

基于多源数据融合的地质勘探信息智能分析

文 | 张宗军 蒋佳俊

战略性矿产资源勘查正面临环境恶劣、研究程度低等诸多困境，传统地质填图方法受地形条件和人工经验方面制约，难以满足高精度识别的实际需求。地球物理、地球化学和遥感数据分别刻画岩体物性、元素分布和光谱特征，不过单一数据源无法实现对矿产资源的精确识别。构建多源数据融合体系，结合深度学习算法，成为突破传统勘查技术瓶颈的关键路径。

多源地质数据融合分析理论

地质勘探数据在观测原理、物理单位和空间尺度方面存在明显不同。地球物理数据主要测量岩体密度、磁化率、电阻率等物理属性，其空间分辨率由测网间距决定且通常在百米到千米范围。地球化学数据通过分析水系沉积物样本中的元素浓度获取，区域调查采样间隔一般为1~4km且空间分辨率受限于采样点分布，遥感数据捕捉地表矿物的光谱反射特征，多光谱传感器能提供空间分辨率为10~90m的多波段影像。三类数据的物理单位包含重力异常(mGal)、元素含量百分比、无量纲的光谱反射率等多种计量体系，数据形态表现为离散测点、栅格图层和多维影像等不同格式，这种多维度差异要求，需构建统一的融合分析平台，以实现各类信息协同处理。

单一数据源在岩性识别中受信息维度不足与物性参数多解性限制，识别精度存在固有上限。地球物理数据虽能探测隐伏岩体，但不同岩性可能表现出相似的物性特征，从而导致识别混淆；地球化学数据受粗网格采样限制，难以精细刻画岩体边界细节，水系沉积物反映流域混合信息而非原地成分，元素异常区域与岩体实际位置存在空间偏移；遥感数据易受植被覆盖、地形阴影影响，混合像元造成光谱失真，矿物成分相似的岩性在光谱特征上高度一致，单一遥感影像难以满足高精度填图需求。针对上述局限性，构建数据层-特征层-决策层的三级协同处理框架，数据层会通过标准化处理以及空间对齐操作，把不同来源的数据转换成统一格式和量纲，进而形成融合物理、化学及光谱属性的综合特征集合，特征层要借助深度学习模型的多层非线性变换手段，将原始数据转化成具有深层语义的抽象表达形式，决策层采用端到端的学习方法来完成像元级别的分类任务，能够有效整合地球物理深部探测、地球化学指标分析和遥感空间解译能力，以此增强地质目标的区分度，成功突破单一数据源在识别精度上的瓶颈限制。

深度学习驱动的融合技术实现

异构数据标准化与空间配准

异构地质数据融合的首要任务是消除量纲差异并实现空间对齐，以此建立起统一的数据表征体系。地球物理异常强度、地球化学元素含量和遥感波段反射率分属不同物理量纲，采用Z-score标准化方法来实现归一化处理：

$$z = \frac{x - \mu}{\sigma}$$

其中x为原始观测值，μ为样本均值，σ为标准差，z为标准化后数值。该方法把航磁异常值、主要化学组分含量以及多光谱波段反射率等不同类型数据集转换到标准正态分布空间，也就是均值为0、标准差为1的空间，消除不同量纲对后续模型训练造成的干扰。通过双线性插值技术达成多源数据的空间配准，将各类数据统一重采样至目标网格分辨率，并且依靠地理坐标系构建像元级空间关联，配准过程要严格把控误差，确保空间位置精确对齐，最终形成空间对齐的多维特征矩阵，作为深度学习模型的标准化输入数据。

卷积神经网络特征提取

卷积神经网络依靠多层非线性转换来自提取原始数据至抽象特征，摆脱传统方法依赖人工设计特征的束缚。深层卷积网络利用小型卷积核逐层提取特征，初期网络捕捉边缘与纹理等基础要素，中期网络识别光谱组合及空间分布规律，后期网络融合多源数据的高级语义特征，池化层通过下采样逐步扩大感受野范围，实现从局部细节到全局结构的多尺度特征表达。卷积操作凭借权重共享机制降低参数量，非线性激活函数引入复杂映射能力，让网络深入挖掘地球物理低频异常、地球化学元素梯度、遥感光谱吸收峰等多源数据间的非线性关联关系，自主学习区分不同岩性的特征组合。

端到端融合识别模型构建

全卷积神经网络架构摒弃传统全连接层设计，依靠反卷积上的采样技术，达成从输入到输出的端到端像素级岩性分类。在解码环节，通过多层反卷积操作，逐步把特征图的分辨率恢复到原始输入尺寸，并且利用跳跃连接结构整合编码阶段提取的浅层细节特征和解码阶段形成的深层语义特征，有效保留空间边界信息，避免分类结果过度平滑。

