 左仁广等, 2021; Singer, 2021)等方面都取得了突出的成果.

模态是指事物发生或存在的方式,多模态是指两个或者两个以上模态各种形式的组合(刘建伟等, 2020).深度学习中的多模态融合技术(Multi-modality Fusion Technology, MFT)是模型在分析和识别任务时处理不同形式数据的过程,多模态数据的融合可为模型决策提供更多信息,从而提高决策总体结果的准确率(Zhang *et al.*, 2018; Bhandari *et al.*, 2019; Maimaitijiang *et al.*, 2019; Su *et al.*, 2019; Zheng *et al.*, 2019; Zhou *et al.*, 2020; 何俊等, 2020).地球物理及地球化学数据以不同方式反映了不同地质对象地质特征中的物理和化学特征,不同传感器的遥感影像数据反映了不同地质对象所在区域的光谱及地质构造信息,综合考虑这两种不同模态数据可以获取更多的地质信息,有助于提高地质对象识别能力.

Brandmeier and Chen(2019)通过 Sentinel-2 与

ASTER 数据联合使用,利用卷积神经网络对 Mount Isa 地区(澳大利亚)7种岩性进行预测(填图),预测准确度可以达到 75%.Arabi Aliabad *et al.*(2019)利用陆地卫星 7 号影像,用模糊 ARTMAP 人工神经网络绘制 Yazd-Shirkoooh 流域局部地质图,并结合现实 Google Earth 地图,通过比较,实现对已有 1:100 000 比例尺的地质调查地质图进行修正.其中模糊 ARTMAP 人工神经网络模型 11 个类别的分类精度可达到 89%(Kappa 系数).Arabi Aliabad *et al.*(2019)使用遥感光学影像进行地质填图,未考虑地质信息对地质填图的作用,Brandmeier and Chen(2019)虽然尝试同时使用地球物理数据和遥感影像数据进行岩性测绘,但是未考虑两种数据的差异,采用了相同的特征提取方式,反而干扰了测绘结果,使得预测准确度明显下降.

综上所述,本文设计了一种基于多模态特征融合的地质体识别模型,综合考虑待填图区域的地球物理地球化学数据和遥感影像数据,使用深度神经网络(deep neural networks, DNN)和卷积神经网络(convolutional neural networks, CNN)分别提取两种不同模态数据的数据特征,然后将两种特征向量进行特征拼接得到多模态特征,最终通过全连接神经网络完成特征融合并进行地质体分类输出,该模型能实现准确的地质体分布预测,进而更好地辅助地质填图.

2 研究区介绍

研究区地处内蒙古自治区东部,东北接兴安盟科尔沁右翼中旗,南邻扎鲁特旗政府所在地鲁北镇,向西南毗邻赤峰市阿鲁科尔沁旗,向西北可达煤炭基地霍林郭勒市及乌拉盖草原景区.

研究区属低山丘陵区,海拔高度一般在 300~600 m,最高峰 748.8 m,相对高差平均 200 m,为浅切割地形区,多数山顶较浑圆,山谷开阔,牧草及灌木丛生,基岩裸露一般,部分为掩盖区.

该研究区(1:50 000 地质填图)的主要目标任务是重点查明扎鲁特盆地北部晚古生代及中生代地层层序特征及时空演化规律;查明火山岩岩相及火山构造特征,探讨构造岩浆作用及其与成矿作用的关系;查明与探讨环境演化及生态景观变迁特征.目前,工作区内共划分 14 个岩石地层单位和 7 个成因类型填图单位,涉及 231 种不同填图单位的岩性.

3 多模态特征融合的地质体识别主要方法与流程

3.1 样本制作

为了提升模型的准确率和泛化能力,监督学习型的深度神经网络通常都需要足量的标签样本,如深度神经网络模型常用的数据集 MNIST (LeCun *et al.*, 1998)、ImageNet (Krizhevsky *et al.*, 2012; Rusakovsky *et al.*, 2015)、CIFAR (Krizhevsky and Hinton, 2009) 等图片数量均已达到上万张。

在地质填图中,地质路线的地质点、分段路线、点和点间界线描述数据是识别地质体的主要数据和知识(在数字填图中简称 PRB 数据)。为了充分学习野外路线的知识,即地质对象的标签,本试验把地质分界点转成地质路线上的点,在分段路线上的中点采一个点,如果分段路线大于 105 m,则每 105 m 采一个点,地质属性逻辑上隶属地质界线的属性之一,连同加上地质岩性控制点、地质样品点等一同构成了地质路线岩石地层单位及岩性的 PRB 的标签。根据 1:5 万地质填图技术要求,各类地质点在地质手图上所标定的点位与实地位置误差一般不得大于 50 m,故以定点为中心、半径 50 m 范围内的采样点可视为同一地质体。因此,本实验对研究区中 PRB 采样点以 50 m 为半径进行样本扩充得到实验所需的 PRB 样本点。表 1 列出了部分地质体 PRB 采样点及 PRB 样本点数量。

3.2 坐标转换

实验数据中,PRB 样本点的位置坐标是真实地理坐标,地球物理地球化学(地物地化)数据和遥感影像数据的位置坐标是栅格坐标,因此,为了获取

PRB 样本点的地物地化数据和遥感影像数据,需要对真实地理坐标进行坐标转换,转换公式如下:

$$W = \left\lfloor \frac{X - X_0}{\text{Ratio}_X} \right\rfloor, \quad (1)$$

$$H = \left\lfloor \frac{Y - Y_0}{\text{Ratio}_Y} \right\rfloor, \quad (2)$$

式中:(X, Y)代表 PRB 样本点的真实地理坐标, (X_0, Y_0)代表地物地化数据和遥感影像数据的起始地理坐标, Ratio_X 和 Ratio_Y 分别为地物地化数据和遥感影像数据在 X 轴和 Y 轴方向的分辨率值, (H, W)代表公式转换后的栅格坐标。

