

引用格式: 郑栋宇, 王轶军, 白毅帆, 等. 2026. 从卷积神经网络到多模态大模型的沉积岩薄片智能分析[J]. 沉积学报. doi: 10.14027/j.issn.1000-0550.2026.029.
ZHENG DongYu, WANG YiJun, BAI YiFan, et al. 2026. From Convolutional Neural Networks to Multi-Modal Large Models for Intelligent Analysis of Rock Thin Sections [J]. Acta Sedimentologica Sinica, doi: 10.14027/j.issn.1000-0550.2026.029.
DOI: 10.14027/j.issn.1000-0550.2026.029 CSTR: 32268.14/j.cjxb.62-1038.2026.029

从卷积神经网络到多模态大模型的沉积岩薄片智能分析

郑栋宇^{1,2,3,4}, 王轶军³, 白毅帆⁴, 吴思萱⁵, 胡修棉⁶, 侯明才^{1,2}

1. 成都理工大学沉积地质研究院, 成都 610059
2. 自然资源部深时地理环境重建与应用重点实验室, 成都理工大学, 成都 610059
3. 计算机与网络安全学院, 成都理工大学, 成都 610059
4. 成都理工大学地球物理学院, 成都 610059
5. 中国石油天然气股份有限公司西南油气田分公司勘探开发研究院, 成都 610041
6. 南京大学地球科学与工程学院, 南京 210023

摘要 【意义】沉积岩薄片矿物组成与结构特征的准确识别, 是物源分析、沉积环境重建和储层评价的重要基础。传统偏光显微镜人工鉴定方法可靠性较高, 但存在效率低、主观性强和标准化程度不足等问题, 难以满足数据驱动沉积学研究对定量化和可重复分析的需求。【进展】围绕沉积岩岩石矿物智能识别的发展历程与技术演进, 系统梳理了传统图像处理、机器学习和深度学习方法在岩石薄片图像分析中的应用进展, 重点总结了卷积神经网络和视觉 Transformer 在分类、检测与分割任务中的应用特点及其对沉积学特征识别的适配方式。进一步归纳了弱监督、半监督和无监督学习在降低数据标注依赖方面的技术优势, 以及生成式模型在数据扩展、类别均衡和稀缺样本补充中的应用潜力; 同时分析了视觉大模型、语言大模型和多模态框架在融合地质知识、提升模型泛化能力和增强解释一致性方面的发展趋势。【结论与展望】总体而言, 岩石矿物智能识别正由单一模型应用向低标注学习、生成式模型与多模态大模型融合的综合分析范式转变。未来应加强标准化岩石薄片数据集建设、地质先验知识引入、模型跨域适配和可解释性评价, 构建具有高泛化能力与地质解释一致性的沉积岩智能分析体系, 为沉积学研究提供高效、可靠的技术支撑。

关键词 沉积岩智能分析; 岩石薄片; 卷积神经网络; 大模型; 人工智能; 深度学习

第一作者 郑栋宇, 男, 1992 年出生, 博士, 副教授, 沉积学大数据与人工智能, E-mail: zhengdongyu@cdut.edu.cn

中图分类号: TE142 文献标志码: A 文章编号: 1000-0550 (2026) 00-0000-00

0 引言

沉积岩记录了地表过程、地壳演化以及能源与矿产资源形成的重要信息。其中, 沉积岩的颗粒组成与结构特征是开展物源分析、沉积环境判识以及储层质量评价的基础。自经典的沉积岩分类与物源分析体系建立以来, 基于偏光显微镜薄片观察的岩石矿物识别一直是沉积岩石学研究的核心内容。通过系统识别骨架颗粒的类型及其结构特征, 可为沉积体系演化、构造背景判别及油气储层质量评价提供重要约束 (Dunham, 1962; Folk, 1966; Dickinson and Suczek, 1979; Bloch *et al.*, 2002; Garzanti *et al.*, 2007)。传统岩石矿物识别主要依赖人工薄片鉴定与点计数方法, 即在单偏光 (Plane-polarized light, PPL) 和正交偏光 (Cross-polarized light, XPL) 条件下, 对单个薄片 300~400 个骨架颗粒进行逐点观察和统计 (Weltje, 2002)。该方法能够获得可靠的颗粒成分与粒径数据, 长期以来被视

为岩相分析的“金标准”。然而,该过程高度依赖岩石学家的经验积累,工作量大、效率低,且在样品数量快速增长的情况下,难以保证结果的一致性与可重复性。随着沉积学研究逐步向盆地尺度、区域对比及数据驱动方向发展,传统人工方法在效率和标准化方面的局限性日益凸显(Joseph *et al.*, 2016)。

在此背景下,岩石薄片分析逐渐向数字化方向发展。早在 20 世纪 90 年代,研究者已尝试利用数字薄片图像,结合 PPL、XPL 及元素面扫描等多源信息,对颗粒和孔隙进行识别(Terribile and FitzPatrick, 1992; Launea *et al.*, 1994)。随后,二值分割、分水岭算法、高斯滤波、区域生长等传统计算机视觉方法被引入岩石薄片图像分析,用于颗粒与孔隙的分离及边界识别(Francus, 1998; Goodchild and Fueten, 1998; Barraud, 2006; Gorsevski *et al.*, 2012; Asmussen *et al.*, 2015; 王志忠等, 2017; 刘春等, 2018)。尽管这些方法在推动岩石学分析数字化方面发挥了重要作用,但其通常依赖人工设定特征,对图像质量和参数选择较为敏感,且在复杂矿物组合条件下难以实现可靠的类型识别。

21 世纪以来,机器学习算法逐步应用于岩石矿物识别研究领域。国际上早期围绕碳酸盐岩纹理识别、显微岩石图像自动分类、石灰岩图像视觉分类开展的一系列研究,验证了图像处理、神经网络、支持向量机等技术实现岩石图像自动判别的可行性(Marmo *et al.*, 2005; Chatterjee, 2013; Młynarczuk *et al.*, 2013); 同期国内也出现依托概率神经网络完成岩石薄片图像分类识别的相关探索(程国建等, 2013)。后续研究进一步将朴素贝叶斯、支持向量机、K-means 聚类、反向传播神经网络等算法运用于薄片图像矿物颗粒自动识别与边界检测工作(Maitre *et al.*, 2019)。但上述传统机器学习方法仍存在明显局限:模型输入依赖人工提取颜色、纹理、形态学等图像特征,识别精度高度受制于人工特征工程的设计质量,很难适配岩石薄片内部复杂、非均质程度高的光学特征。

近年来,深度学习技术的发展为岩石矿物智能识别提供了全新研究范式,以卷积神经网络(Convolutional Neural Networks, CNNs)为代表的模型无需人工提取特征,可自动从原始图像中学习特征,为岩石矿物自动识别提供了高效技术路径(徐述腾和周永章, 2018; 张野等, 2018; Reichstein *et al.*, 2019; Su *et al.*, 2020; Tang *et al.*, 2020; Saxena *et al.*, 2021; Zheng *et al.*, 2022)。但直接应用通用 CNN 模型存在明显局限,难以适配岩石薄片的复杂特征。随着人工智能技术的进一步发展,以视觉 Transformer(Vision Transformer, ViT)为核心的新一代视觉模型,凭借全局关系建模的优势被应用于岩石图像分析,展现出良好潜力(Dosovitskiy *et al.*, 2021; Cao *et al.*, 2022; Koeshidayatullah *et al.*, 2022; Hou *et al.*, 2024; Zheng *et al.*, 2024a, 2024b),但其应用仍受数据规模、标注成本等因素限制,相关适配研究有待深入。

针对岩石薄片标注数据稀缺这一核心瓶颈,弱监督、半监督学习方法逐步被探索应用,旨在降低对精细人工标注的依赖(刘烨和吕锦涛, 2023)。同时,生成式模型的引入,为人工生成岩石显微图像、扩充训练数据集提供了新途径(Ferreira *et al.*, 2022)。此外,多

模态学习与大规模基础模型的兴起，进一步为突破数据瓶颈、提升模型性能提供了新路径（Luo *et al.*, 2025）。这类模型依托海量跨领域数据预训练，可整合多源信息、缓解样本不足问题，同时结合多模态融合技术，能够更好地适配岩石薄片的复杂光学特征，为构建泛化能力强、可解释性好的智能识别系统提供支撑（Fan *et al.*, 2025）。总体而言，岩石矿物智能识别正逐步从单一模型的直接应用，迈向以模型适配、弱监督/半监督学习、生成式模型辅助、大模型与多模态融合为核心的新一代智能分析范式。

