



地球科学

Earth Science

ISSN 1000-2383, CN 42-1874/P

《地球科学》网络首发论文

题目：矿产预测大模型
作者：师路易，左仁广
收稿日期：2025-09-15
网络首发日期：2025-09-17
引用格式：师路易，左仁广. 矿产预测大模型[J/OL]. 地球科学.
<https://link.cnki.net/urlid/42.1874.P.20250916.1824.002>



网络首发：在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

出版确认：纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

doi:10.3799/dqkx.2025.190

矿产预测大模型

师路易¹, 左仁广^{*1}

1. 中国地质大学（武汉）地质过程与成矿预测全国重点实验室，湖北 武汉 430074

摘要：大数据人工智能驱动的矿产预测已成为矿产勘查的重要方向，但现有方法普遍存在泛化能力弱、迁移性差、可解释性不足等问题，难以实现跨区域应用。以“预训练-微调”为核心范式的大模型在自然语言处理、计算机视觉等领域展现出卓越的跨任务迁移和强泛化性，为突破上述瓶颈提供了可行路径。研发矿产预测大模型，对创新找矿范式、提升勘查效率具有重要探索价值，是智能矿产预测发展的新方向。本文系统梳理了大模型的发展历程与构建流程，聚焦大语言模型、视觉大模型及多模态大模型的技术特性。在此基础上，结合现有矿产预测大模型，剖析了技术路径与局限，探讨了面向矿产预测语言、视觉和多模态大模型的构建思路，分析了构建矿产预测大模型面临的挑战，为进一步研发矿产预测大模型提供参考。

关键词：矿产预测；矿产资源潜力评价；大模型；矿产预测大模型

中图分类号：P628

收稿日期：2025-09-15

Foundation model for mineral prospectivity mapping

Shi Luyi¹, Zuo Renguang^{*1}

1. State Key Laboratory of Geological Processes and Mineral Resources, China University of Geosciences (Wuhan), Wuhan 430074, China

Abstract: Mineral prospectivity mapping (MPM) driven by big data and artificial intelligence (AI) represents a cutting-edge approach in mineral exploration. However, traditional methods typically suffer from limited generalization ability, poor transferability, and insufficient interpretability, resulting in the inability to achieve stable cross-regional applications. Foundation models, based on the “pretrain and fine-tune” paradigm, have demonstrated excellent cross-task transfer and strong generalization ability in the fields such as natural language processing and computer vision, offering a promising path to overcome the aforementioned bottlenecks. The development of foundation models for MPM holds significant potential to revolutionize traditional models and improve exploration efficiency, representing a new research direction for intelligent MPM. This study systematically reviews the state-of-the-art of the development and construction processes of foundation models, focusing on the technical characteristics of large language models, visual foundation models, and multimodal foundation models. This study also summarized the limitations of existing foundation models for MPM, and explored the construction process of MPM foundation models from a perspective of language-based, visual-based and multimodal-based foundation models, and discussed the challenges of developing MPM foundation models, providing a reference for development of MPM foundation models.

Keywords: Mineral exploration; Mineral prospectivity mapping; Foundation model; Foundation model for mineral prospectivity mapping

基金项目：国家自然科学基金（42530801, 42425208）

作者简介：师路易（2002-），男，硕士研究生，主要从事数学地质与矿产勘查方面的研究。E-mail: 1821408640@qq.com

通讯作者简介：左仁广（1981-），男，博士，教授，博士生导师，主要从事数学地质与矿产勘查方面的研究。E-mail: zrguang@cug.edu.cn

0 引言

矿产预测的核心目标是评估研究区矿床的空间分布、类型、规模与质量，为矿产勘查决策提供科学依据 (Taylor and Steven, 1983; Agterberg, 1989; Carranza, 2009)。找矿数据具有多源异构、多维、非线性、多尺度，以及稀缺等特点 (Zuo, 2016; 2020)，人工智能算法凭借其强大的非线性特征提取与模式识别能力，已被广泛应用于识别和提取这些数据中的矿化空间模式及非线性关联 (Zuo and Carranza, 2011; Zuo et al., 2023)。根据找矿数据输入的形式，这些方法可分为基于像素(如随机森林) (Zhang et al., 2016; Xiong et al., 2018; Wang and Zuo, 2022; Yin et al., 2022)、基于图像(卷积神经网络, Convolutional Neural Network, CNN) (Li et al., 2021; Yang et al., 2023; Yang and Zuo, 2024) 及基于拓扑图(如图神经网络)(Xu and Zuo, 2023; Zuo and Xu 2023; Shi et al., 2024) 的三大类。然而，仅依赖数据驱动模型普遍存在泛化能力弱、可移植性差、可解释性不足等问题。学者们通过耦合地质知识辅助模型学习成矿规律，在一定程度上缓解了这些难题 (Zuo et al., 2023; Xu and Zuo, 2023; Yang and Zuo, 2024; 左仁广等, 2024; 左仁广, 2025)。但这些模型通常只能在同一研究区完成训练与预测，难以迁移至未经训练的区域 (Shi and Zuo, 2025a)，这一不足制约了人工智能算法在矿产预测领域的推广应用。

基础模型能够通过在大规模数据集上进行预训练，学习通用表示，并在下游任务上使用少量数据进行微调来适配该任务 (Bommasani et al., 2022; Awais et al., 2025)。当基础模型的参数规模达到一定程度，也被称为大模型 (本文统称为大模型)。大模型根据其输入数据可分为大语言模型 (Radford and Narasimhan, 2018; Devlin et al., 2019; Lewis et al., 2019; Brown et al., 2020; Bommasani et al., 2022; Touvron et al., 2023)、视觉大模型 (Dosovitskiy et al., 2020; He et al., 2021; Z. Liu et al., 2021; Bao et al., 2022) 和多模态大模型 (Radford et al., 2021; Rombach et al., 2021; Kirillov et al., 2023)，被应用于对话系统 (Achiam et al., 2023; Bai et al., 2023; A. Liu et al., 2024; Gong et al., 2025)、图片生成 (Rombach et al., 2021; Gong et al., 2025)、视频生成 (Y. Liu et al., 2024; Gong et al., 2025) 等多样化任务，其强大的泛化能力与下游任务适配性使其成为各领域的最新技术范式。在地球科学领域，也涌现出了多个适配地学场景的大模型，如地球物理大模型 (Liu and Ma, 2024; Sheng et al., 2025a, 2025b; Guo et al., 2025)、遥感大模型 (Hu et al., 2023; F. Liu et al., 2024; Guo et al., 2024; Hong et al., 2024)、地理信息系统大模型 (Zhang et al., 2023; Zhang et al., 2024) 及地质大语言模型 (雷丹泓和刘建忠, 2025; Lin et al., 2023; Deng et al., 2024; 之江实验室, 2025) 等。

在矿产预测领域，现有人工智能算法泛化能力的不足已成为其发展的关键瓶颈 (左仁广, 2025)，而大模型的出现为突破这一瓶颈带来了转机。找矿大数据包含空间数据与文本数据，大语言模型具备强大的语义特征提取与推理能力，视觉大模型则擅长空间信息提取，二者的结合使矿产预测大模型的构建成为可能。当前，已有三项代表性研究探索了这一方向：其一为矿产智能体 (MineAgent)(Yu et al., 2024)，该模型以通用大模型为骨干，通过 Agent 系统引导大模型完成控矿构造的自动提取；其二为矿产预测空间大模型 (Geospatial Foundation Models for Mineral Prospectivity Mapping, GFM4MPM) (Daruna et al., 2024)，作为典型的视觉大模型，它基于北美及澳大利亚的地质和地球物理数据从零构建，并利用铅锌矿床数据进行有监督微调以圈定找矿靶区；其三为地球化学异常识别大模型 (Foundation Model for Geochemical Anomaly Identification, FM4GAI)(Shi and Zuo, 2025b)，同样是视觉大模型，它基于区域地球化学数据进行从零构建，

分别利用迁移学习和小样本学习的方式进行有监督微调,在不同找矿场景下圈定矿致地球化学异常图。然而, MineAgent 未采用专业找矿数据进行微调,导致其预测精度有限; GFM4MPM 采用 One - Hot 编码处理文本数据造成了语义信息丢失,且存在输入特征固化难以适配其他研究区等问题; FM4GAI 通过提取单一地球化学特征图的空间特征,使其能够被应用到不同研究区。

本文总结了语言大模型、视觉大模型和多模态大模型的研究现状,梳理了大模型构建的技术路线,探讨了矿产预测大模型的架构设计,讨论了矿产预测大语言模型、视觉大模型及多模态大模型构建过程中面临的挑战,旨在为开发兼具跨区域泛化能力与专业地质适配性的矿产预测大模型提供参考。

1 大模型概述

大模型是指具有大量参数的深度神经网络模型 (Bommasani et al., 2022; Awais et al., 2025),其思想源于机器学习中的迁移学习,将从源域中学习到的知识迁移到目标域中进行应用,旨在减少模型对目标域标注数据的依赖 (Krizhevsky et al., 2012)。该思想早在深度信念网络中得到了应用 (Hinton et al., 2006)。深度信念网络由多层受限玻尔兹曼机 (Restricted Boltzmann Machine, RBM) 组成,通过逐层预训练 RBM,学习数据的层次化特征,使用有标签数据对整个网络进行优化,从而适配具体任务(Hinton and Salakhutdinov, 2006)。大模型通常需要从大规模数据集上进行预训练来获得较强的泛化能力,然后在下游任务数据集上微调后即具有出色的性能 (Devlin et al., 2019; He et al., 2021)。大模型根据其输入数据形式,通常分为大语言模型,视觉大模型和多模态大模型 (Bommasani et al., 2022)。