3.3 数据归一化

归一化即将数据映射为 $[0, 1]$ 或者 $[-1, 1]$ 之间的小数,将有量纲数据转换为无量纲数据,消除数据之间的量纲影响,使不同单位的数据具有可比性。因为地物地化数据为研究区中 PRB 样本点的不同化学元素的含量值与高精磁测 ΔT 磁异常值,而遥感影像数据为不同传感器拍摄的研究区内的影像数据,两者属于不同量纲的数据。因此对地物地化数据和遥感影像数据进行归一化预处理。本实验使用的归一化公式如下:

$$\tilde{x} = \frac{x - x_{\min}}{x_{\max} - x_{\min}}, \quad (3)$$

式中: x 表示样本点对应的地物地化数据或遥感影像数据, x_{\min} 代表研究区中相应的地物地化数据或遥感影像数据的最小值, x_{\max} 代表研究区中相应的地物地化数据或遥感影像数据的最大值。

3.4 多模态特征融合地质体识别模型设计

实验采用地物地化数据和遥感影像数据两种不同模态的数据。遥感影像数据与图像数据类似,具有局部相关性,CNN 能够有效提取数据间的局部相关性,因此应用 CNN 对以 PRB 样本点为中心一定范围内的遥感影像数据进行特征提取;地物地化数据是每个 PRB 样本点单点数据,考虑到地物地化数据是离散点数据,因此应用 DNN 对样本点的地物地化数据进行特征提取。综上所述,本文提出了一种基于多模态特征融合的地质体识别模型(如图 1 所示),首先,对地物地化数据和遥感影像数据进行数据预处理;然后分别使用 CNN 和 DNN 进行特征提取;之后将两种数据特征进行特征拼接得到多模态特征向量,然后通过 DNN 实现特征融合;最后,使用 Softmax 分类器进行分类输出。

3.4.1 特征提取 本实验截取以样本点为中心、半径 25 m 的外接矩形对应的遥感影像作为该样本点

表 1 PRB 采样点及样本点数量

Table 1 The number of PRB sampling points and PRB sample points

地质体	PRB 采样点	PRB 样本点
K1b 流纹质晶玻屑凝灰岩	13	5 200
J3mn 安山质熔结角砾岩晶屑凝灰岩	14	5 600
J3m 流纹质角砾晶屑玻屑凝灰岩	15	6 000
J3m 流纹质晶玻屑凝灰岩	16	6 400
K1b 流纹质角砾岩熔岩	17	6 800
J3m 石泡流纹岩	17	6 800
J3mn 角闪安山岩	18	7 200
J3m 含角砾流纹岩	20	8 000
Qh1eol 土黄色风成砂	21	8 400
J3m 流纹斑岩	24	9 600

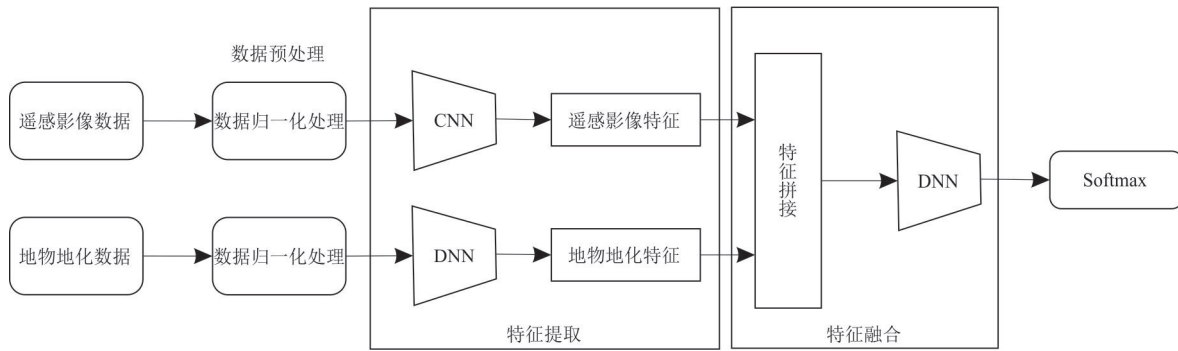


图 1 多模态特征融合地质体识别模型整体结构

Fig.1 The overall structure of multi-modal feature fusion geological body recognition model