1 卷积神经网络与视觉 Transformer 的直接调用与适配性调整

深度学习是一类依托多层网络结构、以数据驱动方式自动挖掘数据内在特征，进而完成分类、识别、回归等复杂任务的人工智能技术，其中计算机视觉是深度学习应用最广泛、技术体系最成熟的核心领域。卷积神经网络凭借局部感受野、权值共享、层级化特征提取的独特优势，在图像分析任务中具备极高的效率与精度，因此被大量应用于岩石薄片图像智能解译，主要覆盖三大核心任务：（1）图像分类，如判别岩石样品为碎屑岩或碳酸盐岩；（2）目标检测，如通过边界框精准定位鲕粒灰岩中的鲕粒目标；（3）图像分割，如精细勾勒砂岩中碎屑颗粒的轮廓并完成矿物类型判别（图 1；蔡宇恒等，2020；郝慧珍等，2021；徐圣嘉等，2022；杜崇娇等，2023；余晓露等，2023；张杰等，2023；张利军等，2024；邓乃尔等，2025；刘恒等，2025）。

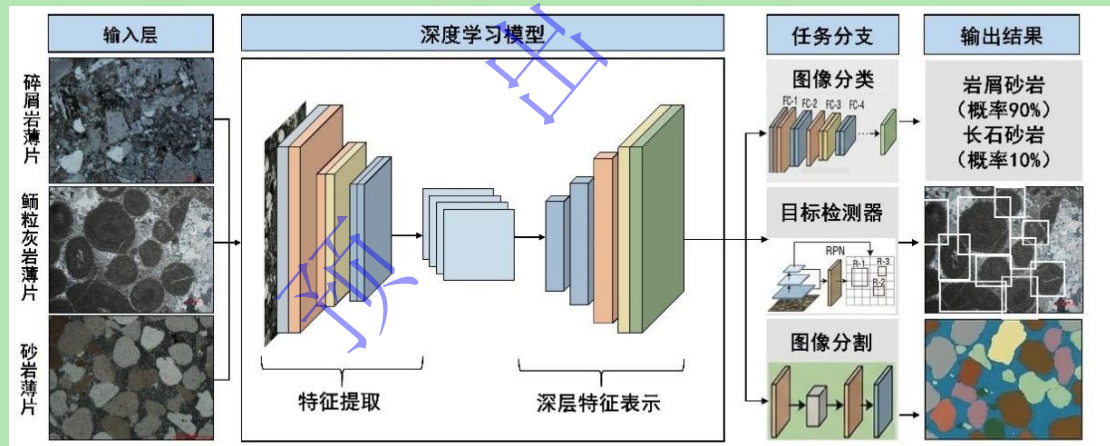


图 1 岩石薄片图像分类、目标检测与图像分割中的典型应用示意

Fig. 1 Representative applications of rock thin-section images in image classification, object detection, and image segmentation

2020 年前后，沉积岩岩石薄片图像分析通常直接移植通用模型，如 ResNet、U-Net、DeepLabV3 等（徐述腾和周永章，2018；张野等，2018；蔡宇恒等，2020；Su *et al.*, 2020；Tang *et al.*, 2020；Saxena *et al.*, 2021；徐圣嘉等，2022；Zheng *et al.*, 2022；杜崇娇等，2023；余晓露等，2023；张杰等，2023；刘恒等，2025）。随着相关研究持续深入，随后的研究呈现出“直接移植通用模型”向“针对岩石矿物特性的适配性调整”转变的明显趋势。这类适配性调整主要体现在三个方面：（1）在模型选择与任务范式层面，相关研究逐渐由通用的 CNN 基础模型，转向根据沉积学目标在“分类—检测—分割”不同任务范式中选择更合适的基准模型；（2）在方法层面，从单纯追求分类精度转向挖掘模型判别过程中所蕴含

的地质意义,强调结果与岩相学认知的一致性;(3)在数据层面,针对岩石薄片在偏光条件、成像机制及矿物光学响应上的特殊性,对模型输入和特征融合方式进行针对性设计。上述调整推动深度学习方法在沉积学中的角色由辅助分类工具逐步转向服务于岩石矿物定量分析与地质解释的技术手段。

1.1 模型选择的适配性调整

在深度学习方法引入沉积地质学研究的过程中,早期工作往往倾向于将成熟的通用 CNN 模型直接应用于薄片图像分析,而较少区分不同沉积学问题在本质上所对应的视觉任务类型。这种统一模型解决多类问题的做法在一定程度上推动了自动识别研究的起步,但也逐渐暴露出方法适配性不足的问题。随着应用经验的积累,研究人员开始认识到,沉积学中不同研究目标,例如岩相判别、颗粒计数、孔隙识别或结构组构分析,在信息需求和输出形式上存在显著差异,其对应的计算机视觉任务类型亦不相同,因而需要采用在相应任务上更为高效和稳定的模型框架。

在目标检测范式下, Koeshidayatullah *et al.* (2020) 基于 YOLOv3 对碳酸盐岩薄片进行针对性训练,实现了在同一图像中对鲕粒、生屑等颗粒类型以及胶结物和孔隙结构的自动定位与分类,并进一步统计其数量与空间分布特征。该研究展示了目标检测模型在处理颗粒尺度明确、边界相对清晰的沉积结构单元时,能够将深度学习输出直接转化为沉积学中常用的定量指标,为组分比例分析与储层评价提供数据支撑。类似地, Liu and Song (2020) 面向碳酸盐微相分析中的化石和非生物颗粒识别,构建了包含多类骨骼颗粒与非生物颗粒的图像数据集,并比较多种深度卷积神经网络的识别效果,表明深度学习在碳酸盐微相分类和古生态解释辅助分析中具有较高应用潜力。此外, Dos Anjos *et al.* (2021) 将深度学习用于碳酸盐岩 micro-CT 图像岩性分类; Bihani *et al.* (2022) 则面向泥岩 SEM 图像开展语义分割,说明相关方法已从偏光薄片扩展到多尺度岩石显微图像分析。在图像分割任务中, Saxena *et al.* (2021) 系统比较了 VGG16、VGG19 以及 DeepLab v3+ 等主流分割框架,并结合砂岩薄片矿物边界连续、孔隙结构复杂及类别不均衡的特点进行评估。结果表明, DeepLab v3+ 在保持颗粒边界连续性与减少噪声区域方面优于传统 VGG 系列网络,在视觉一致性与地质可解释性方面表现更优。在图像分类任务中, Zheng *et al.* (2022) 针对碎屑岩中常见重矿物锆石阴极发光 (CL) 图像类型判别问题,将研究目标明确限定于整颗晶体尺度上的成因分类,从而选用与任务属性高度匹配的分类基准模型。其基于 ResNet-18、ResNet-50 与 VGG-19 等主流 CNN 架构开展对比实验,结果表明,在样本规模有限的条件下,结合迁移学习的 ResNet-18 在精度、收敛稳定性与计算效率之间实现了较优平衡。随着模型架构的演进, Vision Transformer 因其在全局特征建模与复杂空间结构表达方面的优势,被逐渐引入需要精细结构刻画的薄片分割任务。Zheng *et al.* (2024b) 基于 Vision Transformer 构建分割框架,实现了沉积颗粒与显微结构的像素级识别,使模型输出形式更接近沉积学者在显微镜下对颗粒边界进行描绘与统计的传统分析过程。这种模型范式的转

变，不仅提升了分割结果在空间表达上的完整性，也为矿物比例计算与显微结构定量分析提供了更直接的数据基础。

1.2 研究思路的适配性调整

尽管近年来自动化图像分析方法在岩石薄片识别中的性能不断提升，但无论是深度学习模型还是传统图像处理方法，其分析结果是否真正符合沉积学和岩相学的判据，始终是一个需要被审视的核心问题。在沉积学研究中，岩石类型的判定并非单纯基于图像纹理差异，而是依赖矿物组成、颗粒类型、结构组构及其空间组合关系等具有明确地质意义的特征。因此，方法本身是否能够围绕这些关键解释单元展开分析，往往比结果精度的提升更为重要。在这一背景下，部分研究开始将关注重点由“模型能否给出正确结果”，转向“分析过程是否与沉积学认知一致”。这种转向既体现在对深度学习模型判别依据的解释与验证上，也体现在针对具体沉积学问题，对传统图像分析方法进行重新设计与适配的探索中。相关工作不再简单依赖模型输出结论，而是试图通过引入可解释性分析或基于地质对象特征的规则约束，使自动分析过程能够与沉积学者的显微判读逻辑建立对应关系。