Softmax 分类器为每个像素位置输出各类岩性的概率分布，模型优化目标为最小化加权交叉熵损失函数：

$$L = - \sum_{i=1}^N w_i \cdot y_i \cdot \log(\hat{y}_i)$$

其中 N 为岩性类别数， w_i 为类别权重系数（针对样本不均衡问题设置）， y_i 为真实标签， \hat{y}_i 为预测概率。优化器运用自适应学习率策略来迭代更新网络参数，训练过程借助数据增强技术扩充样本的多样性，以此提升模型对旋转缩放等几何变换的鲁棒性与泛化能力。

岩性智能识别实验验证

多源数据集成处理

多源数据融合处理时，要把重力、航磁、地球化学、多光谱遥感影像等异构数据源整合成统一的训练样本集。地球物理数据经过校正处理之后提取其中的异常信息，地球化学数据借助空间插值的方式生成元素分布图层，多光谱影像经过辐射定标和大气校正，从而得到地表反射率，利用影像融合技术，结合不同传感器的光谱与空间优势。从地质图中提取各类岩性单元的空间分布范围并转换为栅格标签数据，合理划分训练集、验证集和测试集来构建含多维特征向量的融合数据集，为监督学习模型训练提供标注样本支撑。

融合模型训练与参数优化

模型训练依靠监督学习的方法，利用标注数据来驱动网络参数优化，从而逐步增强岩性识别的效果。针对样本分布不均问题，通过为不同岩性分配权重系数，以此均衡各类别在损失计算里的影响，在卷积层之后添加批量归一化层，能够加快收敛速度并缓解梯度消失问题，而 Dropout 层是通过随机抑制神经元来防止过拟合现象，采用动态衰减的学习率策略，在训练过程当中逐步调整学习率，进而推动模型逼近最优解。早停机制是通过监测验证集的表现，在连续多个周期无提升时终止训练，并且保留最佳模型权重，借助 GPU 的并行计算能力，可以高效处理大规模的地质数据。基于 VGG-19 等经典架构构建全卷积网络，能够深度解析多源数据的复杂非线性关联，最终实现端到端的像素级岩性分类。

识别精度定量评估

对于量化模型的性能进行准确衡量，评估体系涵盖了混淆矩阵、准确率、召回率以及 F1 分数等多种指标。以高原造山带特定类型过铝质岩体识别为例，深度学习融合模型针对淡色花岗岩的识别准确率达到了 96.0%，其召回率达到了 94.3%，F1 分数为 95.1%，和随机森林等传统方法（准确率 87.5%）相比提高了 8.5 个百分点，其他岩性单元的具体表现情况为，花岗质片麻岩准确率为 91.2%，黑云母石英片

岩准确率为 88.7%，大理岩准确率达到 93.5%，片麻岩准确率为 89.8%，第四系沉积物准确率高达 97.2%，总体分类精度为 92.4%，Kappa 系数达到了 0.908（见表 1），误判样本大多存在于岩体边缘过渡带这个区域，该区域因为不同岩性矿物组成相似度比较高，从而导致光谱与地球化学特征出现混淆的现象。深度学习模型所生成的岩性分布图边界连续性指标从随机森林的 0.63 提升到了 0.91，空间聚合度提高了 43%，有效地克服了浅层机器学习造成的椒盐噪声问题。识别结果表明，多数岩性单元和地质图的吻合度相对较高，高潜力区主要集中在造山带沉积岩系与深变质基底的结构接触带，这充分证实了多源数据融合与深度学习技术在复杂地质环境中岩性识别的实用性。

表 1 不同岩性单元识别精度对比

岩性单元	准确率 (%)	召回率 (%)	F1 分数	样本数量
目标岩体 A	96.0	94.3	0.951	312
花岗质片麻岩	91.2	89.5	0.903	256
变质片岩	88.7	87.2	0.880	189
大理岩	93.5	92.1	0.928	145
片麻岩	89.8	88.6	0.892	198
第四系沉积物	97.2	96.8	0.970	223
总体	92.4	91.5	0.919	1323

来源：济南市勘察测绘研究院

结束语

综合运用地球物理、地球化学、遥感数据的独特优势，多源信息融合技术弥补了单一数据源在复杂地质体识别上的不足。数据层融合会对异构数据开展标准化处理以及空间配准工作，特征层融合依靠深度学习网络自动提取多尺度判别特征，决策层融合负责构建端到端像素级分类模型。实验结果表明，全卷积神经网络融合模型的识别精度显著高于传统方法，有效化解了高原山区等恶劣环境下地质填图方面的难题，为战略性矿产勘查提供了高效有力的技术支持，推动地质勘探信息的智能分析朝着更高精度方向不断发展。

作者简介：张宗军 济南市勘察测绘研究院
蒋佳俊 济南市勘察测绘研究院

责任编辑：孙姗姗 投稿邮箱：zhouhl@staff.ccidnet.com