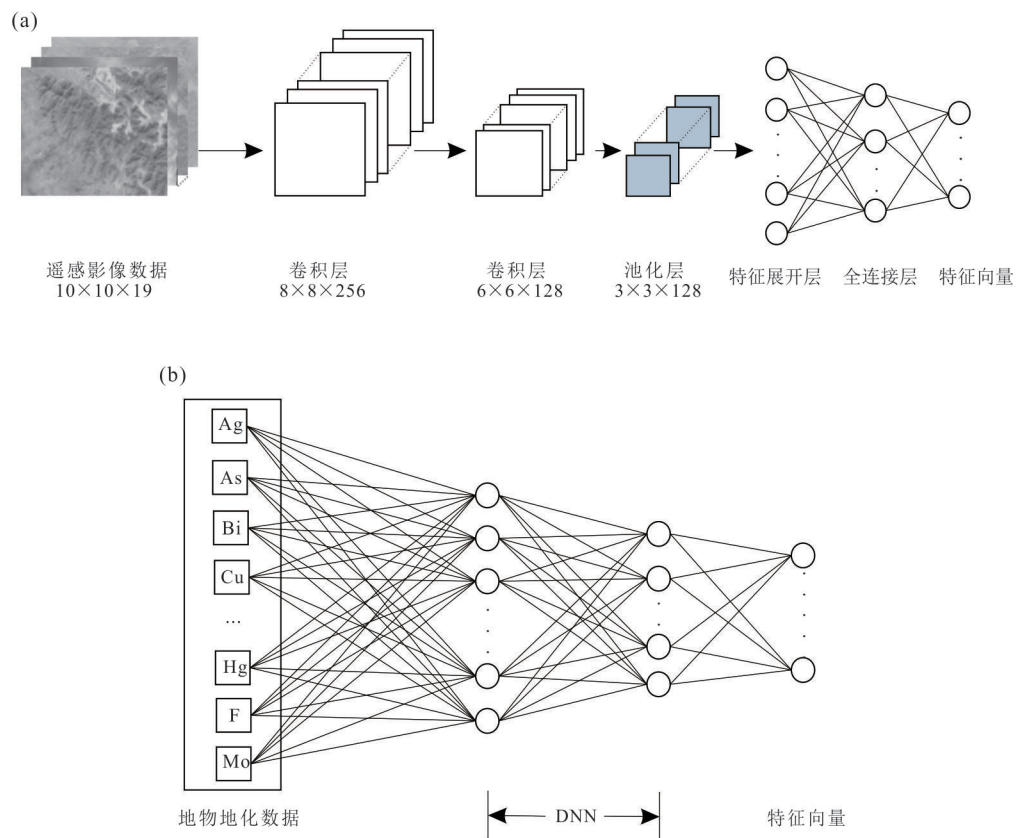


图 2 特征提取网络结构

Fig.2 Feature extraction network structure

a. 遥感影像数据特征提取; b. 地物地化数据特征提取

的遥感影像数据,每个样本点对应的单波段遥感影像数据像素大小为 10×10 . 针对遥感影像数据,本实验设计卷积神经网络提取数据特征,卷积神经网络由两层卷积、一层池化及一层全连接组成,具体结构如图 2a 所示. 已有研究表明,卷积神经网络中 3×3 的卷积核能减小模型参数量同时有效增加模型对数据特征的学习能力 (Simonyan and Zisserman, 2015), 因此实验中卷积核大小设置为 3×3 , 卷

积核个数分别为 256、128,卷积核移动步长均为 1. 平均池化操作通常聚焦图像数据的背景信息,最大池化操作通常聚焦图像数据的纹理信息,因此实验选择最大池化以保留数据特征中的纹理信息. 三层卷积、池化操作后,网络通过特征展开输出 1 152 维特征向量,最后通过一层全连接进行更深层次的特征提取,得到 128 维特征向量.

为了获取地物地化数据的高阶特征表达,本实验设计 3 层 DNN 提取数据特征,具体结构如图 2b 所示.其中,3 层全连接层的神经元个数分别为 256、128、128,最终生成 128 维特征向量.

为了增强网络的非线性表达能力,本实验中所有卷积层、全连接层后的激活函数都选用整流线性单元(rectified linear unit, ReLU).

3.4.2 特征融合 本实验中,遥感影像数据反映了样本点附近一定范围内的空间信息与光谱信息,地物地化数据反映了样本点的地球物理与地球化学元素分布信息,两种数据从不同方面反映了样本点所在位置的地质体特征,通过特征融合的方式整合两种数据的信息,从而提高地质体识别的准确率.

特征融合是将多个不同的特征向量通过特定的映射输出一个新的特征向量.特征融合方法主要包含以下 3 种:①按照人工规则对不同特征向量进行线性融合;②计算不同特征向量的相似矩阵,根据相似度进行融合;③直接进行特征拼接,然后通过神经网络进行特征融合.其中,基于神经网络的特征融合方法能够自动学习特征权重,衡量不同特征的重要程度,更好地实现特征融合.因此,本实验使用基于神经网络的特征融合方法,特征融合过程如图 3 所示. N 维遥感影像特征 $[x_1, x_2, \dots, x_n]$ 与 M 维地物地化数据 $[y_1, y_2, \dots, y_m]$ 进行特征拼接,得到 $N+M$ 维特征向量,然后输入全连接层进行特征融合,该全连接层神经元个数为 256.

3.4.3 损失函数 实验采用分类问题中常用的交叉熵损失函数来对模型参数进行优化,并在交叉熵损失函数的基础上使用 L2 正则化防止模型在训练过程中发生过拟合现象,增强模型的泛化能力.损失函数公式如下:

$$\text{Loss} = -\sum_{i \in C} \hat{y}_i \log(y_i) + \lambda \|\theta\|^2, \quad (4)$$

式中: C 为分类类别, \hat{y}_i 为真实类别, y_i 为预测类别, $\|\theta\|^2$ 为 L2 正则项, λ 为正则项系数.