在计算机视觉领域，为揭示深度学习模型的判别依据，研究者通常采用类激活映射（Class Activation Mapping, CAM）（Selvaraju *et al.*, 2020）及其改进方法，对模型在判别过程中重点关注的图像区域进行可视化。该类方法通过分析模型在分类决策阶段对不同空间位置特征的响应强度，直观展示哪些区域对最终判别结果贡献最大，从而为理解模型的“关注点”提供依据。Zheng *et al.*（2024b）将这一研究思路引入沉积地质学，把研究重点转向模型判别依据的地质合理性验证。通过引入 Grad-CAM 可视化方法与多尺度感受野，作者对深度学习模型在进行岩相或岩类判别时所关注的薄片区域进行直观展示，进而分析模型是否主要聚焦于矿物颗粒、结构组构及其空间分布等具有明确沉积学意义的特征，而非与岩相无关的背景信息或成像噪声。该研究不仅比较了不同模型在分类精度上的表现差异，更重要的是从判别依据层面评估了模型输出结果与岩相学认知之间的一致性，其研究思路随后被快速推广到钻井岩心与岩石矿物识别任务（Lin *et al.*, 2025; Di Martino *et al.*, 2026）。

1.3 数据层面的适配性调整

在进一步的研究中，深度学习方法在沉积学中的适配性调整逐步深入至岩石薄片成像数据本身的物理与光学特性层面。与自然图像或工业视觉数据不同，岩石薄片图像通常依赖偏光显微镜获取，不同偏光条件下矿物的光学响应差异显著，这种差异直接影响模型对颗粒边界、内部结构及显微特征的识别稳定性。例如，在正交偏光条件下，矿物消光现象可能导致颗粒难以与背景区分，从而引发欠分割；而在单偏光条件下，部分透明或弱对比矿物内部结构不明显，容易与背景区域发生混淆。若仅依赖单一成像条件进行建模，模型判别结果往往受制于成像模式本身，而难以稳定反映真实的沉积学结构信息。

针对这一问题，研究人员开始在数据层面对输入信息进行针对性设计，通过多偏光、多视角或多模态数据的联合建模增强模型对沉积学关键信息的感知能力（图2）。以 Ge *et al.*（2021）的研究为代表，该工作并未直接引入复杂的分割或检测框架，而是从碎屑颗粒在不同偏光条件下呈现出的互补光学特征出发，构建了同时输入单偏光与正交偏光薄片图像的双输入深度学习模型。该模型在特征提取阶段分别对两种偏光条件下的图像进行编码，并通过动态特征融合模块在空间与通道维度上自适应整合两类信息，从而强化模型对具有判别意义的干涉条纹、消光特征和矿物内部结构的关注。在此基础上，后续研究进一步将多偏光思想从分类任务拓展至图像分割层面。Zheng *et al.*（2024a）提出了一种面向沉积颗粒识别的自动图像分割方法。该研究采用基于 Vision Transformer 的模型结构，并引入双通道特征融合策略，在特征提取阶段对同一视角下不同偏光条件获取的薄片图像进行多层次融合，从而增强模型对颗粒边界与内部结构的整体判别能力。其的核心在于，通过在统一视角和空间对应关系下整合不同偏光图像信息，减弱单一偏光条件对分割结果的制约，提高颗粒级分割的稳定性。Zeng *et al.*（2026）的研究进一步将双模态思想由“同一视角下的特征互补”拓展至“不同成像条件下结果一致性”的问题。与前述方法侧重于特征层面的联合建模不同，这项研究针对单偏光与正交偏光图像在光学响应机制上的本质差异，设计了专门的多模态融合分割框架，通过显式建模不同偏光条件下矿物表现形式的差异与互补关系，提高分割结果在不同成像条件之间的空间一致性。该方法在保证分割精度提升的同时，有效减少了由偏光条件变化引起的结果波动，使图像分割结果更接近沉积学者在显微镜下综合多偏光信息进行判读的实际过程。

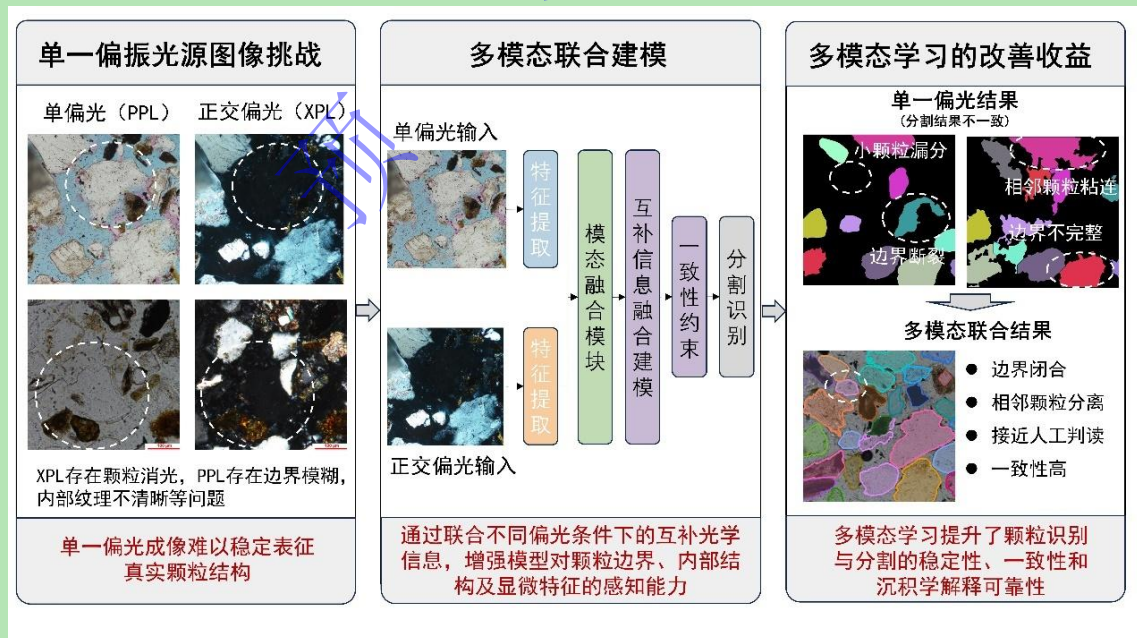


图2 针对岩石薄片图像物理和光学特性复杂性的解决方案与应用示例
Fig. 2 Solutions to the challenges posed by the complex physical and optical characteristics of rock thin-section images, together with representative application examples

2 研究趋势：从神经网络向通用大模型的过渡

如前所述，近年来深度学习方法在岩石薄片分析与沉积岩识别领域取得了一系列具有代表性的研究进展。通过在研究对象、判别依据及成像特性等层面的适配性调整，相关工作在特定岩类和成像条件下实现了较为可靠的自动识别与定量分析，初步验证了数据驱动方法在沉积学研究中的可行性与应用潜力。然而，现有研究亦暴露出若干共性局限：其一，多数方法仍高度依赖像素级或颗粒级的精细人工标注，而高质量薄片标注需由具备岩相学背景的研究人员逐一完成，时间成本和人力成本较高，难以支撑大规模数据集的构建；其二，不同类型岩石及其沉积—成岩背景在矿物组成与显微结构特征上差异显著，现有模型往往依赖特定区域、岩类或尺度条件进行训练，在跨区域、跨岩类及跨尺度应用中泛化能力有限；其三，沉积过程与成因机理等地质先验信息尚未得到系统融入，制约了自动分析结果在地质解释层面的深化应用（周嵘等，2023；吴博等，2024；Luo *et al.*, 2025）。

针对上述问题，本文归纳了三类具有代表性的研究方向：其一，引入弱监督、半监督及无监督等低标注依赖学习策略，以降低对精细人工标注的依赖，充分挖掘已有但未标注薄片数据的利用价值；其二，借助生成式模型对样本分布进行扩展，补充稀有岩相与类别不平衡样本，为模型训练和沉积学分析提供更完整的数据支撑；其三，通过引入视觉大模型、语言大模型及多模态框架，探索将图像信息与地质知识相结合的新范式，以提升模型的泛化能力与地质解释一致性。上述三类研究思路如图 3 所示，表 1 进一步汇总了各类方法的代表性模型与现阶段研究难点。