1.1 大语言模型

自然语言处理领域的核心问题是如何将自然语言表征为计算机能理解的数值特征,从而能够利用该特征完成各种自然语言任务,如对话系统、翻译任务、命名实体识别等 (Nadkarni et al., 2011; Chowdhary, 2020)。传统的 One - Hot 编码表征方式将每个输入词定义为相互正交的高维稀疏向量,导致词与词之间彼此独立,无法高效地表达词语间的语义关系。可通过预训练将词语映射到蕴含语义和上下文关系的向量空间,如 Word2Vec (Mikolov et al., 2013)、GloVe (Pennington et al., 2014) 等。该领域早期使用的模型通常为浅层的机器学习或循环神经网络及其变体 (Collobert et al., 2011; Mikolov, et al., 2013; Sutskever et al., 2014; Tang et al., 2015),这些模型往往计算效率低、全局语义捕捉能力不足、可拓展性差(Vaswani et al., 2017)。Transformer 模型架构 (图 1) 显著提升了长距离依赖的建模能力,使模型可以理解更复杂的语言结构和语义关联 (Vaswani et al., 2017)。Transformer 具有较强的可拓展性、通用性和灵活性,使得大规模的预训练模型成为可能,推动了大模型迈入了“预训练 - 微调”的范式 (Zhao et al., 2023)(图 2)。

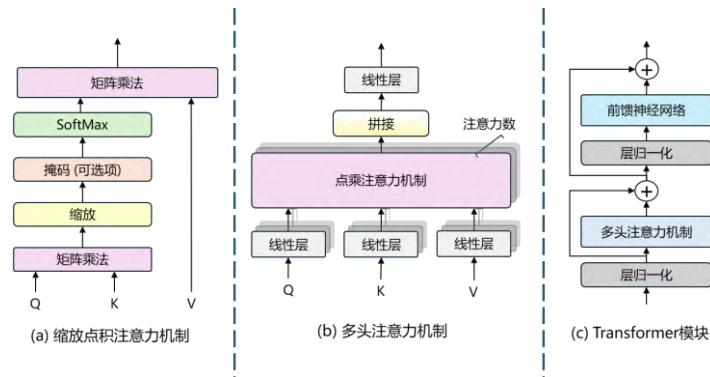


图 1. Transformer 架构示意图 (修改自 Vaswani et al., 2017)

Figure 1. An architecture of the Transformer architecture (Vaswani et al., 2017).

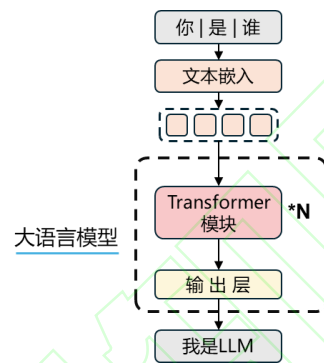


图 2. 大语言模型架构示意图

Figure 2. An architecture of the large language model

预训练任务决定了模型从自然语言中学习文本语义信息的方式，如 BERT (Devlin et al., 2019) 和 GPT (Radford and Narasimhan, 2018; Brown et al., 2020)。BERT 采用掩码语言建模 (Masked Language Model, MLM)(图 3a) 和下一个句子预测 (Next Sentence Prediction, NSP)(图 3b)学习文本中的语义信息，通过添加任务特定的输出层，可将预训练的模型参数 (即语义理解能力) 迁移到下游任务，如文本分类、问答系统、命名实体识别、自然语言推理等任务。GPT 使用自回归语言建模 (Causal Language Modeling, CLM) 学习语言生成规律 (图 3c)，采用提示词工程来引导模型生成目标输出，或通过基于人类反馈的强化学习来提升模型输出结果的质量 (Ouyang et al., 2022)，最终应用于文本生成、对话系统、代码生成等下游任务 (Achiam et al., 2023)。BERT 和 GPT 分别代表了重构式和生成式大语言模型，从理解和生成两个方向创新了自然语言处理的方法，成为大语言模型构建的基础 (Zhou et al., 2023)。在应用时，这些模型通过挂载知识库和检索增强来缓解生成式大语言模型的幻觉现象 (Lewis et al., 2019; Liang et al., 2025)。同时，采用思维链、自我验证和迭代推理技术来增强大语言模型的推理能力，优化其在需要复杂逻辑的任务中的应用 (Wei et al., 2022)。大语言模型在性能提升的同时，其参数量也越来越庞大，需要消耗的计算资源也越来越多。可以通过参数共享、知识蒸馏、混合精度训练和混合专家模型 (Mixture of Expert) 来提升模型的计算效率，降低其运营成本 (Jacobs et al., 1991; Fedus et al., 2022)。如今，大语言模型正从单一模态迈向多模态，从特定任务的大语言模型迈向更全能、更高效的通用大语言模型 (Yin et al., 2024)。

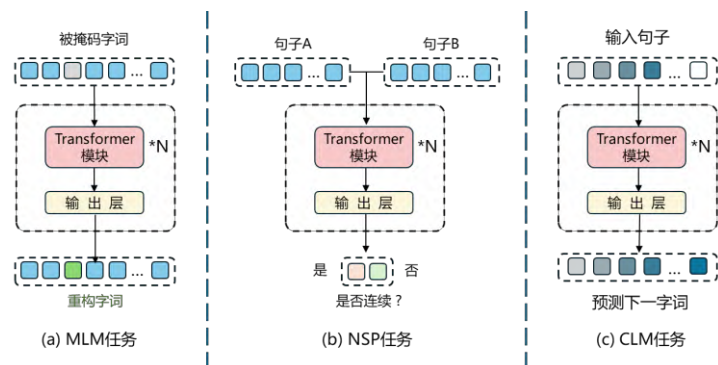


图 3. 大语言模型预训练任务示意图

Figure 3. An architecture of pre-training tasks for large language models

1.2 视觉大模型

计算机视觉领域的核心问题在于如何从图像或视频中提取并理解视觉信息，使计算机能够感知和解析视觉内容，并解决图像识别、目标检测、三维重建等视觉任务 (Voulodimos et al., 2018; Szeliski, 2022)。早期依赖特征工程的浅层机器学习模型计算复杂且难以应对复杂场景 (Viola and Jones, 2001; Lowe, 2004; Dalal and Triggs, 2005; Lazebnik et al., 2006; Krizhevsky et al., 2012)。CNN (LeCun et al., 1989, 1998; Krizhevsky et al., 2012) 的兴起及 ImageNet 等大规模数据集 (Deng et al., 2009) 的发布共同推动了预训练和微调范式的发展。VGG (Simonyan and Zisserman, 2014)、ResNet (He et al., 2016) 等架构在大规模数据集上进行预训练后，迁移至小规模数据集上取得了良好的效果。然而 CNN 仍需克服全局建模能力弱、灵活性差、计算复杂度高等局限性 (Azulay and Weiss, 2019)。2020 年，Google 团队提出了视觉 Transformer (Visual Transformer, ViT) (Dosovitskiy et al., 2020)，将图片划分为序列数据 (像素块)，借助 Transformer 架构的长距离建模能力，高效完成图像分类任务 (图 4)。

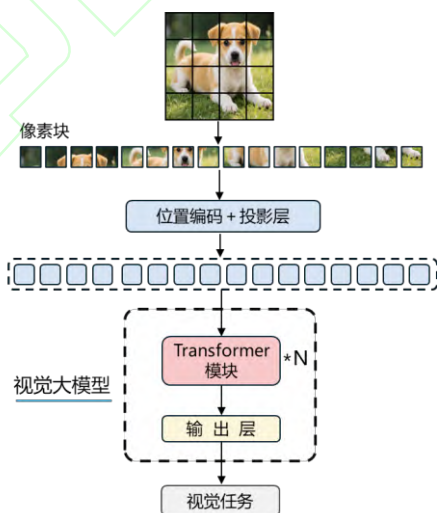


图 4. 视觉大模型示意图 (修改自 Dosovitskiy et al., 2020)

Figure 4. An architecture of the visual foundation model (Dosovitskiy et al., 2020).

ViT 的出现，使该领域模型的架构方向从 CNN 转向了 Transformer (Dosovitskiy et al., 2020; Z. Liu et al.,

2021)。同时，ViT 采用无监督预训练任务如掩码自编码器 (Masked AutoEncoder, MAE)(He et al., 2021)和扩散模型 (Diffusion Models, DM)(Ho et al., 2020; Rombach et al., 2021)(图 5)增强模型的视觉特征提取能力。两种预训练任务分别代表了视觉特征提取和生成两种类型的视觉大模型。在预训练视觉大模型的基础上，添加对应任务的分类器后，进行有监督微调可让模型完成图像分类、目标检测、视频理解、图像生成等视觉任务 (Kirillov et al., 2023; J. Zhang et al., 2024)。

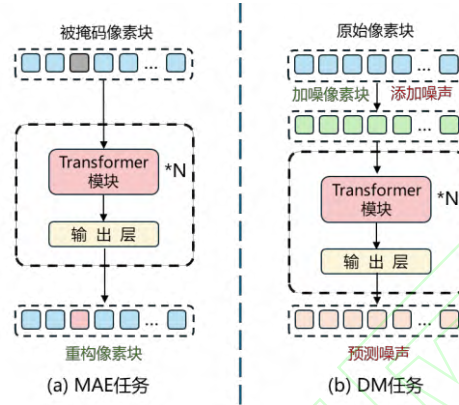


图 5. 视觉大模型预训练任务示意图

Figure 5. An architecture of pre-training tasks for visual foundation models.