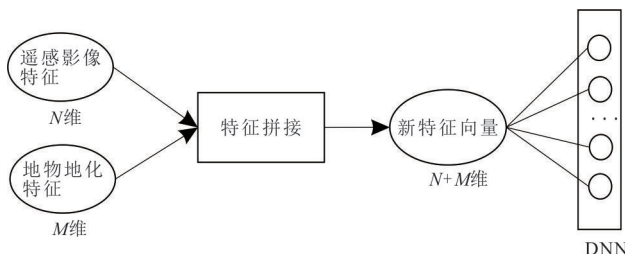


图 3 特征融合网络结构

Fig.3 Feature fusion network structure

4 具体应用实例效果与评价分析

4.1 实验数据集

本实验已知研究区中的 231 类地质体(指填图单位+岩性)总计 3 953 个 PRB 采样点信息,在保证样本点遥感影像区域内标签一致性的前提下进行半径 50 m 的样本扩充,扩充后得到 395 060 个样本点.表 2 展示了部分 PRB 采样点信息,表中色标代表地质体在预测图中的 RGB 值.

本实验使用的数据为研究区中样本点的地物地化数据和遥感影像数据(表 3).其中,地物地化数据包含 1:20 万地球化学数据(As、Hg 等元素)、1:5 万地球化学数据(Ag、Cu 等元素)和地面高精度磁测数据;遥感影像数据包含高分三号卫星(GF-3)、高分一号卫星(GF-1)、资源三号卫星(ZY-3)、Landsat8 卫星及数字高程模型(DEM)数据,Landsat8 卫星数据采用 4 种不同波段组合数据,波段组合分别为 752、753、432、632(波段组合中的数字代表波段号).

高分三号卫星是中国高分专项工程的一颗遥感卫星,是由中国航天科技集团公司研制的 C 频段多极化合成孔径雷达(synthetic aperture radar, SAR)成像卫星.合成孔径雷达具有不受光照和气候限制,全天候、全天时工作的特性(涂宽等, 2019),对地表覆盖区有一定的穿透性,能穿透地表浅覆盖层,获取覆盖层以下的地质特征(Pour *et al.*, 2016;涂宽等, 2019).因此,它可以弥补光学遥感无法获取浅覆盖层以下的地质构造信息的缺陷 Carter and Clift (2008).

DEM 是对研究区地形的数字化模拟,是用一组有序数值阵列表示地面高程的一种实体地面模型(李晨伟等, 2018).DEM 数据可以进一步补充地质体的线性构造信息,有利于不同地质体的区分.

4.2 实验环境和参数设置

本研究主要的硬件环境为 Intel Core™ i7-6700 CPU(64GB 内存)和 TITAN RTX(24 GB 显存),软件环境为 Ubuntu18.04 LTS 64 位操作系统,模型框架选择使用 TensorFlow 后端的 Keras.模型优化器选择带动量的随机梯度下降(stochastic gradient descent, SGD),batchsize 设置为 1 024,迭代次数 epoch 设置为 2 000 次,学习率设置为 10^{-3} ,学习率衰减系数为 6, L2 正则化系数为 10^{-3} .

表 2 PRB 采样点信息

Table 2 PRB sampling point information

GEOPOINT	采样点类型	XX	YY	填图单位+岩性	色标
DJ126	分段路线	325 660.43	4 976 502.71	$\xi\pi$ 正长斑岩	255,0,0
DK111	B 标本	326 438.48	4 961 608.46	$\lambda\pi$ 流纹斑岩	255,0,0
DT534	岩性控制点	310 839.70	4 956 034.54	$\eta\sigma\kappa$ K1 石英二长斑岩	255,64,25
DL111	b 薄片	324 676.34	4 956 893.87	$\delta\mu$ 闪长玢岩	0,255,255
DS352	岩性控制点	310 212.49	4 962 644.40	Qp3w 乌尔吉组风成砂土	255,255,128
DJ112	分段路线	329 900.84	4 975 344.51	Qh3eol 风成砂土	255,255,181
DP049	分段路线	317 272.82	4 957 868.69	P3l 岩屑细砂岩	242,216,67
DJ541	岩性控制点	326 278.26	4 971 615.54	P3l 片岩	242,216,81
DJ014	B 标本	330 298.39	4 968 650.78	P3l 构造角砾岩	105,91,7
DO5251	分段路线	319 140.89	4 965 597.51	J3m 流纹质浆屑角砾凝灰岩	178,254,242

表 3 地物地化数据及遥感影像数据具体类别

Table 3 Specific categories of geophysical data, geochemical data and remote sensing image data

数据类型	分辨率	波段/元素数
1:20 万地球化学	5 m×5 m	39
1:5 万地球化学	5 m×5 m	12
地面高精度磁测	5 m×5 m	1
高分三号卫星	10 m×10 m	3
高分一号&资源三号卫星	2 m×2 m	3
数字高程模型	30 m×30 m	1
Landsat8 卫星	30 m×30 m	3*4

4.3 模型评价指标

考虑到研究区中地质体类别数较多,属于深度学习多分类问题,因此选择多分类问题中常用的准确率(accuracy, ACC)及 F1 分数(F1-score)作为模型评价指标从不同角度进行评估.各评价指标的计算如公式(5)~(9)所示:

准确率 ACC 为所有测试集样本点中模型预测正确的样本点的占比,主要用来衡量模型整体的有效性:

$$ACC = \frac{\text{Num}_{\text{correct}}}{\text{Num}_{\text{total}}} \quad (5)$$