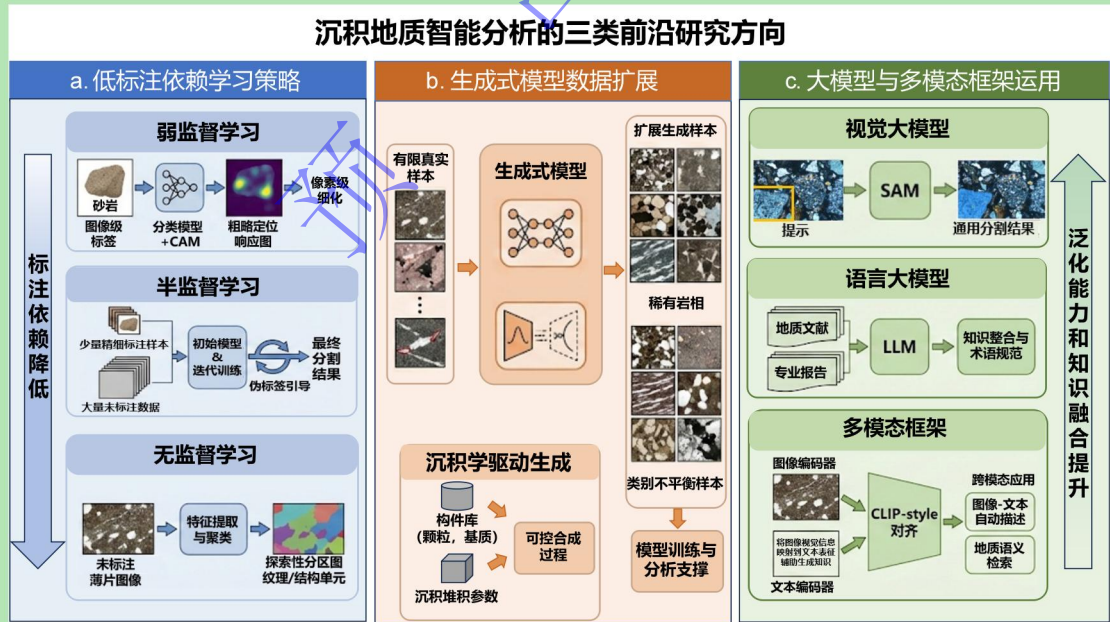


图 3 沉积岩岩石图像智能分析的三类前沿研究方向示意图

Fig. 3 Schematic illustration of three emerging research directions for intelligent analysis of sedimentary rock images

表 1 沉积岩岩石图像智能分析的三类前沿研究问题与难点分析
Table 1 Three emerging research directions for intelligent analysis of sedimentary rock images: key challenges and research gaps

研究方向	主要解决的问题	代表性应用模型	当前研究难点
低标注依赖学习策略	缓解岩石薄片分析对像素级、颗粒级精细人工标注的高度依赖；提高大量未标注或弱标注薄片数据的利用率；降低构建训练集的人力与时间成本	弱监督学习：分类模型 + CAM/注意力响应图，实现从图像级标签到像素级粗定位； 半监督学习：初始模型 + 伪标签迭代 + 一致性训练/自训练； 无监督学习：特征提取与聚类、原型学习、自监督表征学习，用于探索性分区或潜在类别发现	弱监督定位结果往往较粗，边界精度不足； 半监督方法对伪标签质量敏感，错误易累积传播； 无监督结果与沉积学语义之间对应关系不稳定，可解释性不足； 不同岩类、不同尺度下迁移能力仍有限
生成式模型数据扩展	针对样本量不足、稀有岩相缺失、类别不平衡等问题扩展训练数据分布；补充难获取或低频岩相样本；为模型训练和沉积学分析提供更完整的数据支撑	GAN/扩散模型/条件生成模型：生成薄片图像、视野外重建、类别条件控制生成； 沉积学驱动生成：将颗粒、基质、孔隙、层理等先验约束融入生成过程； 合成数据增强：生成样本 + 下游分类/检测/分割联合训练	生成样本的真实性与地质合理性难以同时保证； 易出现“视觉逼真但沉积学不合理”的伪特征； 生成数据与真实数据之间存在域差异，过度依赖可能导致模型偏移； 稀有岩相的可控生成、类别平衡生成和定量评价标准仍不成熟
大模型与多模态框架应用	提升模型在跨区域、跨岩类、跨尺度条件下的泛化能力；缓解传统模型仅依赖视觉纹理、难以融入地质知识的问题；增强自动分析结果与地质解释的一致性	视觉大模型：SAM 等用于通用分割、提示式分割和标注辅助； 语言大模型：LLM 用于地质文献、专业报告与知识规则整合； 多模态框架：CLIP-style 图文对齐、图像编码器 + 文本编码器、薄片图像描述与检索、图像-文本联合推理	通用大模型与薄片图像之间存在明显领域差异，直接迁移效果受限； 高质量图文配对数据和专业语料仍稀缺；地质知识如何结构化注入模型仍不成熟； 多模态输出虽更具解释性，但其结论的专业可靠性、可验证性与可控性仍需加强

2.1 弱监督、半监督与无监督学习：监督策略的演进与低标注学习框架

监督学习与弱监督、半监督、无监督学习的核心差异在于对标注信息的完整性、精确性及依赖程度不同：监督学习需依赖全量、精确的标签信息完成模型训练，而弱监督、半监督与无监督学习则通过弱化标注要求（如采用低粒度标签、部分标注样本）或完全脱离人工标注（仅利用数据自身结构特征），降低对精细标注的依赖。前文介绍的无论是基于目标检测进行颗粒与结构单元定位统计，还是通过图像分割实现矿物与孔隙的像素级刻画，均属于依赖地质专家提供的大量高质量精细标注的监督学习。该策略在验证深度学习方法有效性与提升任务精度方面发挥了重要作用，但其高昂的标注成本和时间投入限制了在大规模薄片数据及跨区域研究中的推广。因此，越来越多研究逐步转向以降低人工精细标注依赖为核心目标的学习范式，通过引入低粒度标签、未标注样本或利用数据自身结构信息，提高数据利用效率，缓解精细标注不足带来的应用瓶颈。

弱监督学习通过以图像级岩类或岩相标签替代像素级精细标注，从根本上降低了数据标注成本。相较于逐像素勾绘矿物边界与孔隙形态，图像级标签仅需对整幅薄片给出类别或沉积组构判定，标注效率显著提高，而无需构建大规模精细分割数据集。在此基础上，Li *et al.*（2026a）提出的弱监督图像分割框架首先利用图像级标签训练分类模型，并通过类激活映射（Selvaraju *et al.*, 2017; Zheng *et al.*, 2024b）提取模型在判别过程中关注的关

键区域，将原本仅包含整体类别信息的监督信号转化为空间响应图。随后，研究通过知识蒸馏与基于纹理相似度的空间传播策略，对高置信区域进行扩展与边界细化，使模型逐步逼近像素级分割结果，实现由“图像级语义”向“像素级结构”的逐层映射，在无需大量精细标注的前提下完成关键结构单元的分割。

半监督学习通过“少量精细标注+大量未标注数据”的协同训练机制，仅对少量代表性薄片进行精细标注以建立初始模型，再利用未标注数据生成高置信伪标签并迭代优化，使模型在更大数据范围内学习稳定的矿物纹理与结构特征。Han and Liu (2024) 提出的砂岩薄片半监督分割框架，基于 GL-SLIC 超像素划分图像单元，仅对少量碎片进行人工标注训练分类器，并结合伪标签筛选与判别器复核逐步扩充训练集，最终实现矿物与孔隙的全图分割。实验表明，该方法在极少标注条件下即可获得稳定且边界清晰的结果，部分指标优于完全监督 U-Net 模型，体现了半监督策略通过有效利用未标注数据减少对大规模精细标注依赖的优势。

无监督学习则完全不依赖人工标注，其目标在于利用薄片图像自身所包含的纹理特征、结构差异及空间组合关系，对矿物或结构单元进行自动分组与模式发现。以薄片图像无监督分割为例，Barbosa *et al.* (2024) 提出了一种无监督砂岩薄片自动分割方法，通过无监督卷积神经网络学习像素级纹理表征，并结合超像素与图划分策略获得结构分区结果。该方法在颗粒识别与粒度分布统计方面与传统筛分实验结果具有较好一致性，为矿物统计分析 & 专家快速标注提供了可靠的候选方案。

综上，弱监督、半监督与无监督学习通过低标注策略，有效缓解了高质量岩石薄片标注数据稀缺问题，充分盘活了海量未标注地质图像资源。但这类方法核心仍围绕数据特征聚类展开，易受地质体异质性影响，在复杂场景下类别易混淆，且存在聚类结果地质意义不明确等问题。未来研究需要进一步融合地质先验知识、优化特征表征与跨域适配机制，进一步提升模型在复杂地质场景中的泛化能力。

2.2 生成式模型：数据生成的演进与稀缺样本增强

在监督策略不断优化的基础上，尽管精细标注压力得到一定缓解，但高质量样本数量与类型覆盖不足的问题仍然存在。不同岩类、沉积环境与成岩阶段对应的显微结构差异显著，且部分岩相与特殊组构在实际样本中出现频率较低，单纯依赖现有数据难以充分表征沉积岩显微结构的复杂性。在此背景下，生成式模型通过学习数据整体分布并合成具有统计一致性的新增样本，可在一定程度上缓解样本稀缺、类别不平衡及分布覆盖不足等问题，为沉积学数据体系的完善与模型训练提供重要支撑。以 Ransinangue *et al.* (2025) 提出的沉积学驱动生成框架 SynSection 为例，该研究首先从少量真实薄片中提取并标注大量颗粒实例及连续背景纹理块，构建可复用的“颗粒—基质”构件库；其中，薄片由基质与胶结物共同构成的连续背景相被视为承载颗粒的纹理场，并采用轻量化 StyleGAN-2 (Karras *et al.*, 2020)