1.3 多模态大模型

多模态数据是指由多种异质信息源 (如文本、图像、音频、视频或传感器数据等) 构成的组合表征形式，每种模态的数据在系统中承担特定功能，从不同的角度反映目标任务的特征。多模态融合的核心在于打破单一信息形式的局限，通过整合多模态信息，实现对场景或任务更全面、准确的理解 (Baltrusaitis et al., 2019)。多模态大模型是一种能够处理、理解和生成多种模态信息的人工智能模型 (Bommasani et al., 2022)。在各种模态的信息中，自然语言作为一种复杂度高的数据，其蕴含的内容不仅包括表层的词汇语义和句法结构，更涉及深层的语用意图、文化语境以及动态的语境依赖关系 (Brown et al., 2020; Chowdhary, 2020)。自然图像则不具备自然语言的复杂性，仅使用图像训练的视觉大模型往往缺乏较强的推理能力 (Radford et al., 2021)。因此，可借助大语言模型来提升视觉大模型的推理能力和零样本学习能力，如 Contrastive Language-Image Pre-training (CLIP) (Radford et al., 2021; Achiam et al., 2023; Y. Liu et al., 2024; Gong et al., 2025)。通过对比学习的方式，使图像与文本描述的潜在特征相关联 (图 6)，借助大语言模型的推理能力实现对自然图像的零样本分类。在此基础上，将嵌入的视觉特征与文本特征同时输入到预训练的大语言模型，实现大语言模型对视觉数据的理解。这类以大语言模型为骨干，通过多模态编码器和生成器扩增其能力边界的大模型被称为多模态大语言模型 (Yin et al., 2024)(图 7)。

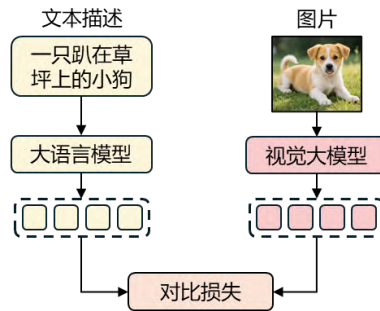


图 6. 多模态大模型预训练任务示意图 (修改自 Radford et al., 2021)

Figure 6. An architecture of pre-training tasks for multimodal foundation models (after Radford et al., 2021).

目前，多模态大语言模型有两种构建方式，其一是统一嵌入解码器方法：借助多模态编码器和投影层，将多模态数据的特征投影至与文本特征相似的特征空间，经过拼接后输入到统一的解码器（大语言模型）中，实现多模态特征融合（Gemma Team, 2024）；其二是跨模态注意力架构方法：该方式将输入多模态编码器和投影层的多模态特征以交叉注意力的方式嵌入到大语言模型的多头注意力机制中，并与文本特征进行交叉注意力计算，从而实现多模态特征融合（Grattafiori et al., 2024）。在上述过程中需要使用 CLIP 预训练任务将多模态特征进行对齐，随后可针对该多模态大语言模型的功能来选择合适的预训练任务，如多模态图像生成（基于 DM 的预训练任务），图文对话（基于 CLM 的预训练任务）等。

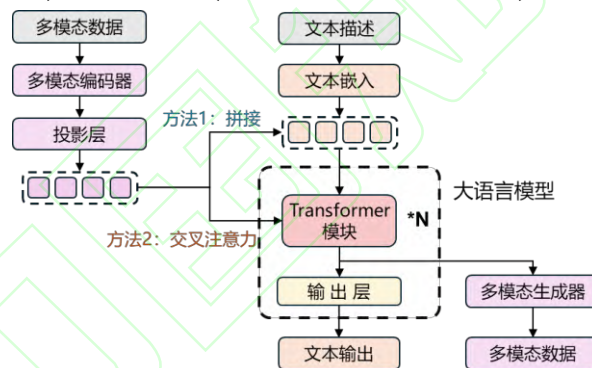


图 7. 多模态大语言模型架构

Figure 7. An architecture of the multimodal large language model

1.4 大模型构建

大模型的构建流程通常包括数据准备、模型预训练、模型微调和模型部署应用等四个核心环节。这些环节相互关联，构成了从原始数据到可部署模型的完整技术链条。

(1) 数据准备。数据集是大模型知识体系的基石，其质量决定了模型的性能。数据集的选择需遵循任务适配性原则：(a) 面向通用知识库构建时，优先选用维基百科、Common Crawl 等语义丰富且经过初步清理的文本数据；(b) 针对代码生成任务，通常采用 GitHub 公开代码库、CodeSearchNet 等专业代码数据库；(c) 针对图文对话任务时，应构建带有图像的文本对话数据集（Schuhmann et al., 2022）。随后对数据进行预处理，如对文本进行数据清理、数据增强、分词和编码；对图像数据进行图像归一化、尺寸标准化和通道转换等（Dosovitskiy et al., 2020）。

(2) 模型预训练。预训练任务的设计直接影响了模型的能力边界，其核心在于通过任务 - 目标对齐原则，将预训练任务与模型的应用目标进行匹配。预训练的目标是使模型从大规模的数据集中学习通用技能，如 MLM 和 MAE 通过掩码重构使模型能够分别学习文本和图像中的语义信息和分布规律，提升模型的理解能力 (Devlin et al., 2019; He et al., 2021); CLM 和 DM 根据不同的生成形式分别对真实文本或图像进行建模，拟合其分布规律，提升模型的生成能力(Radford and Narasimhan, 2018; Rombach et al., 2021); CLIP 利用对比损失，将多模态数据的特征空间进行对齐，提升模型多模态感知能力 (Radford et al., 2021)。

(3) 模型微调。微调阶段旨在使预训练模型适应下游任务。参数高效微调方式主要包括：(a) 基于附加参数的方法：在预训练模型的基础上插入小规模的可训练参数，在微调时冻结预训练网络中插入的可训练参数，如适配器微调(Houlsby et al., 2019)、前缀微调(Li and Liang, 2021; X. Liu et al., 2021)等；(b) 基于参数选择的方法：在微调时选择性地调整预训练模型中的部分参数如偏置项、注意力权重等 (Zaken et al., 2021); (c) 低秩适配方法：引入新的低秩矩阵并与预训练模型的权重矩阵相乘来代替原始的权重矩阵，微调时冻结原始权重矩阵，通过更新低秩矩阵的方式来节约算力，如 LoRA (Hu et al., 2022)、AdaLoRA (Zhang et al., 2023) 等。

(4) 模型部署应用。在实际使用中，大模型的部署需要考虑如何稳定运行和快速响应。通常会把模型封装成标准的服务接口，方便其他系统调用，并通过负载均衡和自动调整计算资源的方式，来应对不同时间的访问压力，保证系统稳定运行 (Dong and Xie, 2024)。为了让模型响应更快、并行处理更多请求，可以采用一些优化技术，比如将多个请求合并成批次一起处理，或把常用的计算结果暂时保存起来，避免重复运算，从而提高处理速度、降低延迟 (Poddar et al., 2022)。同时，还需要建立一套监控系统，持续观察模型的准确度、响应时间以及对服务器资源的占用情况，及时发现问题并进行优化。对于那些数据不断更新的应用场景，还可以让模型在运行过程中逐步学习新数据，不断调整自身，从而保持预测能力的稳定和有效 (Feng et al., 2023)。此外，可通过知识蒸馏 (Hinton et al., 2015) 的方式将大模型知识迁移到小模型，或通过量化 (Jacob et al., 2018)、剪枝 (Han et al., 2015) 等模型压缩技术降低计算复杂度。

2 矿产预测大模型

找矿大数据具有显著的多模态特征，涵盖地质 (地层、构造、岩浆岩、矿床等)、地球化学、地球物理、遥感等空间数据，以及学术文献、地质报告等非结构化文本数据 (左仁广等, 2020; 左仁广, 2021)。尽管大量研究利用空间数据分析技术识别矿化模式并开展矿产预测，但现有方法存在局限性。一方面，模型训练高度依赖研究区的成矿动力学背景和成矿特征，输入数据的固定顺序与局部特征的独特性导致模型泛化能力不足，在迁移到成矿地质背景差异较大的区域时预测性能大幅下降 (Shi and Zuo, 2025a)。另一方面，文本数据 (如文献、报告) 常被作为静态先验知识或结果解释工具 (Yang and Zuo, 2024; Xu et al., 2025a; Zuo et al., 2025)，未能与空间数据实现动态融合，成矿规律与空间模式间的隐含非线性关联难以被充分挖掘。尽管已有研究尝试将文本转化为知识图谱，高效发现成矿规律和学习成矿特征 (Wang et al., 2022; Yan et al., 2023; Qiu et al., 2023; Yan et al., 2024)，但仍受限于跨区域适配性差、多模态数据协同建模不足、依赖专家经验等问题。这些瓶颈促使研究者转向大模型技术的研究 (Y. Zhang et al., 2023)，通过跨区域泛化、多模态深度融

合及知识增强等创新路径，为矿产预测提供更智能、更高效的解决方案。

2.1 矿产预测大模型研究现状

当前矿产预测领域涌现出两类大模型技术路径，分别是使用 Agent 系统调整多模态大模型输出的矿产智能体 (MineAgent)(Yu et al., 2024) 以及使用地质数据和地球物理数据构建的矿产预测空间大模型 (Geospatial Foundation Models for Mineral Prospectivity Mapping, GFM4MPM)(Daruna et al., 2024)。此外在地球化学异常识别领域，已有研究提出了地球化学异常识别大模型 (Foundation Model for Geochemical Anomaly Identification, FM4GAI)(Shi and Zuo, 2025b)。

MineAgent 基于指令微调策略，通过 Agent 系统与现有多模态大模型，如 GPT - 4 (Achiam et al., 2023), Qwen (Bai et al., 2023) 等交互，引导其完成控矿构造提取至成矿概率计算的全流程任务 (图 8)。该方法无需重新训练大模型，仅通过预设指令流程可生成预测结果，降低了计算资源消耗。然而，由于该模型未基于找矿大数据训练，对控矿构造的识别精度有限，且成矿概率输出缺乏地质知识约束，导致预测结果在复杂地质条件下具有较高的不确定性。