F1-score 是统一精确率(precision, P)和召回率(recall, R)的一个综合指标,可以避免精确率和召回率相差较大的极端情况,较好地反映整体结果.其中,精确率 P 代表预测为类 c 的样本点中标签为 c 的样本点占比;召回率 R 代表标签为类 c 的样本点被模型正确预测的样本点占比.由于本实验地质体类别较多且各类地质体参与模型训练的样本点数不同,因此,将各类地质体的 F1-score 进行加权平均后得到的值(Weighted-F1)作为最终的 F1-score:

$$P = \frac{TP}{TP + FP} \quad (6)$$

$$R = \frac{TP}{TP + FN} \quad (7)$$

$$F1 = \frac{2 \cdot P \cdot R}{P + R} \quad (8)$$

$$\text{Weighted-F1} = \sum_{i \in C} \frac{\text{Num}_i \cdot F1_i}{\text{Num}_{\text{total}}} \quad (9)$$

式中: $\text{Num}_{\text{correct}}$ 代表所有预测正确的测试集样本点数, $\text{Num}_{\text{total}}$ 代表测试集样本点总数,TP代表每类预测正确的测试集样本点数,FP代表模型预测为类 c 但是实际不是类 c 的样本点数,FN代表每类预测错误的测试集样本点数, C 为分类类别, Num_i 代表第 i 类地质体的测试集样本点数, $F1_i$ 代表第 i 类的 F1 分数.

4.4 实验结果及分析

为了保证已知的 PRB 采样点都能作为测试集进行实验,从而更好地体现模型的准确率与泛化能力,本研究采用 K 折交叉验证的方法进行实验.将扩充后的样本点随机平均分为 5 组,每次选择其中 4 组作为训练集,另外 1 组作为测试集,进行 5 次实验,最终将 5 次实验结果的平均值作为最终的实验结果.

4.4.1 遥感影像对比实验 在仅使用遥感影像数据的实验过程中,使用不同的遥感影像数据组合进行地质体识别,具体遥感影像数据组合使用情况如表 4 所示.从表中的实验结果可以看出,SAR 数据与 DEM 数据的加入更有利于地质体识别,因此实验中遥感影像数据选择加入 GF-3 与 DEM 数据.

为了确定每个样本点最佳的单波段遥感影像数据范围,实验中选择不同单波段遥感影像数据范围进行对比实验.表 5 记录的是使用不同单波段遥感影像数据范围的卷积神经网络模型的参数量,每轮模型训练时间以及每组测试集预测时间,表 6 记录的是

表 4 不同遥感影像数据组合对比实验结果

Table 4 Comparison of experimental results of different remote sensing image data combinations

模型名称	数据集	ACC	F1-score
卷积神经网络	GF-3+GF-1+ZY-3+Landsat8	91.05%	91.09%
卷积神经网络	GF-1+ZY-3+Landsat8+DEM	92.40%	92.59%
卷积神经网络	GF-3+GF-1+ZY-3+Landsat8+DEM	93.42%	93.52%

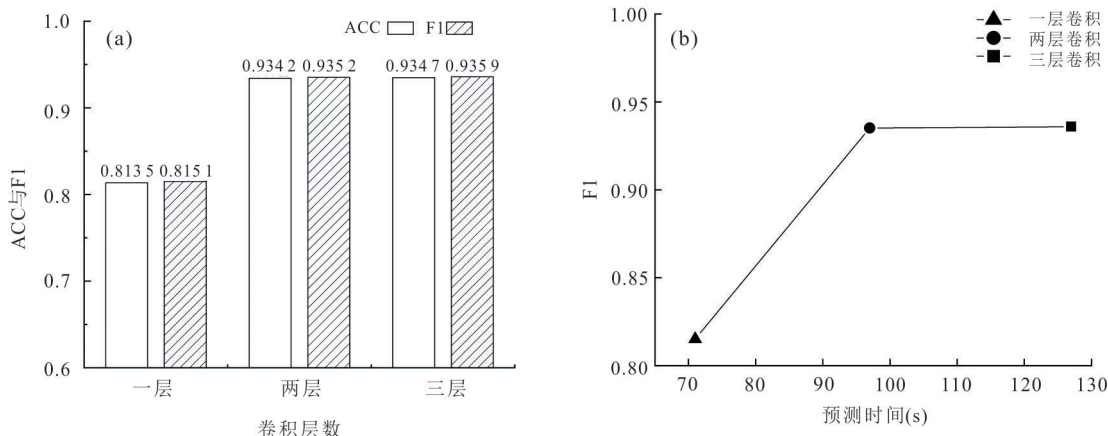


图 4 卷积层数对比实验结果

Fig.4 Convolution layer number comparison experiment results

a. 不同卷积层数ACC与F1值; b. 不同卷积层数预测时间与F1值

单波段遥感影像数据范围分别为 5×5 、 10×10 及 15×15 的对比实验结果. 其中, 遥感影像数据范围指单个样本点对应的单波段遥感影像数据像素大小.

根据表 6 的实验结果可以发现, 遥感影像数据范围为 10×10 相较于遥感影像数据范围为 5×5 在 ACC 与 F1 值上分别提升了 9.62%、9.79%, 说明遥感影像数据范围越大, 包含的遥感影像数据提供的信息量越多, 更有利于不同地质体的区分, 使得地质体识别模型的精度提高. 但是, 遥感影像数据范围也不是越大越好, 从表 6 的实验结果中可以发现, 遥感影像数据范围为 15×15 相较于遥感影像数据范围为 10×10 在 ACC 与 F1 值上仅提升了 0.10% 与 0.08%, 但是参数量过多且模型训练时间过长. 综上所述, 单个样本点对应的单波段遥感影像数据像素大小选择 10×10 .