对其显微纹理特征进行建模与生成,从而获得分辨率与风格可控的背景基底。在此基础上, SynSection 的核心生成过程并非对整幅薄片进行自由采样,而是显式模拟沉积学约束下的二维颗粒堆积(packing)过程,即在图像平面内逐步放置颗粒实例,并利用一组具有沉积学可解释性的参数对生成结果进行控制,包括颗粒类型比例、填充度、空间邻近关系、取向特征及结构非均一性等。针对颗粒贴图与背景基质在颜色分布与尺度特征上的潜在不连续问题,该项研究进一步引入像素分布自适应与多尺度金字塔融合方法,实现显微纹理在颗粒—基质界面处的平滑过渡。实验结果表明,利用 SynSection 合成数据对下游分类与分割网络进行预训练,并结合少量真实薄片进行微调,可显著提升模型在精度、稳定性与泛化能力方面的表现。总体而言, SynSection 的核心贡献在于将生成式模型从“黑箱式纹理采样”转变为具有沉积学可解释性、参数可控性与自动标注能力的工程工具,为缓解薄片样本稀缺与类别不平衡问题提供了一条切实可行的技术路径。

在此基础上,未来生成式研究需要进一步充分考虑沉积学与岩石学知识。例如, SynSection 工作尚未充分考虑岩石薄片在单偏光与正交偏光两种成像模式下的差异性与对应性。两种偏光条件在矿物光学响应与纹理表现上存在显著差异,但在空间位置与结构单元层面具有严格的一一对应关系。如何在生成过程中同步刻画不同偏光条件下的成像特征,并保持其空间结构与语义属性的一致性,是实现多模态薄片生成与联合分析的关键问题。此外,生成结果的评价标准不应局限于视觉逼真度,而应关注其是否符合沉积学与岩石学基本规律。若未能体现合理的矿物共生关系、颗粒接触方式及成岩改造特征,即便图像外观接近真实样本,也难以支撑地质解释。因此,有必要将沉积过程、成岩机理及岩相学经验等地质知识引入生成模型,对结构组织与参数分布进行约束。在充分融入地质机理的前提下,生成式模型方能由数据扩增工具转变为具有解释潜力的沉积地质辅助分析手段。最后,将生成的海量岩石图片与公开沉积岩显微图像数据标准、碳酸盐岩薄片基准数据集相结合,形成可复用、可验证、可对比的开放数据资源(胡修棉等, 2020; Li *et al.*, 2026b),将有力推动沉积岩智能识别。

2.3 大模型与多模态框架:多源信息融合与跨模态地质表征

相较于通过降低标注粒度或生成标注样本规模来缓解数据瓶颈的弱监督、半监督与生成式方法,视觉大模型、语言大模型及多模态框架依托大规模预训练形成的通用表征能力,更直接面向跨区域、跨岩类与跨尺度应用中的泛化问题,并为地质知识的系统引入提供了新的技术路径。

以 Segment Anything Model (SAM) 图像分割大模型及其后续工作为代表的通用模型,代表了分割基础模型在图像理解任务中的重要进展。SAM 通过在大规模通用图像数据上的预训练,构建了以提示驱动为核心的分割机制,使模型在无需针对具体任务重新训练的条件下,仍能对未知类别和复杂场景中的目标进行分割(Kirillov *et al.*, 2023)。在此基础上,Ravi *et al.* (2025) 进一步将分割能力扩展至连续图像序列与视频数据,并引入文本提示等

高层语义信息，使分割过程能够与语义概念建立联系，体现出由纯视觉模型向多模态模型的进化。然而，由于 SAM 模型主要基于自然图像与通用语义进行训练，其对矿物光学特性、颗粒渐变边界及成岩改造结构等沉积学特征缺乏系统学习，在薄片场景中仍易出现过分割和语义不稳定等问题（Zheng *et al.*, 2024a）。因此，在实际使用 SAM 模型的过程中，需要引入地质知识的约束。例如，在对岩石 CT 照片以及钻屑照片进行颗粒分割时，由于颗粒与背景的边界并不显著，直接调用 SAM 无法准确分割所有颗粒，因此需要引入边界特征识别模块来提升颗粒分割的准确性（Wang *et al.*, 2025）。除沉积岩薄片图像存在颗粒边界模糊的核心问题外，单偏光与正交偏光图像的显著光性差异，也对模型精准捕捉矿物特征构成挑战。

此外，大语言模型（Large Language Models, LLMs），如 DeepSeek、GeoGPT 等（Zhang *et al.*, 2024），有望为地质知识的系统化组织与高效调用提供核心技术支撑。这类模型基于 Transformer 架构，经过大规模通用语料预训练后，在语义理解、信息整合与知识迁移方面展现出优异性能（Thulke *et al.*, 2024; Zhang *et al.*, 2024）。在地质等专业领域，通过引入领域专属语料、构建任务导向的指令数据集与专业术语体系，并结合参数高效微调策略，可显著增强模型对专业概念、知识结构及领域推理逻辑的适配能力。在地质学领域，ClimateGPT 通过适配气候领域文献，有效提升了对专业知识的理解能力并降低了事实性错误（Thulke *et al.*, 2024），为 LLMs 在地质领域的应用提供了有益借鉴。视觉与语言大模型的融合，将进一步打破图像与文本之间的语义壁垒，实现二者的统一表征。以 CLIP 为代表的模型，通过在大规模图像—文本配对数据上开展对比学习，使视觉特征与语言语义在统一特征空间中完成对齐（Radford *et al.*, 2021），进而赋予模型强大的跨任务泛化能力。在沉积学应用场景中，已有研究尝试将 CLIP 的视觉—语言对齐机制引入岩石薄片描述任务，通过将薄片图像特征与语言生成空间进行映射对齐，实现了从图像到专业文本的自动生成（Luo *et al.*, 2025）。此类方法无需依赖像素级精细标注，即可将薄片图像中的视觉地质证据，转化为符合地质工作规范的结构化文本信息，为岩石薄片资料的高效整理、智能检索与人机协同分析提供了全新技术路径。

总体而言，当前岩石矿物智能识别正逐步从单一模型的直接应用，迈向以模型适配、弱监督/半监督学习、生成式模型辅助、大模型与多模态融合为核心的新一代智能分析范式，各类技术路径的协同发展，将推动岩石薄片智能识别技术向更精准、高效、实用的方向迈进。

3 基于大模型的沉积岩薄片分割实例验证

尽管 SAM 在通用图像分割领域展现出优异的零样本泛化能力，但其训练数据主要来源于自然场景图像，缺乏对矿物光学特征、颗粒渐变边界以及成岩改造结构等沉积学特征

的系统学习。因此，在岩石薄片场景中容易出现颗粒边界识别不准确、过分割以及类别语义不稳定等问题。

以双偏光沉积岩薄片颗粒分割场景为例，Wang *et al.* (2026) 基于 PPL-XPL 配对图像开展 SAM 适配验证，检验通用分割大模型在沉积岩薄片场景中的适用性。与普通自然图像不同，岩石薄片图像中同一颗粒在单偏光与正交偏光条件下具有互补但差异显著的光学响应：PPL 图像更有利于保留颗粒轮廓和低频结构信息，而 XPL 图像则更能反映干涉色、消光现象和矿物内部纹理。针对这一特点，适配改进的 SAM 框架并未简单沿用通用模型的输入与提示机制，而是通过双偏光特征融合和自动提示生成策略，综合利用 PPL 与 XPL 图像中的互补信息，并将通用分割能力转化为更适合沉积颗粒识别的专业分析工具（图 4）。

该实例表明，视觉大模型在沉积岩薄片分析中的应用重点，不是简单迁移已有通用模型，而是将大模型的通用表征能力与岩石薄片的专业成像特征相结合。通过引入偏光互补信息、颗粒纹理先验和自动提示机制，适配改进的 SAM 框架可为减少人工交互、提高颗粒分割效率和推动岩相图像智能分析提供有益示范。更广泛地看，这类研究也提示未来沉积学大模型的发展需要从“模型调用”转向“领域适配”，即在保留基础模型泛化能力的同时，进一步融合岩石学知识、显微成像机理和地质解释约束，从而提高模型输出结果的专业可靠性和实际可用性。