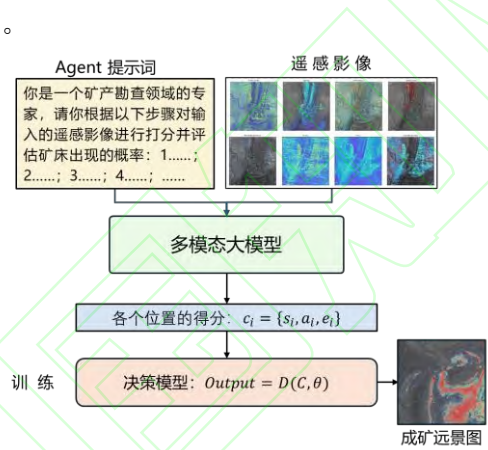


图 8. MineAgent 示意图 (修改自 Yu et al., 2024)

Figure 8. An architecture of MineAgent (Yu et al., 2024)

GFM4MPM (Daruna et al., 2024)以北美及澳大利亚地区的地质、地球物理数据为训练集，从零构建专用于矿产预测的视觉大模型(图 9)。该模型通过 ViT 处理空间数据 (Dosovitskiy et al., 2020)，结合掩码重构任务进行预训练 (He et al., 2021)，并以铅锌矿床空间分布为标签进行有监督微调 (图 9)。尽管其在北美、澳大利亚取得了优于传统方法的预测效果，但依然存在一些局限性。其一是将文本型地质数据转换为 One - Hot 编码，丢失了语义关联性。其二是输入特征固化 (仅依赖特定地质、地球物理数据)，难以迁移到其他成矿地质背景差异较大的研究区。

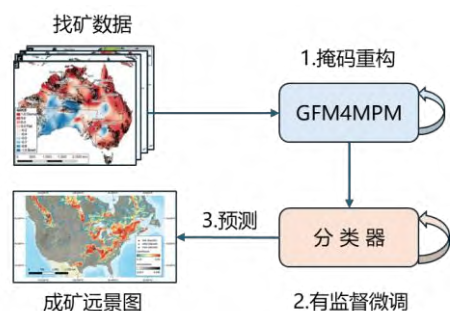


图 9. GFM4MPM 示意图 (修改自 Daruna et al., 2024)

Figure 9. An architecture of GFM4MPM (Daruna et al., 2024).

FM4GAI (Shi and Zuo, 2025b) 同样是从零训练的视觉大模型架构，以 39 种区域地球化学勘查数据为预训练数据集，采用掩码重构作为预训练任务，学习地球化学数据的空间特征，分别使用迁移学习和小样本学习对模型进行有监督微调。同时，将每个地球化学特征拉伸为仅包括单一特征的地球化学图，避免输入特征固化弱化模型的适用性。

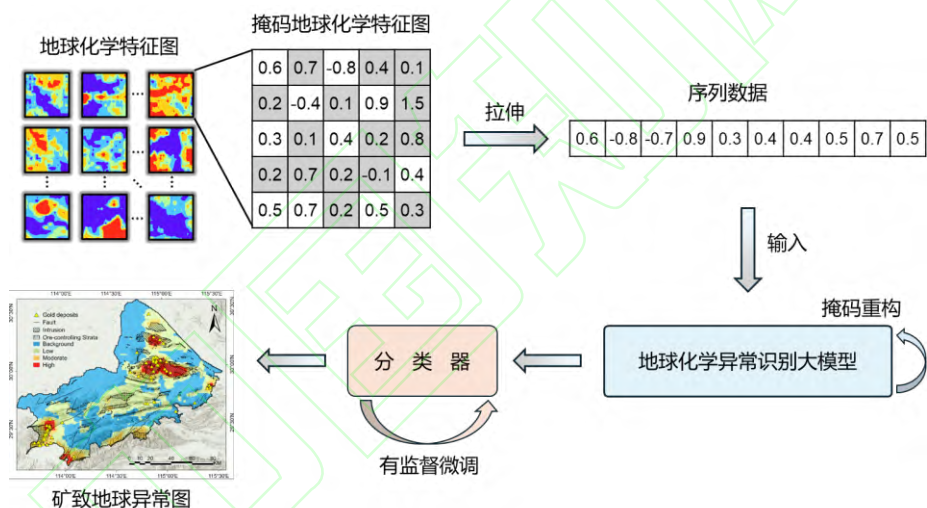


图 10. FM4GAI 示意图 (修改自 Shi and Zuo, 2025b)

Figure 10. An architecture of FM4GAI (Shi and Zuo, 2025b).

如表 1 所示，MineAgent 以低资源消耗为优势，但依赖现有模型的通用能力，缺乏找矿专业性。GFM4MPM 通过专项训练实现高精度预测，却受限于特征固化，缺乏泛化能力和零样本学习能力。两者均在泛化能力与专业适配性之间存在权衡，亟需融合知识驱动与数据驱动的混合架构以突破现有瓶颈。FM4GAI 与 GFM4MPM 类似，通过专项训练实现高精度预测，但空间特征提取的基本单位是单个地球化学特征图，使其能够应用于各种研究区，适用性强，但同样缺乏零样本学习能力。

表 1. MineAgent、GFM4MPM 和 FM4GAI 技术路线对比

Table 1. Comparison of technical roadmaps: MineAgent, GFM4MPM and FM4GAI

路线	代表	训练数据	主要优势	主要局限	适用性
----	----	------	------	------	-----

多模态大模型 + Agent	MineAgent	多波段遥感数据	训练成本低，实现简单，流程自动化	未用专业找矿数据系统微调；对隐伏构造识别有限；概率输出缺少地质约束	适合快速部署与流程自动化
从零构建矿产预测视觉大模型	GFM4MPM	北美/澳大利亚地质与地球物理数据，铅锌矿监督标签	空间模式学习强，目标任务表现好	文本信息 One - Hot 编码致语义丢失；输入特征固化，跨区适配性受限	适合构造数据区域的高精度预测
从零构建地球化学异常识别大模型	FM4GAI	39 种地球化学特征	学习单个地球化学特征图的空间模式	缺少成矿知识约束，结果可解释性差，不确定性高	可用于具有不同地球化学特征的研究区

2.2 矿产预测大模型构建

矿产预测主要涉及文本数据与空间数据。依据数据形式，矿产预测大模型可划分为语言大模型、视觉大模型、多模态大模型三类。

2.2.1 矿产预测语言大模型

找矿文本大数据是矿产预测语言大模型的核心输入，涵盖与矿产勘查相关的文本信息，主要包括学术文献与地质报告。这类数据的核心价值在于其蕴含对现有矿床成因的归纳分析和成矿特征的描述，但其非结构化特性为模型训练带来挑战 (Zhou et al., 2021; Ma, 2022)。矿产预测语言大模型则聚焦于找矿文本大数据处理与成矿知识挖掘的专用大模型，其核心功能是从非结构化文本中提取、整合与成矿相关的规律知识 (Fu et al., 2025)。

矿产预测语言大模型的具体构建过程如下：

(1) 数据准备。收集高质量的找矿文本数据集，如从国内外期刊上收集找矿相关论文、从专业渠道获取找矿相关的地质报告等。若要从零开始构建该语言大模型，则需要收集 1.4 中所提到的通用文本语料，来丰富模型的预训练数据集，使其能覆盖各领域的基本知识，增强模型的泛化能力。可以利用现有大语言模型对数据集进行高效预处理，如数据清洗、数据增强以及构建对话数据集等。为使模型能充分理解找矿文本，应在现有词表的基础上建立专业术语词表。

(2) 模型预训练。预训练时，可首先选择 MLM 和 NSP 对模型进行预训练，增强模型的文本理解能力。在此基础上，选择 CLM 任务来训练模型的文本生成能力 (图 3)。

(3) 模型微调。使用 1.4 中提到的基于人类反馈的强化学习方法来优化预训练模型的输出，使其更合理。此外使用高质量的找矿文本及其构建的对话数据集再次对其参数进行微调，增加其对专业知识的理解。为减少微调时所消耗的计算资源，使用参数高效微调方法对模型参数进行调整。

(4) 模型部署应用。除 1.4 中提到的部署手段以外，可预先使用提示词工程来引导模型对找矿文本的批量处理，如批量文献解析与知识总结、自动构建知识图谱等 (图 11)。同时，为减少模型出现幻觉，编造虚假信息，可选择检索增强生成 (Lewis et al., 2019) 和知识增强生成 (Liang et al., 2025) 等手段挂载地质知识库或知识图谱，使模型能够根据知识库中的内容对用户进行更准确地反馈。

尽管使用专业找矿文本预训练的语言大模型相较于其它通用大模型 (如 GPT、Qwen 等) 对找矿文本数据的适配性较高，但是从零开始训练模型会导致训练成本高昂 (Bommasani et al., 2022)。此外，仅将找矿文本作为训练数据，导致其知识来源过于单一，使其泛化能力不足。因此，在构建垂直语言大模型时，常选择开源的通用大模型，如 Qwen (Bai et al., 2023)、LLaMA (Touvron et al., 2023)、DeepSeek (Liu et al., 2024) 等为基底，对其已有的语言理解和生成能力加以利用 (图 11)，不仅能够减少对找矿文本的数据量的依赖，同时能够减少模型训练成本。

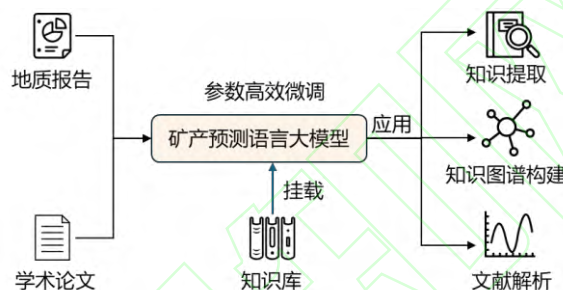


图 11. 矿产预测语言大模型示意图

Figure 11. An architecture of the large language model for mineral prospectivity mapping