4.4.2 卷积层数对比实验 为了验证卷积神经网络结构的有效性与可靠性, 实验对比了不同的卷积层数对网络预测精度的影响(表 7). 从实验结果(图 4a)可以发现, 两层卷积相较于一层卷积, ACC 与 F1 值分别提升了 12.07% 与 12.01%, 而三层卷积相较于两层卷积, ACC 与 F1 值仅提升了 0.05% 与 0.07%, 但是每组测试集预测时间增长了 30.93%

表 5 不同遥感影像数据范围的模型参数量、计算及预测时间

Table 5 The amount of model parameters, calculation and prediction time for different remote sensing image data ranges

遥感影像数据范围	参数量	训练时间 (s/轮)	预测时间 (s)
5×5	197 223	1	62
10×10	516 455	2	97
15×15	778 599	3	139

表 6 不同遥感影像数据范围对比实验结果

Table 6 Comparison of experimental results of different remote sensing image data ranges

遥感影像数据范围	ACC	F1-score
5×5	83.56%	83.52%
10×10	93.18%	93.31%
15×15	93.28%	93.39%

(图 4b). 因此, 根据对比实验的结果, 最终选择两层作为卷积神经网络的卷积层数.

4.4.3 消融对比实验 在确认了每个样本点最佳的单波段遥感影像数据范围后, 将卷积神经网络模型、深度神经网络模型与提出的多模态特征融合模型进行消融实验对比, 具体的模型和数据使用情况

如表 8 所示. 其中, 深度神经网络模型和卷积神经网络模型是单独使用地物地化数据或者遥感影像数据, 多模态特征融合模型同时使用地物地化数据和遥感影像数据.

3 种模型的实验结果如表 8 所示, 从实验结果可以发现, 使用多模态特征融合之后, 模型性能与仅用 DNN 模型或者 CNN 模型的结果对比, ACC 提升了 14.08%、2.79%, F1 值提升了 14.42%、2.69%. 结果表明, 地物地化数据特征和遥感数据特征之间有一定的互补关系, 在两者进行特征融合后可以更为精准地识别地质体类型, 使得模型的性能得到进一步的提升.

另外, 精确率越高表明模型对相应地质体有更好的识别能力, 表 9 记录了部分地质体在卷积神经

网络模型和多模态特征融合模型下的精确率, 结果表明, 地物地化数据的加入, 可以提供更多的地质信息, 有助于区分不同地质体.

为验证多模态特征融合模型的识别效果, 本实验对研究区局部区域的地质体进行了预测(图 5), 通过预测图与实测图对比可以看出, 绝大部分地质体分类及形态展现出了较好的一致性. 但是, 由于研究区属于浅覆盖地区, 使得部分实测地质界线点定点为大致位置(部分误差会超过 100 m), 导致预测图中有部分地质体出现分类与实测图空间位置矛盾的情况, 需要野外验证和进一步完善.

5 结论

(1) 基于多模态特征融合的地质体识别方法采用了多尺度(不同精度、不同比例尺)、多模态的地球物理、地球化学及遥感影像数据作为 PRB 标签整体样本的输入, 不但加大了标签样本数据的多样性, 同时对获取更多的地质信息进而提升地质对象识别能力更为有效;

(2) 通过不同遥感影像数据范围对比实验以及卷积层数对比实验, 确定 PRB 样本点的遥感影像数据范围、CNN 模型结构等, 最终构建了一个基于多模态特征融合的地质体识别模型. 该模型综合考虑待填图区域的地球物理、地球化学数据和遥感影像数据, 利用深度神经网络与卷积神经网络分别提取两种不同模态数据的特征, 最后利用全连接神经网络将特征拼接后的多模态特征进行特征融合来实现准确的地质体识别, 进而更好地辅助地质填图;

表 7 不同卷积层数的卷积核参数

Table 7 Convolution kernel parameters of different convolution layer numbers

模型名称	卷积层数	卷积核大小	每层卷积核个数
卷积神经网络	1	3×3	[256]
卷积神经网络	2	3×3	[256, 128]
卷积神经网络	3	3×3	[256, 128, 64]

表 8 各模型数据使用情况及实验结果

Table 8 Data usage of each model and experimental results

模型名称	地物地化数据	遥感影像数据	ACC	F1-score
深度神经网络	√	×	82.13%	81.79%
卷积神经网络	×	√	93.42%	93.52%
多模态特征融合模型	√	√	96.21%	96.21%

表 9 不同模型精确率

Table 9 Precision of different models

地质体	卷积神经网络	多模态特征融合模型
K1b 流纹斑岩	84.45%	91.52%
J3m 沉凝灰岩	86.06%	90.27%
P2z 粉砂质板岩	86.35%	94.05%
P3l 中细粒长石岩屑砂岩	86.97%	91.38%
J3m 粗安岩	87.12%	93.39%
K1b 流纹质角砾晶屑玻屑凝灰岩	87.92%	95.43%
δμ 闪长玢岩	88.16%	93.09%
K1m 角闪辉石安山岩	88.77%	94.50%
J3m 流纹质含角砾岩屑晶屑熔结凝灰岩	88.83%	92.96%
J3m 流纹质沉角砾凝灰岩	94.99%	98.75%