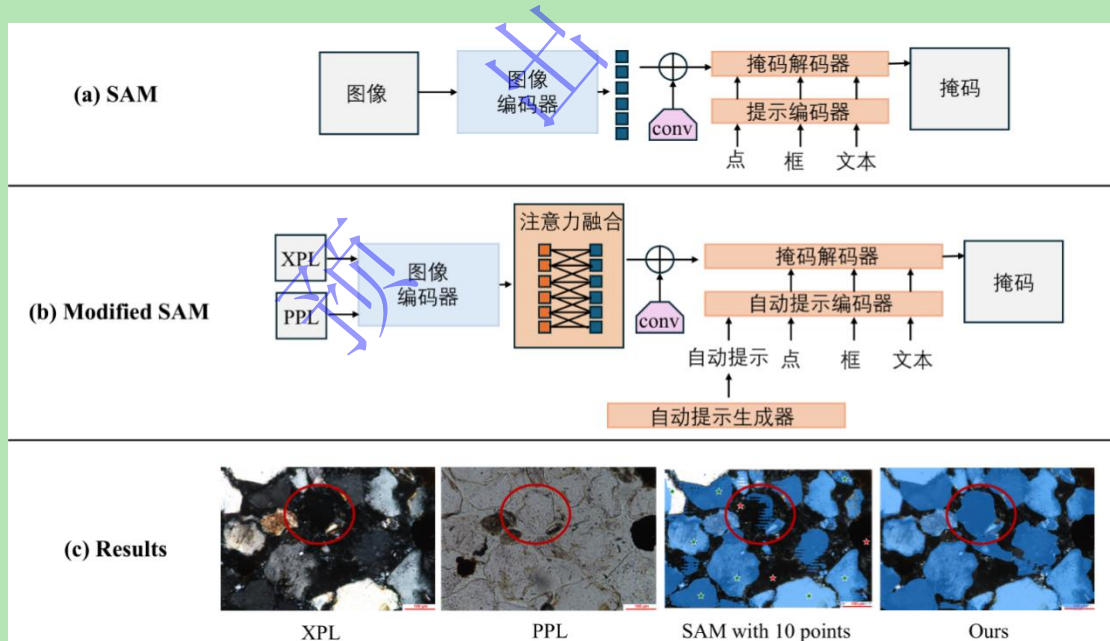


图 4 适配改进 SAM 模型结构及其在河流砂薄片颗粒分割中的应用示意图（据 Wang *et al.*, 2026 修改）

(a) 传统 SAM 模型结构；(b) 面向双偏光薄片图像适配的改进 SAM 模型结构；(c) 基于适配改进 SAM 的颗粒分割示意

Fig. 4 Schematic illustration of the adapted Segment Anything Model (SAM) framework and its application to grain segmentation in fluvial sandstone thin-section images (modified from Wang *et al.*, 2026)

(a) Original SAM architecture; (b) Adapted SAM architecture for paired plane-polarized light (PPL) and cross-polarized light (XPL) thin-section images; (c) Example of grain segmentation using the adapted SAM framework

4 总结与展望

在过去数年中，沉积学显微岩相照片的智能识别技术取得了跨越式发展与显著提升。早期研究多以直接调用通用模型为主，随着研究的深入，科研人员逐步针对岩石、矿物的光学特征、边界信息及结构组分等核心地质特征，对模型模块进行针对性优化与改进，有效提升了智能模型在沉积学领域的识别精度与适配性，推动了技术从“通用”向“专用”的转型。

尽管当前技术已取得长足进步，但岩相照片（尤其是高质量标注照片）的稀缺性以及岩石矿物的复杂多样，仍是制约该领域进一步发展的核心瓶颈。基于此，未来研究需重点探索弱监督、半监督乃至无监督学习方法，逐步降低传统监督学习对精细人工标注的高度依赖，突破数据稀缺带来的技术局限。与此同时，合理运用生成式模型生成高质量显微岩相照片，有望有效扩充岩相图像数据集规模，满足深度学习模型对海量训练数据的需求，为模型精度的进一步提升提供数据支撑。此外，当前在各领域广泛应用的大模型与多模态模型，亦是未来沉积学显微岩相智能识别的重点研究方向。科研工作者可依托现有技术基础，“站在巨人的肩膀上”，利用 SAM 等经过海量数据预训练的图像大模型，针对沉积岩显微图像的独特特征进行针对性微调，以提升模型在复杂薄片场景中的识别精度、泛化能力与地质解释可靠性。同时，通过融合显微图像信息与地质文本信息，引入岩石学、沉积学和成岩演化等多源地质先验知识，进一步提升模型的地质适配性与识别可靠性。具体而言，可将矿物组合特征、颗粒粒度与分选特征、颗粒接触关系、孔隙类型及成岩作用序列等专业知识编码为知识图谱、规则约束或文本提示，与视觉特征联合建模；同时结合地层层位、沉积环境、岩相类型及区域地质背景等宏观地质信息，对模型预测结果进行约束与校验。通过构建“图像特征—地质知识—分类决策”协同学习框架，实现从单纯视觉识别向地质机理驱动识别的转变，为未来构建具有跨盆地、跨层系泛化能力的沉积岩显微图像智能识别通用大模型奠定坚实基础。

回望历史，19 世纪 H.C. Sorby 等地质先贤首次通过显微镜观察到，外观朴素无奇的岩石矿物竟能呈现出绚丽多彩的干涉色，这份发现的惊喜，为显微岩相学的诞生埋下了种子。此后，F.J. Pettijohn、R.L. Folk、R.J. Dunham、W.R. Dickinson 等学者进一步推广显微岩相学研究，逐步建立并完善了现代沉积学的研究范式，为后续相关研究提供了坚实的理论与方法支撑。随着深度学习等人工智能技术的飞速迭代与深度渗透，我们有理由相信，在不远的未来，通用大模型将实现显微岩相照片的自动化精准识别，并进一步延伸至沉积环境判识、物源分析等核心沉积学研究领域。当这一目标达成之时，必将成为沉积学发展史上具有里程碑意义的重要时刻，为该领域的研究注入全新活力，推动沉积学研究进入智能化、高效化的全新阶段。

致谢 本论文的撰写契机源于 2025 年 11 月由中国地质大学杜远生教授、余文超教授等组织的第五届“中国大地构造沉积学与深时气候沉积学”学术会议。在论文写作过程中,得到多位前辈专家的宝贵建议与悉心指导。

参考文献 (References)