2.2.2 矿产预测视觉大模型

找矿空间大数据通常包括地质、地球化学、地球物理及遥感数据，这些数据均蕴含丰富的空间信息 (Zuo, 2020)。综合多源空间数据的矿化信息，可减少单一找矿数据的不确定性，提高找矿成效。但由于不同数据类型的矿化信息特征各异，且研究区数据种类与数量差异显著，导致单一模型难以适配所有场景。因此，矿产预测视觉大模型设计需强调可扩展性。一种可行的方案是采用混合专家系统架构 (Jacobs et al., 1991)，为每类空间数据设置独立子模型，专注提取单一来源的矿化信息 (图 12)。

矿产预测视觉大模型的具体构建过程如下：

(1) 数据准备。找矿空间大数据呈现出多源特性，不同来源的数据具有不同的格式。需要根据各类空间数据的特点对其进行表征和归一化，如地球化学数据可被表征为图像或者拓扑图 (Xu et al., 2025b)。在应对缺失的空间数据时，应进行缺失值处理，然后对空间数据分别进行预处理并嵌入。

(2) 模型预训练。在预训练任务选择上，优先采用能充分提取找矿空间特征的任务。MAE 通过将输入数据进行掩码并预测缺失区域，使模型从已知信息中挖掘缺失区域特征，提升模型对视觉信息的理解能力。针对矿产预测视觉大模型的混合专家系统架构，在第一阶段由各子模型独立使用 MAE 进行预训练，使各子模型能够理解各自领域的的数据。在第二阶段的预训练中，使用同一位置的各类找矿空间大数据同时进行 MAE

预训练，如对某一来源的全部数据进行掩码并使用其他来源的空间数据对其进行重构，使矿产预测视觉大模型能够理解多源找矿空间大数据。

(3) 模型微调。冻结预训练模型参数，添加可训练的任务适配器，使用已知矿床标注的找矿空间数据集进行有监督微调，将高维找矿特征转换为成矿概率，来适配相应的任务。

(4) 模型部署应用。除 1.4 中提到的方式外，还可以为每一种空间数据的输入设置独立的接口以及微调方法，保证其在应用时的便利。

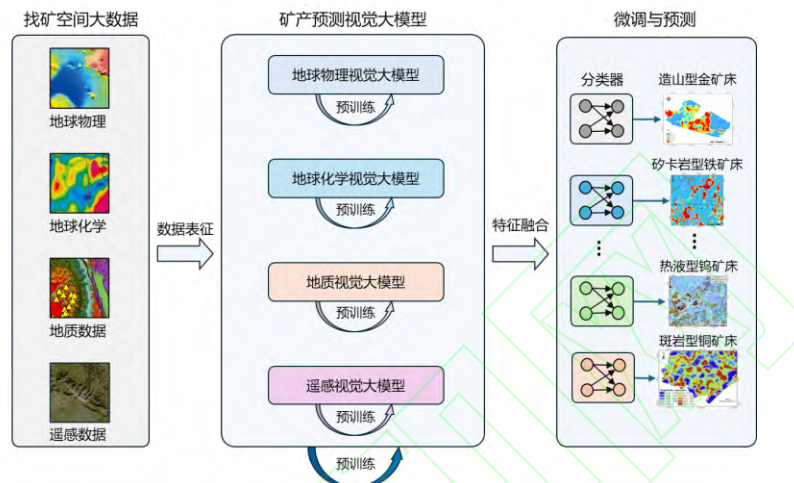


图 12. 矿产预测视觉大模型示意图

Figure 12. An architecture of the visual foundation model for mineral prospectivity mapping

尽管“无监督预训练 + 有监督微调”的策略虽能利用有限的标注数据实现小样本训练，但其仅依赖数据驱动的特性存在一定的局限性。一方面，少量矿床（点）微调往往难以支持无标注区域的矿床发现。另一方面，缺乏成矿知识约束的模型可能产生与地质事实不一致的预测结果，且结果往往难以解释。

2.2.3 矿产预测多模态大模型

多模态大语言模型为矿产预测大模型提供了一个可行路径，矿产预测多模态大模型将以具有推理能力的矿产预测语言大模型为骨干，并借助矿产预测视觉大模型的空间信息感知能力，使其能从未训练区域中自动提取多源找矿信息并圈定成矿远景区和找矿靶区。

矿产预测多模态大模型的具体构建过程如下：

(1) 数据准备。构建如下预训练数据集：(a) 对齐数据集：构建找矿语料 - 空间数据对 (图 13a)，对齐文本数据和空间数据的特征空间；(b) 对话数据集：使用现有大语言模型来构建对话数据集 (图 13b)，增强模型的图文对话能力；(c) 空间数据生成数据集：结合对话数据集，将某一个空间数据作为输出目标来构建该数据集 (图 13c)。同时构建矿产预测微调数据集 (图 14)，可根据现有科技论文文献作为基础，使用现有大语言模型进行优化获得；成矿远景图和找矿靶区图可采用深度学习模型获取，汇总后构建该微调数据集，使模型具备空间决策能力。



图 13. 找矿文本 - 空间预训练数据集

Figure 13. Mineral exploration text-spatial pre-training dataset.



图 14. 找矿文本 - 空间微调数据集

Figure 14. Mineral exploration text-spatial fine-tuning dataset

(2) 模型预训练。该模型可在上文提到的矿产预测语言大模型和矿产预测视觉大模型的基础上进行构建 (图 15)，并额外添加成矿远景图和找矿靶区图生成器来进行空间决策。模型的预训练任务包括三种：(a) 找矿文本-空间对齐任务：使用 CLIP 任务对齐找矿文本和空间数据的特征空间；(b) 找矿文本生成任务：使用 CLM 任务，让模型从输入找矿数据中挖掘并生成对应文本描述；(c) 空间数据生成任务：根据提供的找矿信息，输出对应空间数据，训练模型生成找矿空间数据的能力，使矿产预测大模型具有初步空间决策能力。

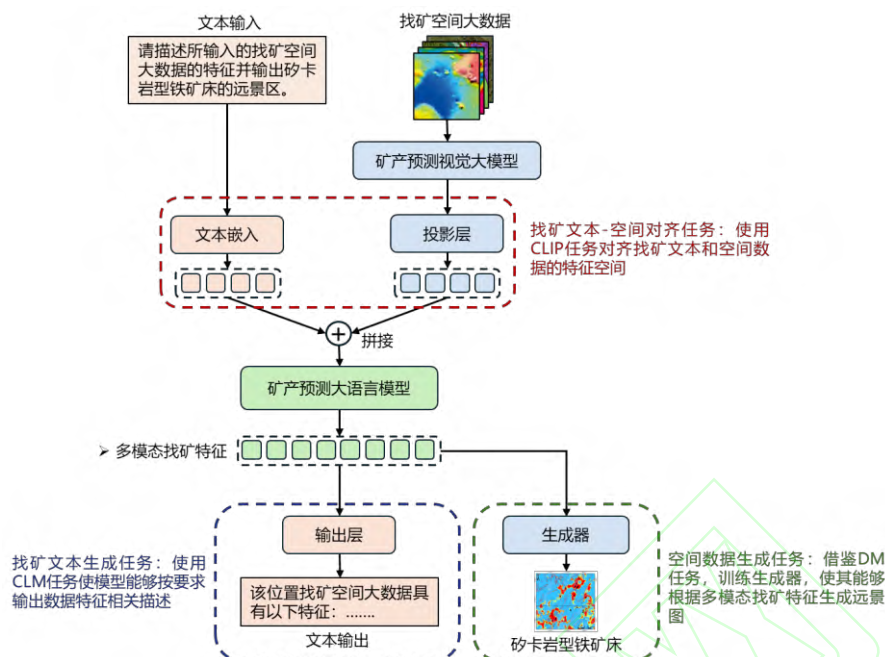


图 15. 矿产预测多模态大模型预训练示意图

Figure 15. An architecture of pre-training for the mineral prospectivity mapping multimodal foundation model.

(3) 模型微调。使用找矿文本-空间微调数据集 (图 14)，对图 15 中虚线内的浅层神经网络进行微调。同时使用基于人类反馈的强化学习来提升该模型文本输出的合理性。

(4) 模型部署应用：除 1.4 中提到的方式外，可为每一种空间数据的输入设置独立的接口以及微调方法。可预先使用提示词工程来引导模型对找矿文本的批量处理，如批量文献解析与知识总结、自动构建地质知识图谱等 (图 11)。同时，为减少模型出现幻觉，编造虚假信息，可选择检索增强生成 (Lewis et al., 2019) 和知识增强生成 (Liang et al., 2025) 等手段挂载知识库或知识图谱，使模型能够根据知识库中的内容对用户进行准确地反馈。

最终目标是构建一个以对话系统为核心的矿产预测多模态大模型，能够根据用户输入的文献资料或找矿空间数据，发现成矿知识和圈定找矿远景区或找矿靶区。该模型不仅具备跨区域泛化能力，还能通过语言交互实现知识驱动的智能矿产预测。

3 矿产预测大模型构建的挑战与建议

矿产预测大模型已逐渐成为突破传统人工智能方法局限的重要研究方向，其核心优势在于能够融合文本与空间等多模态数据，可提升跨区域泛化能力和地质任务适配性。然而，从 MineAgent 到 GFM4MPM 和 FM4GAI 等探索性实践来看，现有研究在数据集、特征表示、模型架构等方面仍存在如下挑战：

(1) 数据集。高质量、结构化的找矿大数据集获取难度大，且预处理流程复杂。本文所提出的矿产预测大模型需依赖多源找矿数据 (包括文献文本、地质、地球化学、地球物理、遥感数据等)。在预训练任务中，需大量优质的文本-空间数据对，而此类数据需耗费大量人力进行收集、清洗、标注，并依赖专家知识构建，显著增加了数据准备成本。由于大模型的知识来源依赖训练数据，数据获取的局限性直接制约了模型性能。为缓解文本数据预处理阶段的人力消耗，可借助现有大语言模型，如 Qwen、DeepSeek、GPT 等作为工具，