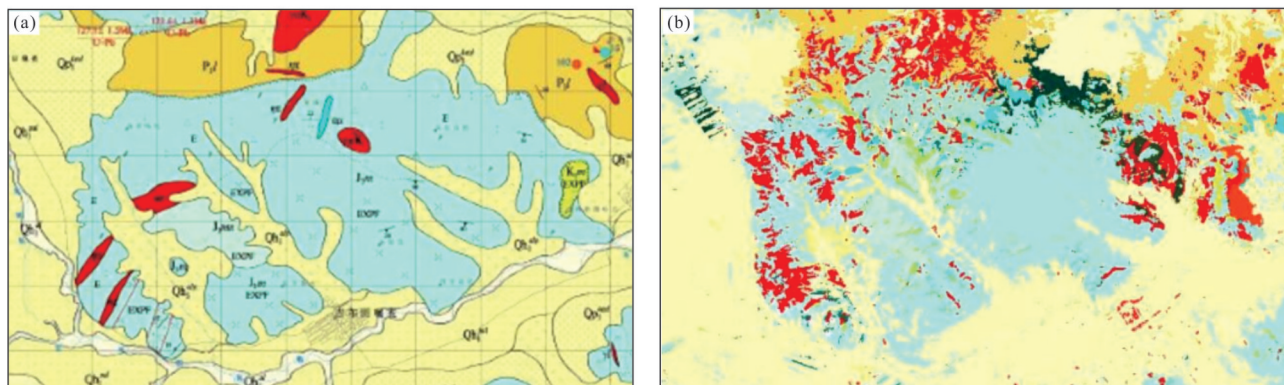


图 5 研究区局部实测图与预测图

Fig.5 Part of the measured map and prediction map of the study area

a. 研究区局部实测图; b. 研究区局部预测图

(3)与使用单一模态信息的 DNN 模型或者 CNN 模型相比,本文提出的基于多模态特征融合的地质体识别模型在准确率和 F1 值上提升了 14.08% 和 14.42% 以及 2.79% 和 2.69%,证明了所提出模型的优越性能和分类预测的准确性。

目前,该方法正在云南、贵州沉积岩区、安徽变质区、青海造山带开展进一步示范应用实验,根据实验结果进一步改进与完善该方法。随着方法的优化,将改善现有地质调查工作模式,形成新型地质填图模式。

References

- Anderson, T. I., Vega, B., Kovscek, A. R., 2020. Multimodal Imaging and Machine Learning to Enhance Microscope Images of Shale. *Computers & Geosciences*, 145: 104593–104606. <https://doi.org/10.1016/j.cageo.2020.104593>
- Arabi Aliabad, F., Shojaei, S., Zare, M., et al., 2019. Assessment of the Fuzzy ARTMAP Neural Network Method Performance in Geological Mapping Using Satellite Images and Boolean Logic. *International Journal of Environmental Science and Technology*, 16(7): 3829–3838. <https://doi.org/10.1007/s13762-018-1795-7>
- Bhandari, D., Paul, S., Narayan, A., 2019. Multimodal Data Fusion and Prediction of Emotional Dimensions Using Deep Neural Network. *Computational Intelligence: Theories, Applications and Future Directions*, 799(2): 215–228. https://doi.org/10.1007/978-981-13-1135-2_17
- Brandmeier, M., Chen, Y., 2019. Lithological Classification Using Multi-Sensor Data and Convolutional Neural Networks. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2/W16: 55–59. <https://doi.org/10.5194/isprs-archives-xlii-2-w16-55-2019>
- Carter, A., Clift, P. D., 2008. Was the Indosinian Orogeny a Triassic Mountain Building or a Thermotectonic Reactivation Event? *Comptes Rendus Geoscience*, 340(2–3): 83–93. <https://doi.org/10.1016/j.crte.2007.08.011>
- Cheng, G. J., Guo, W. H., 2017. Rock Images Classification by Using Deep Convolution Neural Network. *Journal of Physics: Conference Series*, 887(1): 012089. <https://doi.org/10.1088/1742-6596/887/1/012089>
- Gao, S., 2020. A Review of Recent Researches and Reflections on Geospatial Artificial Intelligence. *Geomatics and Information Science of Wuhan University*, 45(12): 1865–1874(in Chinese with English abstract).
- He, J., Zhang, C.Q., Li, X.Z., et al., 2020. Survey of Research on Multimodal Fusion Technology for Deep Learning. *Computer Engineering*, 46(5): 1–11(in Chinese with English abstract).
- Krizhevsky, A., Hinton, G., 2009. Learning Multiple Layers of Features from Tiny Images. *Handbook of Systemic Autoimmune Diseases*, 1(4).
- Krizhevsky, A., Sutskever, I., Hinton, G. E., 2012. ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems*, 25(2): 1097–1105. <https://doi.org/10.1145/3065386>
- LeCun, Y., Bottou, L., Bengio, Y., et al., 1998. Gradient-Based Learning Applied to Document Recognition. *Proceedings of the IEEE*, 86(11): 2278–2324. <https://doi.org/10.1109/5.726791>
- Li, C. W., Zhang, R. S., Zhang, Z. T., et al., 2018. Tectonic Interpretation and Analysis Based on Multisource Remote Sensing Data: A Case Study of Jitai River in Chayi, Tibet. *Remote Sensing Technology and Applica-*