- 蔡宇恒, 滕奇志, 涂秉宇. 2020. 基于深度学习的岩石铸体薄片图像孔隙自动提取[J]. 科学技术与工程, 20(28): 11685-11692. [Cai Yuheng, Teng Qizhi, Tu Bingyu. 2020. Automatic extraction of pores in thin slice images of rock castings based on deep learning[J]. Science Technology and Engineering, 20(28): 11685-11692.]
- 程国建, 杨静, 黄全舟, 等. 2013. 基于概率神经网络的岩石薄片图像分类识别研究[J]. 科学技术与工程, 13(31): 9231-9235. [Cheng Guojian, Yang Jing, Huang Quanzhou, et al. 2013. Rock image classification recognition based on probabilistic neural networks[J]. Science Technology and Engineering, 13(31): 9231-9235.]
- 邓乃尔, 徐浩, 周文, 等. 2025. 基于深度学习的岩石矿物智能识别研究进展与发展趋势[J]. 成都理工大学学报(自然科学版), 52(1): 64-93. [Deng Naier, Xu Hao, Zhou Wen, et al. 2025. Research progress and development trends of intelligent recognition of rock minerals based on deep learning[J]. Journal of Chengdu University of Technology (Science & Technology Edition), 52(1): 64-93.]
- 杜崇娇, 余晓露, 彭金宁, 等. 2023. 基于深度学习的碳酸盐岩颗粒分类识别[J]. 成都理工大学学报(自然科学版), 50(5): 629-636. [Du Chongjiao, Yu Xiaolu, Peng Jinning, et al. 2023. Research on classification and recognition of carbonate particles based on deep learning[J]. Journal of Chengdu University of Technology (Science & Technology Edition), 50(5): 629-636.]
- 郝慧珍, 顾庆, 胡修棉. 2021. 基于机器学习的矿物智能识别方法研究进展与展望[J]. 地球科学, 46(9): 3091-3106. [Hao Huizhen, Gu Qing, Hu Xiumian. 2021. Research advances and prospective in mineral intelligent identification based on machine learning[J]. Earth Science, 46(9): 3091-3106.]
- 胡修棉, 赖文, 许艺炜, 等. 2020. 沉积岩显微数字图像数据的获取与信息收集标准[J]. 中国科学数据, 5(3): 10-20. [Hu Xiumian, Lai Wen, Xu Yiwei, et al. 2020. Standards for digital micrograph of the sedimentary rocks[J]. China Scientific Data, 5(3): 10-20.]
- 刘春, 许强, 施斌, 等. 2018. 岩石颗粒与孔隙系统数字图像识别方法及应用[J]. 岩土工程学报, 40(5): 925-931. [Liu Chun, Xu Qiang, Shi Bin, et al. 2018. Digital image recognition method of rock particle and pore system and its application[J]. Chinese Journal of Geotechnical Engineering, 40(5): 925-931.]
- 刘恒, 王浩铮, 冯林峰, 等. 2025. 基于改进 YOLOv5s 实现岩石薄片中心副矿物实时视觉识别与定位[J]. 成都理工大学学报(自然科学版), 52(5): 888-899. [Liu Heng, Wang Haozheng, Feng Linfeng, et al. 2025. Real-time visual recognition and accessory mineral localization in rock thin sections based on improved YOLOv5s[J]. Journal of Chengdu University of Technology (Science & Technology Edition), 52(5): 888-899.]
- 刘烨, 吕锦涛. 2023. 基于超像素与半监督的岩石图像分割与识别[J]. 工程科学与技术, 55(2): 171-183. [Liu Ye, Lü Jintao. 2023. Semi-supervised rock image segmentation and recognition based on superpixel[J]. Advanced Engineering Sciences, 55(2): 171-183.]
- 王志忠, 王永增, 马连成, 等. 2017. 基于分水岭和形态学重构的矿石图像分割方法[J]. 物联网技术, 7(3): 89-91. [Wang Zhizhong, Wang Yongzeng, Ma Liancheng, et al. 2017. A mineral image segmentation method based on watershed and morphological reconstruction[J]. Internet of Things Technologies, 7(3): 89-91.]
- 吴博, 李永胜, 王睿, 等. 2024. 基于自训练的多标签岩石薄片分类方法[J]. 黄金, 45(2): 61-67. [Wu Bo, Li Yongsheng, Wang Rui, et al. 2024. Multi-label rock ore slice classification approach based on self-training[J]. Gold, 45(2): 61-67.]
- 徐圣嘉, 苏程, 朱孔阳, 等. 2022. 基于深度学习的岩石薄片矿物自动识别方法[J]. 浙江大学学报(理学版), 49(6): 743-752. [Xu Shengjia, Su Cheng, Zhu Kongyang, et al. 2022. Automatic identification of mineral in petrographic thin sections based on images using a deep learning method[J]. Journal of Zhejiang University (Science Edition), 49(6): 743-752.]
- 徐述腾, 周永章. 2018. 基于深度学习的镜下矿石矿物的智能识别实验研究[J]. 岩石学报, 34(11): 3244-3252. [Xu Shuteng, Zhou Yongzhang. 2018. Artificial intelligence identification of ore minerals under microscope based on deep learning algorithm[J]. Acta Petrologica Sinica, 34(11): 3244-3252.]
- 余晓露, 李龙龙, 蒋宏, 等. 2023. 融合图像处理与深度学习的亮晶颗粒灰岩岩相学分析应用[J]. 石油实验地质, 45(5): 1026-1038. [Yu Xiaolu, Li Longlong, Jiang Hong, et al. 2023. Application of sparry grain limestone petrographic analysis combining image processing and deep learning[J]. Petroleum Geology & Experiment, 45(5): 1026-1038.]
- 张杰, 沈安江, 胡安平, 等. 2023. 基于深度学习的碳酸盐岩薄片人工智能鉴定方法研究[J]. 海相油气地质, 28(4): 337-348. [Zhang Jie, Shen Anjiang, Hu Anping, et al. 2023. Research on artificial intelligence identification approach for carbonate thin sections based on deep learning[J]. Marine Origin Petroleum Geology, 28(4): 337-348.]
- 张利军, 鲁文豪, 张建东, 等. 2024. 基于深度学习的镜下岩石、矿物薄片识别[J]. 地学前缘, 31(3): 498-510. [Zhang Lijun, Lu Wenhao, Zhang Jiandong, et al. 2024. Rock and mineral thin section identification based on deep learning[J]. Earth Science Frontiers, 31(3): 498-510.]
- 张野, 李明超, 韩帅. 2018. 基于岩石图像深度学习的岩性自动识别与分类方法[J]. 岩石学报, 34(2): 333-342. [Zhang Ye, Li Mingchao, Han Shuai. 2018. Automatic identification and classification in lithology based on deep learning in rock images[J]. Acta Petrologica Sinica, 34(2): 333-342.]
- 周嵘, 吴朝东, 张亚楠. 2023. 基于生成对抗网络的砂岩薄片图像视野外重建[J]. 北京大学学报(自然科学版), 59(2): 231-241. [Zhou Rong, Wu Chaodong, Zhang Ya'nan. 2023. Outpainting reconstruction of sandstone thin-section image based on generative adversarial network[J]. Acta Scientiarum Naturalium Universitatis Pekinensis, 59(2): 231-241.]
- Asmussen P, Conrad O, Günther A, et al. 2015. Semi-automatic segmentation of petrographic thin section images using a "seeded-region growing algorithm" with an application to characterize weathered subarkose sandstone[J]. Computers & Geosciences, 83: 89-99.
- Barbosa R T C M, Faria E L, Klatt M, et al. 2024. Unsupervised segmentation for sandstone thin section image analysis[J]. Computational Geosciences, 28(6): 1049-1057.
- Barraud J. 2006. The use of watershed segmentation and GIS software for textural analysis of thin sections[J]. Journal of Volcanology and Geothermal Research, 154(1/2): 17-33.
- Bihani A, Daigle H, Santos J E, et al. 2022. MudrockNet: Semantic segmentation of mudrock SEM images through deep learning[J]. Computers & Geosciences, 158: 104952.
- Bloch S, Lander R H, Bonnell L. 2002. Anomalously high porosity and permeability in deeply buried sandstone reservoirs: Origin and predictability[J]. AAPG Bulletin, 86(2): 301-328.
- Cao H, Wang Y Y, Chen J, et al. 2022. Swin-Unet: Unet-like pure transformer for medical image segmentation[C]//Proceedings of computer vision – ECCV 2022 workshops. Tel Aviv: Springer, 205-218.