经提示词工程后对文本语料批量进行清洗、标注等预处理步骤。

(2) 找矿大数据表征。找矿大数据具有多源，异构，多维度，多时空分辨率，缺失值多等特征，如何高效表征这些数据作为大模型的输入是构建大模型的一个挑战。知识图谱作为一种结构化数据，通过构建找矿知识图谱来表征找矿大数据，将找矿大数据的多源特征作为属性并使用边来表示其异构特性。

(3) 预训练任务设计。现有预训练任务难以适配矿化特征提取需求。直接借鉴自然语言处理与计算机视觉领域的掩码重构、生成式扩散模型等任务，可能因找矿数据的特殊性 (如地质异质性、多模态耦合) 导致效果不佳。因此，需开发针对找矿场景的定制化预训练任务 (如基于地质规律的多模态对齐任务)，以提升模型对矿化特征的识别能力。

(4) 模型框架设计。传统方法难以高效处理找矿大数据的特征高维性与结构复杂性，亟需将现有的通用多模态大模型发展至矿产预测领域。由于从头训练大模型的计算开销巨大，因此需要利用现有通用大模型为基座，并引入矿产预测视觉大模型，为通用大模型提供多模态找矿数据感知能力。该方式可大幅缩短训练时间，从而推动矿产预测大模型的实际应用。

4 结论与展望

随着大模型在自然语言处理与计算机视觉领域的突破性进展，其强大的泛化能力与多任务适应性引发了学术界与工业界的广泛关注。矿产预测大模型作为语言-空间多模态融合的典型应用，通过整合文献文本中的成矿知识与找矿空间数据中的矿化特征，创新了从“数据驱动”到“数据驱动 + 知识引导”的新型找矿范式。具体而言，该模型具备“成矿知识挖掘与成矿规律提取”和“空间预测与找矿远景区或找矿靶区生成”两大核心优势。然而，矿产预测大模型的实际应用在数据层面、方法层面等仍面临多重挑战。

矿产预测大模型的未来研究方向需聚焦以下六个方面：(1) 构建大规模找矿数据集：系统收集准确丰富的大规模找矿文本 - 空间数据集，为矿产预测大模型构建提供坚实的数据支撑；(2) 探索找矿空间数据表征形式：创新多源空间数据表征方法，完整保留其中蕴含的找矿信息；(3) 定制化模型架构设计：针对性设计适配多源找矿数据的模型架构，实现模型对找矿信息的充分提取；(4) 定制化预训练任务设计：开发基于地质规律的多模态对齐任务，增强模型对地质异质性的适应能力；(5) 知识增强框架构建：通过地质知识库与大模型的深度融合 (如检索增强生成)，实现从“数据驱动”到“数据驱动 + 知识引导”的范式转变；(6) 动态更新机制探索：引入流数据处理框架与增量学习策略，赋予模型实时响应新数据的能力，提升预测的时效性与区域适配性。

致谢

感谢中国科学技术大学伍新明教授，中南大学邓浩副教授、江西理工大学孙涛副教授、中国地质大学 (北京) 张振杰副教授、中国地质大学 (武汉) 王成彬副教授等对本文初稿的有益讨论和建议。

References

- Achiam, J., Adler, S., Agarwal, S., et al., 2023. Gpt-4 Technical Report. arXiv:2303.08774. Retrieved from <http://arxiv.org/abs/2303.08774>.
- Agterberg, F. P., 1989. Computer Programs for Mineral Exploration. *Science*, 245(4913): 76-81. doi: 10.1126/science.245.4913.76.
- Awais, M., Naseer, M., Khan, S., et al., 2025. Foundation Models Defining a New Era in Vision: A Survey and Outlook. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(4): 2245-2264. doi: 10.1109/TPAMI.2024.3506283.
- Azulay, A., Weiss, Y., 2019. Why Do Deep Convolutional Networks Generalize So Poorly to Small Image Transformations? *Journal of Machine Learning Research*, 20(184): 1-25.
- Bai, J., Bai, S., Chu, Y., et al., 2023. Qwen Technical Report. arXiv:2309.16609. Retrieved from <http://arxiv.org/abs/2309.16609>.
- Baltrusaitis, T., Ahuja, C., Morency, L.-P., 2019. Multimodal Machine Learning: A Survey and Taxonomy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(2): 423–443. doi: 10.1109/TPAMI.2018.2798607
- Bao, H., Dong, L., Piao, S., et al., 2022. BEiT: BERT Pre-Training of Image Transformers. arXiv: 2106.08254. Retrieved from <http://arxiv.org/abs/2106.08254>.
- Bommasani, R., Hudson, D. A., Adeli, E., et al., 2022. On the Opportunities and Risks of Foundation Models. arXiv: 2108.07258. Retrieved from <http://arxiv.org/abs/2108.07258>.
- Brown, T. B., Mann, B., Ryder, N., et al., 2020. Language Models Are Few-shot Learners. In: *Advances in Neural Information Processing Systems*, 33, 1877-1901.
- Carranza, E. J. M., 2009. Geochemical Anomaly and Mineral Prospectivity Mapping in GIS. *Economic Geology*, 104(6): 890-890. doi: 10.2113/gsecongeo.104.6.890.
- Chowdhary, K., 2020. Natural Language Processing. In: *Fundamentals of Artificial Intelligence*. Springer India, New Delhi, 603-649. doi: 10.1007/978-81-322-3972-7_19.
- Collobert, R., Weston, J., Bottou, L., et al., 2011. Natural Language Processing (Almost) From Scratch. *Journal of machine learning research*, 12(7): 2493-2537.
- Dalal, N., Triggs, B., 2005. Histograms of Oriented Gradients for Human Detection. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), vol. 1, 886-893. doi: 10.1109/CVPR.2005.177.
- Daruna, A., Zadorozhnyy, V., Lukoczki, G., et al., 2024. GFM4MPM: Towards Geospatial Foundation Models for Mineral Prospectivity Mapping. In: *Proceedings of the 32nd ACM International Conference on Advances in Geographic Information Systems*, Association for Computing Machinery, New York, NY, USA, 565-568. doi: 10.1145/3678717.3691268.
- Deng, C., Zhang, T., He, Z., et al., 2024. K2: A Foundation Language Model for Geoscience Knowledge Understanding and Utilization. In: *Proceedings of the 17th ACM International Conference on Web Search and Data Mining*, 161-170. doi: 10.1145/3616855.3635772.
- Deng, J., Dong, W., Socher, R., et al., 2009. Imagenet: A Large-scale Hierarchical Image Database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, 248-255. doi: 10.1109/CVPR.2009.5206848.
- Devlin, J., Chang, M.-W., Lee, K., et al., 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 4171-4186. doi: 10.18653/v1/N19-1423.
- Dosovitskiy, A., Beyer, L., Kolesnikov, A., et al., 2020. An Image is Worth 16x16 Words: Transformers for Image

- Recognition at Scale. arXiv:2010.11929. Retrieved from <http://arxiv.org/abs/2010.11929>.
- Dong, H., Xie, S., 2024. Large Language Models (LLMs): Deployment, Tokenomics and Sustainability. arXiv: 2405.17147. Retrieved from <https://doi.org/10.48550/arXiv.2405.17147>.
- Fedus, W., Zoph, B., Shazeer, N., 2022. Switch Transformers: Scaling to Trillion Parameter Models with Simple and Efficient Sparsity. *Journal of Machine Learning Research*, 23(120): 1–39.
- Feng, X., Liu, S., Chen, H., et al., 2023. Continual Trajectory Prediction with Uncertainty-Aware Generative Memory Replay. In: 2023 IEEE International Conference on Data Mining (ICDM), 1013–1018. doi: 10.1109/ICDM58522.2023.00116.
- Fu, Y., Wang, M., Wang, C., et al., 2025. GeoMinLM: A Large Language Model in Geology and Mineral Survey in Yunnan Province. *Ore Geology Reviews*, 182: 106638. doi: 10.1016/j.oregeorev.2025.106638
- Gemma Team, Kamath, A., Ferret, J., et al., 2025. Gemma 3 Technical Report. arXiv: 2503.19786. Retrieved from <https://doi.org/10.48550/ARXIV.2503.19786>.
- Gong, L., Hou, X., Li, F., et al., 2025. Seedream 2.0: A Native Chinese-English Bilingual Image Generation Foundation Model. arXiv:2503.07703. Retrieved from <http://arxiv.org/abs/2503.07703>.
- Grattafiori, A., Dubey, A., Jauhri, A., et al., 2024. The Llama 3 Herd of Models. arXiv: 2407.21783. Retrieved from <https://doi.org/10.48550/ARXIV.2407.21783>.
- Guo, X., Lao, J., Dang, B., et al., 2024. SkySense: A Multi-Modal Remote Sensing Foundation Model Towards Universal Interpretation for Earth Observation Imagery. arXiv:2312.10115. Retrieved from <http://arxiv.org/abs/2312.10115>.
- Guo, Z., Wu, X., Liang, L., et al., 2025. Cross-Domain Foundation Model Adaptation: Pioneering Computer Vision Models for Geophysical Data Analysis. *Journal of Geophysical Research: Machine Learning and Computation*, 2(1): e2025JH000601. doi: 10.1029/2025JH000601.
- Han, S., Pool, J., Tran, J., et al., 2015. Learning Both Weights and Connections for Efficient Neural Network. In: Advances in Neural Information Processing Systems, vol. 28, Curran Associates, Inc.
- Han, Z., Gao, C., Liu, J., et al., 2024. Parameter-Efficient Fine-Tuning for Large Models: A Comprehensive Survey. arXiv: 2403.14608. Retrieved from <https://doi.org/10.48550/arXiv.2403.14608>.
- He, K., Zhang, X., Ren, S., et al., 2016. Deep Residual Learning for Image Recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 770-778.
- He, K., Chen, X., Xie, S., et al., 2021. Masked Autoencoders Are Scalable Vision Learners. arXiv:2111.06377. Retrieved from <http://arxiv.org/abs/2111.06377>.
- Hinton, G. E., Osindero, S., Teh, Y.-W., 2006. A Fast Learning Algorithm for Deep Belief Nets. *Neural Computation*, 18(7): 1527-1554. doi: 10.1162/neco.2006.18.7.1527.
- Hinton, G. E., Salakhutdinov, R. R., 2006. Reducing the Dimensionality of Data with Neural Networks. *Science*, 313(5786): 504-507. doi: 10.1126/science.1127647.
- Hinton, G., Vinyals, O., Dean, J., 2015. Distilling the Knowledge in a Neural Network. arXiv:1503.02531. Retrieved from <https://doi.org/10.48550/arXiv.1503.02531>.
- Ho, J., Jain, A., Abbeel, P., 2020. Denoising Diffusion Probabilistic Models. arXiv: 2006.11239. Retrieved from <https://arxiv.org/abs/2006.11239>.
- Hong, D., Zhang, B., Li, X., et al., 2024. SpectralGPT: Spectral Remote Sensing Foundation Model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(8): 5227-5244. doi: 10.1109/TPAMI.2024.3362475.
- Houlsby, N., Giurgiu, A., Jastrzebski, S., et al., 2019. Parameter-Efficient Transfer Learning for NLP. arXiv: 1902.00751. Retrieved from <https://doi.org/10.48550/ARXIV.1902.00751>.
- Hu, E. J., Shen, Y., Wallis, P., et al., 2022. Lora: Low-rank Adaptation of Large Language Models. arXiv:2106.09685. Retrieved from <https://doi.org/10.48550/arXiv.2106.09685>.
- Hu, Y., Yuan, J., Wen, C., et al., 2023. RSGPT: A Remote Sensing Vision Language Model and Benchmark.