- tion, 33(4): 657–665(in Chinese with English abstract).
- Liu, J.W., Ding, X.H., Luo, X.L., 2020. Survey of Multimodal Deep Learning. *Application Research of Computers*, 37(6): 1601–1614(in Chinese with English abstract).
- Maimaitijiang, M., Sagan, V., Sidike, P., et al., 2019. Soybean Yield Prediction from UAV Using Multimodal Data Fusion and Deep Learning. *Remote Sensing of Environment*, 237: 111599. <https://doi.org/10.1016/j.rse.2019.111599>
- Pires de Lima, R., Duarte, D., Nicholson, C., et al., 2020. Petrographic Microfacies Classification with Deep Convolutional Neural Networks. *Computers & Geosciences*, 142: 104481. <https://doi.org/10.1016/j.cageo.2020.104481>
- Pour, A. B., Hashim, M., Makoundi, C., et al., 2016. Structural Mapping of the Bentong–Raub Suture Zone Using PALSAR Remote Sensing Data, Peninsular Malaysia: Implications for Sediment–Hosted/Orogenic Gold Mineral Systems Exploration. *Resource Geology*, 66(4): 368–385. <https://doi.org/10.1111/rge.12105>
- Ran, X. J., Xue, L. F., Zhang, Y. Y., et al., 2019. Rock Classification from Field Image Patches Analyzed Using a Deep Convolutional Neural Network. *Mathematics*, 7(8): 755–770. <https://doi.org/10.3390/math7080755>
- Russakovsky, O., Deng, J., Su, H., et al., 2015. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, 115(3): 211–252. <https://doi.org/10.1007/s11263-015-0816-y>
- Simonyan, K., Zisserman, A., 2015. Very Deep Convolutional Networks for Large-Scale Image Recognition".: arXiv: 1409.1556. <https://arxiv.org/abs/1409.1556>
- Singer, D. A., 2021. How Deep Learning Networks could be Designed to Locate Mineral Deposits. *Journal of Earth Science*, 32(2): 288–292. <https://doi.org/10.1007/s12583-020-1399-2>
- Su, Y. H., Fan, K., Bach, N., et al., 2019. Unsupervised Multi-Modal Neural Machine Translation. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 15–20, 2019, Long Beach, CA, USA. IEEE, 10474–10483. <https://doi.org/10.1109/CVPR.2019.01073>
- Tu, K., Wen, Q., Chen, H., et al., 2019. New Method of Structural Interpretation in Meadow Covering Based on GaoFen-3 Pol-SAR Images. *Journal of Remote Sensing*, 23(2): 243–251(in Chinese with English abstract).
- Wang, X., Huang, Q. Y., Celikyilmaz, A., et al., 2019. Reinforced Cross-Modal Matching and Self-Supervised Imitation Learning for Vision–Language Navigation. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 15–20, 2019, Long Beach, CA, USA. IEEE, 6622–6631. <https://doi.org/10.1109/CVPR.2019.00679>
- Xu, Y. Y., Li, Z. X., Xie, Z., et al., 2020. Prediction of Copper Mineralization Based on Semi-Supervised Neural Network. *Earth Science*, (12): 4563–4573(in Chinese with English abstract).
- Zhang, Y., Song, B., Du, X. J., et al., 2018. Vehicle Tracking Using Surveillance with Multimodal Data Fusion. *IEEE Transactions on Intelligent Transportation Systems*, 19(7): 2353–2361. <https://doi.org/10.1109/TITS.2017.2787101>
- Zheng, M. Y., Tie, Y., Qi, L., et al., 2019. Dynamic Gesture Recognition Based on the Multimodality Fusion Temporal Segment Networks. 2019 8th International Symposium on Next Generation Electronics (ISNE). October 9–10, 2019, Zhengzhou, China. IEEE, 1–3. <https://doi.org/10.1109/ISNE.2019.8896438>
- Zhou, T. X., Su, R., Guo, Y., et al., 2020. A Multimodality Fusion Network Based on Attention Mechanism for Brain Tumor Segmentation. 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI). April 3–7, 2020, Iowa City, IA, USA. IEEE, 377–380. <https://doi.org/10.1109/ISBI45749.2020.9098392>
- Zuo, R.G., 2021. Data Science–Based Theory and Method of Quantitative Prediction of Mineral Resources. *Earth Science Frontiers*, 28(3): 49–55(in Chinese with English abstract).
- Zuo, R. G., Peng, Y., Li, T., et al., 2021. Challenges of Geological Prospecting Big Data Mining and Integration Using Deep Learning Algorithms. *Earth Science*, 46(1): 350–358(in Chinese with English abstract).

附中文参考文献

- 高松, 2020. 地理空间人工智能的近期研究总结与思考. 武汉大学学报(信息科学版), 45(12): 1865–1874.
- 何俊, 张彩庆, 李小珍, 等, 2020. 面向深度学习的多模态融合技术研究综述. 计算机工程, 46(5): 1–11.
- 李晨伟, 张瑞丝, 张竹桐, 等, 2018. 基于多源遥感数据的构造解译与分析:以西藏察隅吉太曲流域为例. 遥感技术与应用, 33(4): 657–665.
- 刘建伟, 丁熙浩, 罗雄麟, 2020. 多模态深度学习综述. 计算机应用研究, 37(6):1601–1614.
- 涂宽, 文强, 谌华, 等, 2019. GF-3全极化影像在地表浅覆盖区进行地质构造解译的新方法. 遥感学报, 23(2): 243–251.
- 徐永洋, 李孜轩, 谢忠, 等, 2020. 基于半监督神经网络的铜矿预测方法. 地球科学, 45(12): 4563–4573.
- 左仁广, 2021. 基于数据科学的矿产资源定量预测的理论与方法探索. 地学前缘, 28(3): 49–55.
- 左仁广, 彭勇, 李童, 等, 2021. 基于深度学习的地质找矿大数据挖掘与集成的挑战. 地球科学, 46(1): 350–358.