- arXiv:2307.15266. Retrieved from <http://arxiv.org/abs/2307.15266>.
- Jacob, B., Kligys, S., Chen, B., et al., 2018. Quantization and Training of Neural Networks for Efficient Integer-Arithmetic-Only Inference. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2704-2713.
- Jacobs, R. A., Jordan, M. I., Nowlan, S. J., et al., 1991. Adaptive Mixtures of Local Experts. *Neural computation*, 3(1): 79-87. doi: 10.1162/neco.1991.3.1.79.
- Kirillov, A., Mintun, E., Ravi, N., et al., 2023. Segment Anything. arXiv:2304.02643. Retrieved from <http://arxiv.org/abs/2304.02643>.
- Krizhevsky, A., Sutskever, I., Hinton, G. E., 2012. Imagenet Classification with Deep Convolutional Neural Networks. In: Advances in Neural Information Processing Systems, volume 25, Curran Associates, Inc.
- Lazebnik, S., Schmid, C., Ponce, J., 2006. Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories. In: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), vol. 2, 2169-2178. doi: 10.1109/CVPR.2006.68.
- LeCun, Y., Boser, B., Denker, J. S., et al., 1989. Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Computation*, 1(4): 541-551. doi: 10.1162/neco.1989.1.4.541.
- LeCun, Y., Bottou, L., Bengio, Y., et al., 1998. Gradient-Based Learning Applied to Document Recognition. *Proceedings of the IEEE*, 86(11): 2278-2324. doi:10.1109/5.726791.
- Lewis, M., Liu, Y., Goyal, N., et al., 2019. BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension. arXiv:1910.13461. Retrieved from <http://arxiv.org/abs/1910.13461>.
- Lewis, P., Perez, E., Piktus, A., et al., 2020. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. In: Advances in Neural Information Processing Systems, Vol. 33, 9459–9474. Curran Associates, Inc.
- Liang, L., Bo, Z., Gui, Z., et al., 2025. KAG: Boosting LLMs in Professional Domains via Knowledge Augmented Generation. In: Companion Proceedings of the ACM on Web Conference 2025, WWW '25, 334–343, New York, NY, USA: Association for Computing Machinery. doi: 10.1145/3701716.3715240.
- Li, X. L., Liang, P., 2021. Prefix-Tuning: Optimizing Continuous Prompts for Generation. arXiv: 2101.00190. Retrieved from <https://doi.org/10.48550/ARXIV.2101.00190>.
- Liu, A., Feng, B., Xue, B., et al., 2024. Deepseek-v3 Technical Report. arXiv:2412.19437. Retrieved from <http://arxiv.org/abs/2412.19437>.
- Liu, F., Chen, D., Guan, Z., et al., 2024. RemoteCLIP: A Vision Language Foundation Model for Remote Sensing. *IEEE Transactions on Geoscience and Remote Sensing*, 62: 1-16. doi: 10.1109/TGRS.2024.3390838.
- Liu, Q., Ma, J., 2024. Foundation Models for Geophysics: Review and Perspective. arXiv:2406.03163. Retrieved from <https://doi.org/10.48550/arXiv.2406.03163>.
- Li, T., Zuo, R., Xiong, Y., et al., 2021. Random-Drop Data Augmentation of Deep Convolutional Neural Network for Mineral Prospectivity Mapping. *Natural Resources Research*, 30(1): 27–38. doi: 10.1007/s11053-020-09742-z.
- Liu, X., Ji, K., Fu, Y., et al., 2021. P-tuning v2: Prompt Tuning Can Be Comparable to Fine-Tuning Universally Across Scales and Tasks. arXiv:2110.07602. Retrieved from <http://arxiv.org/abs/2110.07602>.
- Liu, Y., Zhang, K., Li, Y., et al., 2024. Sora: A Review on Background, Technology, Limitations, and Opportunities of Large Vision Models. arXiv:2402.17177. Retrieved from <http://arxiv.org/abs/2402.17177>.
- Liu, Z., Lin, Y., Cao, Y., et al., 2021. Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. In: Proceedings of the IEEE/CVF international conference on computer vision, 10012-10022.
- Lin, Z., Deng, C., Zhou, L., et al., 2023. Geogalactica: A Scientific Large Language Model in Geoscience. arXiv:2401.00434. Retrieved from <http://arxiv.org/abs/2401.00434>.
- Lowe, D. G., 2004. Distinctive Image Features from Scale-Invariant Keypoints. *International journal of computer vision*, 60(2): 91-110. doi: 10.1023/B:VISI.0000029664.99615.94.

- Ma, X., 2022. Knowledge Graph Construction and Application in Geosciences: A Review. *Computers & Geosciences*, 161: 105082. doi: 10.1016/j.cageo.2022.105082
- Mikolov, T., Chen, K., Corrado, G., et al., 2013. Efficient Estimation of Word Representations in Vector Space. arXiv:1301.3781. Retrieved from <http://arxiv.org/abs/1301.3781>.
- Nadkarni, P. M., Ohno-Machado, L., Chapman, W. W., 2011. Natural Language Processing: an Introduction. *Journal of the American Medical Informatics Association*, 18(5): 544-551. doi: 10.1136/amiajnl-2011-000464.
- Ouyang, L., Wu, J., Jiang, X., et al., 2022. Training Language Models to Follow Instructions with Human Feedback. arXiv: 2203.02155. Retrieved from <http://arxiv.org/abs/2203.02155>.
- Poddar, L., Szarvas, G., Wang, C., et al., 2022. Deploying a Retrieval based Response Model for Task Oriented Dialogues. arXiv:2210.14379. Retrieved from <https://arxiv.org/abs/2210.14379>.
- Pennington, J., Socher, R., Manning, C. D., 2014. Glove: Global Vectors for Word Representation. In: Proceedings of the 2014 conference on empirical methods in natural language processing, 1532-1543.
- Qiu, Q., Wang, B., Ma, K., et al., 2023. A Practical Approach to Constructing a Geological Knowledge Graph: A Case Study of Mineral Exploration Data. *Journal of Earth Science*, 34(5): 1374-1389. doi: 10.1007/s12583-023-1809-3.
- Radford, A., Narasimhan, K., 2018. Improving Language Understanding by Generative Pre-Training. OpenAI Technical Report. Retrieved from https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf.
- Radford, A., Kim, J. W., Hallacy, C., et al., 2021. Learning Transferable Visual Models From Natural Language Supervision. arXiv: 2103.00020. Retrieved from <http://arxiv.org/abs/2103.00020>.
- Rombach, R., Blattmann, A., Lorenz, D., et al., 2021. High-Resolution Image Synthesis with Latent Diffusion Models. arXiv: 2112.10752. Retrieved from <https://doi.org/10.48550/arXiv.2112.10752>.
- Schuhmann, C., Beaumont, R., Vencu, R., et al., 2022. Laion-5b: An Open Large-Scale Dataset for Training Next Generation Image-Text Models. In: Advances in Neural Information Processing Systems, vol. 35, 25278-25294, Curran Associates, Inc.
- Sheng, H., Wu, X., Gao, H., et al., 2025a. On the Workflow, Opportunities and Challenges of Developing Foundation Model in Geophysics. arXiv: 2504.17384. Retrieved from <https://doi.org/10.48550/ARXIV.2504.17384>.
- Sheng, H., Wu, X., Si, X., et al., 2025b. Seismic Foundation Model: A Next Generation Deep-Learning Model in Geophysics. *Geophysics*, 90(2): IM59-IM79. doi: 10.1190/geo2024-0262.1.
- Shi, L., Xu, Y., Zuo, R., 2024. A Heterogeneous Graph Construction Method for Mineral Prospectivity Mapping. *Natural Resources Research*, 33(4): 1365–1376. doi: 10.1007/s11053-024-10344-2
- Shi, L., Zuo, R., 2025a. Geological Knowledge-Embedding Transfer-Learning Architecture for Geochemical Anomaly Identification. *Mathematical Geosciences*, 57: 821-844. doi: 10.1007/s11004-025-10190-9.
- Shi, L., Zuo, R., 2025b. Foundation Model for Geochemical Anomaly Identification. *Natural Resource Research*, Under Review.
- Simonyan, K., Zisserman, A., 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv:1409.1556. Retrieved from <https://doi.org/10.48550/arXiv.1409.1556>.
- Sutskever, I., Vinyals, O., Le, Q. V., 2014. Sequence to Sequence Learning with Neural Networks. In: Advances in Neural Information Processing Systems, volume 27, Curran Associates, Inc.
- Szeliski, R., 2022. Computer Vision: Algorithms and Applications. Springer Cham. doi: 10.1007/978-3-030-34372-9.
- Tang, D., Qin, B., Feng, X., et al., 2015. Effective LSTMs for Target-Dependent Sentiment Classification. arXiv:1512.01100. Retrieved from <https://doi.org/10.48550/arXiv.1512.01100>.
- Taylor, R. B., Steven, T. A., 1983. Definition of Mineral Resource Potential. *Economic Geology*, 78(6): 1268-1270. doi: 10.2113/gsecongeo.78.6.1268.
- Touvron, H., Lavril, T., Izacard, G., et al., 2023. LLaMA: Open and Efficient Foundation Language Models. arXiv:

2302.13971. Retrieved from <http://arxiv.org/abs/2302.13971>.

- Vaswani, A., Shazeer, N., Parmar, N., et al., 2017. Attention is All you Need. In: *Advances in Neural Information Processing Systems*, vol. 30, Curran Associates, Inc.
- Viola, P., Jones, M., 2001. Rapid Object Detection Using a Boosted Cascade of Simple Features. In: *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, vol. 1, I-I. doi: 10.1109/CVPR.2001.990517.
- Voulodimos, A., Doulamis, N., Doulamis, A., et al., 2018. Deep Learning for Computer Vision: A Brief Review. *Computational Intelligence and Neuroscience*, 2018 (1):7068349. doi: 10.1155/2018/7068349.
- Wang, B., Wu, L., Xie, Z., et al., 2022. Understanding Geological Reports Based on Knowledge Graphs Using a Deep Learning Approach. *Computers & Geosciences*, 168: 105229. doi: 10.1016/j.cageo.2022.105229.
- Wang, Z., Zuo, R., 2022. Mineral Prospectivity Mapping Using a Joint Singularity-Based Weighting Method and Long Short-Term Memory Network. *Computers & Geosciences*, 158: 104974. doi: 10.1016/j.cageo.2021.104974.
- Wei, J., Wang, X., Schuurmans, D., et al., 2022. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. In: *Proceedings of the 36th International Conference on Neural Information Processing Systems*.
- Xu, Y., Zuo, R., 2023. An Interpretable Graph Attention Network for Mineral Prospectivity Mapping. *Mathematical Geosciences*, 56: 169-190. doi: 10.1007/s11004-023-10076-8
- Xu, Y., Zuo, R., Bai, Y., 2025a. Geological Knowledge - Guided Dual - Branch Deep Learning Model for Identification of Geochemical Anomalies Related to Mineralization. *Journal of Geophysical Research: Machine Learning and Computation*, 2(1): e2024JH000468. doi: 10.1029/2024JH000468.
- Xu, Y., Zuo, R., Chen, Z., et al., 2025b. Recent Advances and Future Research Directions in Deep Learning as Applied to Geochemical Mapping. *Earth-Science Reviews*, 270: 105209. doi: 10.1016/j.earscirev.2025.105209.
- Yan, Q., Xue, L., Li, Y., et al., 2023. Mineral Prospectivity Mapping Integrated with Geological Map Knowledge Graph and Geochemical Data: A Case Study of Gold Deposits at Raofeng Area, Shaanxi Province. *Ore Geology Reviews*, 161: 105651. doi: 10.1016/j.oregeorev.2023.105651.
- Yan, Q., Zhao, J., Xue, L., et al., 2024. Mineral Prospectivity Mapping Based on Spatial Feature Classification with Geological Map Knowledge Graph Embedding: Case Study of Gold Ore Prediction at Wulonggou, Qinghai Province (Western China). *Natural Resources Research*, 33(6): 2385-2406. doi: 10.1007/s11053-024-10386-6.
- Yang, F., Wang, Z., Zuo, R., et al., 2023. Quantification of Uncertainty Associated with Evidence Layers in Mineral Prospectivity Mapping Using Direct Sampling and Convolutional Neural Network. *Natural Resources Research*, 32(1): 79-98. doi: 10.1007/s11053-022-10144-6
- Yang, F., Zuo, R., 2024. Geologically Constrained Convolutional Neural Network for Mineral Prospectivity Mapping. *Mathematical Geosciences*, 56: 1605-1628 doi: 10.1007/s11004-024-10141-w.
- Yin, B., Zuo, R., Xiong, Y., 2022. Mineral Prospectivity Mapping via Gated Recurrent Unit Model. *Natural Resources Research*, 31(4): 2065-2079. doi: 10.1007/s11053-021-09979-2
- Yin, S., Fu, C., Zhao, S., et al., 2024. A Survey on Multimodal Large Language Models. *National Science Review*, 11(12): nwae403. doi: 10.1093/nsr/nwae403.
- Yu, B., Shen, T., Na, H., et al., 2024. MineAgent: Towards Remote-Sensing Mineral Exploration with Multimodal Large Language Models. arXiv: 2412.17339. Retrieved from <https://doi.org/10.48550/arXiv.2412.17339>.
- Zaken, E. B., Ravfogel, S., and Goldberg, Y., 2021. BitFit: Simple Parameter-efficient Fine-tuning for Transformer-based Masked Language-models. arXiv: 2106.10199. Retrieved from <https://doi.org/10.48550/ARXIV.2106.10199>.
- Zhang, Q., Chen, M., Bukharin, A., et al., 2023. AdaLoRA: Adaptive Budget Allocation for Parameter-Efficient Fine-Tuning. arXiv: 2303.10512. Retrieved from <https://doi.org/10.48550/ARXIV.2303.10512>.
- Zhang, Y., Wang, Z., He, Z., et al., 2024. BB-GeoGPT: A Framework for Learning a Large Language Model for Geographic Information Science. *Information Processing & Management*, 61(5): 103808. doi:

10.1016/j.ipm.2024.103808.

- Zhang, Y., Wei, C., Wu, S., et al., 2023. Geogpt: Understanding and Processing Geospatial Tasks Through an Autonomous GPT. arXiv:2307.07930. Retrieved from <https://doi.org/10.48550/arXiv.2307.07930>.
- Zhang, Z., Zuo, R., and Xiong, Y., 2016. A Comparative Study of Fuzzy Weights of Evidence and Random Forests for Mapping Mineral Prospectivity for Skarn-Type Fe Deposits in the Southwestern Fujian Metallogenic Belt, China. *Science China Earth Sciences*, 59(3): 556–572. doi: 10.1007/s11430-015-5178-3
- Zhang, J., Huang, J., Jin, S., et al., 2024. Vision-Language Models for Vision Tasks: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(8): 5625–5644. doi: 10.1109/TPAMI.2024.3369699
- Zhao, W. X., Zhou, K., Li, J., et al., 2023. A Survey of Large Language Models. arXiv: 2303.18223. Retrieved from <http://arxiv.org/abs/2303.18223>.
- Zhou, C., Wang, H., Wang, C., et al., 2021. Geoscience Knowledge Graph in the Big Data Era. *Science China Earth Sciences*, 64(7): 1105–1114. doi: 10.1007/s11430-020-9750-4
- Zhou, C., Li, Q., Li, C., et al., 2023. A Comprehensive Survey on Pretrained Foundation Models: A History from BERT to ChatGPT. arXiv:2302.09419. Retrieved from <http://arxiv.org/abs/2302.09419>.
- Zuo, R., Carranza, E. J. M., 2011. Support Vector Machine: A Tool for Mapping Mineral Prospectivity. *Computers & Geosciences*, 37(12): 1967–1975. doi: 10.1016/j.cageo.2010.09.014
- Zuo, R., 2016. A Nonlinear Controlling Function of Geological Features on Magmatic-Hydrothermal Mineralization. *Scientific Reports*, 6(1): 27127. doi: 10.1038/srep27127.
- Zuo, R., 2020. Geodata Science-Based Mineral Prospectivity Mapping: A Review. *Natural Resources Research*, 29(6): 3415–3424. doi: 10.1007/s11053-020-09700-9.
- Zuo, R., Xiong, Y., Wang, Z., et al., 2023. A New Generation of Artificial Intelligence Algorithms for Mineral Prospectivity Mapping. *Natural Resources Research*, 32(5): 1859–1869. doi: 10.1007/s11053-023-10237-w
- Zuo, R., Xu, Y., 2023. Graph Deep Learning Model for Mapping Mineral Prospectivity. *Mathematical Geosciences*, 55(1): 1–21. doi: 10.1007/s11004-022-10015-z
- Zuo, R., Yang, F., Cheng, Q., and Kreuzer, O. P., 2025. A Novel Data-Knowledge Dual-Driven Model Coupling Artificial Intelligence with a Mineral Systems Approach for Mineral Prospectivity Mapping. *Geology*, 53(3): 284–288. doi: 10.1130/G52970.1.

中文参考文献

- 雷丹泓, 刘建忠, 2025. 多模态模型:透视月球演化的奥秘. *知识就是力量*, 2025, (04):12-13.
- 之江实验室, 2025. GeoGPT: 一个非营利性的地球科学探索研究项目. <https://geogpt.zero2x.org.cn/cn>.
- 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.
- 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. doi: 10.13745/j.esf.sf.2020.12.1.
- 左仁广, 成秋明, 许莹, 等, 2024. 可解释性矿产预测人工智能模型. *中国科学: 地球科学*, 54(9): 2917-2928.
- 左仁广, 2025. 智能矿产预测的技术挑战与解决方案. *中国科学:地球科学*, 55(9):3104-3119.