- Chatterjee S. 2013. Vision-based rock-type classification of limestone using multi-class support vector machine[J]. *Applied Intelligence*, 39(1): 14-27.
- Di Martino A, Carlini G, Amorosi A. 2026. Explainable artificial intelligence for sedimentary facies segmentation[J]. *Scientific Reports*, 16(1): 5984.
- Dickinson W R, Suczek C A. 1979. Plate tectonics and sandstone compositions[J]. *AAPG Bulletin*, 63(12): 2164-2182.
- Dos Anjos C E M, Avila M R V, Vasconcelos A G P, et al. 2021. Deep learning for lithological classification of carbonate rock micro-CT images[J]. *Computational Geosciences*, 25(3): 971-983.
- Dosovitskiy A, Beyer L, Kolesnikov A, et al. 2021. An image is worth 16×16 words: Transformers for image recognition at scale[C]//*Proceedings of 9th international conference on learning representations (2021)* (Vienna, Austria). OpenReview.net.
- Dunham R J. 1962. Classification of Carbonate Rocks According to Depositional Texture. In *Classification of Carbonate Rocks*; Ham, W.E., Ed.; AAPG: Tulsa, OK, USA, 1962; pp. 108–121. Fan J S, Yu X L, Di Y C, et al. 2025. A foundation model for rock thin-section images analysis[J]. *Communications Engineering*, 5(1): 9.
- Ferreira I, Ochoa L, Koeshidayatullah A. 2022. On the generation of realistic synthetic petrographic datasets using a style-based GAN[J]. *Scientific Reports*, 12(1): 12845.
- Folk R L. 1966. A review of grain-size parameters[J]. *Sedimentology*, 6(2): 73-93.
- Francus P. 1998. An image-analysis technique to measure grain-size variation in thin sections of soft clastic sediments[J]. *Sedimentary Geology*, 121(3/4): 289-298.
- Garzanti E, Doglioni C, Vezzoli G, et al. 2007. Orogenic belts and orogenic sediment provenance[J]. *The Journal of Geology*, 115(3): 315-334.
- Ge S P, Wang C, Jiang Z W, et al. 2021. Dual-input attention network for automatic identification of detritus from river sands[J]. *Computers & Geosciences*, 151: 104735.
- Goodchild J S, Fueten F. 1998. Edge detection in petrographic images using the rotating polarizer stage[J]. *Computers & Geosciences*, 24(8): 745-751.
- Gorsevski P V, Onasch C M, Farver J R, et al. 2012. Detecting grain boundaries in deformed rocks using a cellular automata approach[J]. *Computers & Geosciences*, 42: 136-142.
- Han Y B, Liu Y. 2024. Intelligent classification and segmentation of sandstone thin section image using a semi-supervised framework and GL-SLIC[J]. *Minerals*, 14(8): 799.
- Hou L, Ma C, Tang W Q, et al. 2024. DDViT: Advancing lithology identification on FMI image logs through a dual modal transformer model with less information drop[J]. *Geoenergy Science and Engineering*, 234: 212662.
- Joseph P, Teles V, Weill P. 2016. Modelling approaches in sedimentology: Introduction to the thematic issue[J]. *Comptes Rendus Géoscience*, 348(7): 473-478.
- Karras T, Laine S, Aittala M, et al. 2020. Analyzing and improving the image quality of StyleGAN[C]//*Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. Seattle: IEEE: 8107-8116.
- Kirillov A, Mintun E, Ravi N, et al. 2023. Segment anything. In: *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*[C]. Piscataway, NJ: IEEE Press: 3992-4003.
- Koeshidayatullah A, Morsilli M, Lehrmann D J, et al. 2020. Fully automated carbonate petrography using deep convolutional neural networks[J]. *Marine and Petroleum Geology*, 122: 104687.
- Koeshidayatullah A, Al-Azani S, Baraboshkin E E, et al. 2022. FaciesViT: Vision transformer for an improved core lithofacies prediction[J]. *Frontiers in Earth Science*, 10: 992442.
- Launeau P, Cruden A R, Bouchez J L. 1994. Mineral recognition in digital images of rocks; a new approach using multichannel classification[J]. *The Canadian Mineralogist*, 32(4): 919-933.
- Li K R, Gao Y J, Ma Y J, et al. 2026a. Weakly supervised semantic segmentation of microscopic carbonates on marginal devices[J]. *Computers & Geosciences*, 207: 106059.
- Li K R, Song J M, Zhang Z Y, et al. 2026b. A dataset and benchmark of carbonate thin-section images for deep learning[J]. *Scientific Data*, 13(1): 340.
- Lin S, Hu Q K, Guo H W, et al. 2025. An efficient and automated classification system for rocks based on visually explainable deep learning[J]. *Lithosphere*, 2025(2): lithosphere_2024_239.
- Liu X K, Song H J. 2020. Automatic identification of fossils and abiotic grains during carbonate microfacies analysis using deep convolutional neural networks[J]. *Sedimentary Geology*, 410: 105790.
- Luo X, Sun J M, Chi P, et al. 2025. ThinGPT: Describing sedimentary rock thin section images with a multimodal large language model[J]. *Petroleum Science*, 22(12): 5020-5033.
- Maitre J, Bouchard K, Bédard L P. 2019. Mineral grains recognition using computer vision and machine learning[J]. *Computers & Geosciences*, 130: 84-93.
- Marmo R, Amodio S, Tagliaferri R, et al. 2005. Textural identification of carbonate rocks by image processing and neural network: Methodology proposal and examples[J]. *Computers & Geosciences*, 31(5): 649-659.
- Młynarczyk M, Górszczyk A, Ślipek B. 2013. The application of pattern recognition in the automatic classification of microscopic rock images[J]. *Computers & Geosciences*, 60: 126-133.
- Radford A, Kim J W, Hallacy C, et al. 2021. Learning transferable visual models from natural language supervision[C]//*Proceedings of the 38th international conference on machine learning*. PMLR, (eds Meila, M. & Zhang, T.) 139, 8748-8763.
- Ransinangue A, Labourdette R, Houzay E, et al. 2025. SynSection: Sedimentology-driven data generation for deep learning applications in carbonates petrography[J]. *Marine and Petroleum Geology*, 182: 107490.
- Ravi N, Gabeur V, Hu Y T, et al. 2025. SAM 2: Segment anything in images and videos[C]//*Proceedings of the thirteenth international conference on learning representations*. Singapore: OpenReview.net.
- Reichstein M, Camps-Valls G, Stevens B, et al. 2019. Deep learning and process understanding for data-driven Earth system science[J]. *Nature*, 566(7743): 195-204.
- Saxena N, Day-Stirrat R J, Hows A, et al. 2021. Application of deep learning for semantic segmentation of sandstone thin sections[J]. *Computers & Geosciences*, 152: 104778.
- Selvaraju R R, Cogswell M, Das A, et al. 2017. Grad-CAM: Visual explanations from deep networks via gradient-based localization[C]//*2017 IEEE international conference on computer vision*. Venice: IEEE, 618-626.
- Selvaraju R R, Cogswell M, Das A, et al. 2020. Grad-CAM: Visual explanations from deep networks via gradient-based localization[J]. *International Journal of Computer Vision*, 128(2): 336-359.
- Su C, Xu S J, Zhu K Y, et al. 2020. Rock classification in petrographic thin section images based on concatenated convolutional neural networks[J]. *Earth Science Informatics*, 13(4): 1477-1484.
- Tang M, Liu Y M, Durlafsky L J. 2020. A deep-learning-based surrogate model for data assimilation in dynamic subsurface flow problems[J]. *Journal of Computational Physics*, 413: 109456.
- Terribile F, FitzPatrick E A. 1992. The application of multilayer digital image processing techniques to the description of soil thin sections[J]. *Geoderma*, 55(1/2): 159-174.
- Thulke D, Gao Y B, Pelsers P, et al. (2024-01-17). ClimateGPT: Towards AI synthesizing interdisciplinary research on climate change[EB/OL]. <https://arxiv.org/abs/2401.09646>.

- Wang Y J, Zheng D Y, Hou M C, et al. 2026. An auto-prompting Segment Anything Model for dual-modal grain segmentation in rock images[J]. *Applied Computing and Geosciences*, 31: 100375.
- Wang Z Q, Hou Z Y, Cao D P. 2025. Enhancing SAM-based digital rock image segmentation via edge-semantics fusion[J]. *Applied Computing and Geosciences*, 28: 100292.
- Weltje G J. 2002. Quantitative analysis of detrital modes: Statistically rigorous confidence regions in ternary diagrams and their use in sedimentary petrology[J]. *Earth-Science Reviews*, 57(3/4): 211-253.
- Zeng W, Wang Y R, Chen C H, et al. 2025. Improved bimodal segmentation of multi-light images based on feature fusion[J]. *Pattern Recognition*, DOI: 10.1016/j.patcog.2025.111947.
- Zhang Y F, Wei C, He Z T, et al. 2024. GeoGPT: An assistant for understanding and processing geospatial tasks[J]. *International Journal of Applied Earth Observation and Geoinformation*, 131: 103976.
- Zheng D Y, Wu S X, Ma C, et al. 2022. Zircon classification from cathodoluminescence images using deep learning[J]. *Geoscience Frontiers*, 13(6): 101436.
- Zheng D Y, Hou L, Hu X M, et al. 2024a. Sediment grain segmentation in thin-section images using dual-modal Vision Transformer[J]. *Computers & Geosciences*, 191: 105664.
- Zheng D Y, Zhong H T, Camps-Valls G, et al. 2024b. Explainable deep learning for automatic rock classification[J]. *Computers & Geosciences*, 184: 105511.

From Convolutional Neural Networks to Multi-Modal Large Models for Intelligent Analysis of Rock Thin Sections

ZHENG DongYu^{1,2,3,4}, WANG YiJun³, BAI YiFan⁴, WU SiXuan⁵, HU XiuMian⁶, HOU MingCai^{1,2}

1. Institute of Sedimentary Geology, Chengdu University of Technology, Chengdu 610059, China

2. Key Laboratory of Deep-Time Geography and Environment Reconstruction and Applications of the Ministry of Natural Resources, Chengdu University of Technology, Chengdu 610059, China

3. College of Computer Science and Cyber Security (Pilot Software College), Chengdu University of Technology, Chengdu 610059, China

4. College of Geophysics, Chengdu University of Technology, Chengdu 610059, China

5. Exploration and Development Research Institute, PetroChina Southwest Oil & Gasfield Company, Chengdu 610041, China

6. School of Earth Sciences and Engineering, Nanjing University, Nanjing 210023, China

Abstract: Accurate identification of mineral composition and microstructural features in sedimentary rock thin sections is an important basis for provenance analysis, reconstruction of depositional environments, and reservoir evaluation. Although traditional manual identification based on polarized light microscopy is highly reliable, it is limited by low efficiency, strong subjectivity, and insufficient standardization. Therefore, it is difficult to meet the growing demand for high-throughput, quantitative, and reproducible analysis in data-driven sedimentological research. Focusing on the development history and technical evolution of intelligent mineral identification in sedimentary rocks, recent progress in traditional image processing, machine learning, and deep learning methods for rock thin-section image analysis is systematically summarized. The applications of convolutional neural networks and Vision Transformers in image classification, object detection, and image segmentation are reviewed, with particular emphasis on their adaptability to the recognition of sedimentological features. The roles of weakly supervised, semi-supervised, and unsupervised learning in reducing dependence on manual annotation are further discussed. The potential of generative models in data augmentation, class balancing, and rare-sample supplementation is also analyzed. In addition, the development trends of vision foundation models, large language models, and multimodal frameworks are examined in terms of geological knowledge integration, model generalization, and interpretive consistency. Intelligent mineral identification is shifting from the application of single models toward an integrated analytical paradigm that combines low-label learning, generative models, and multimodal large models. Future research should strengthen the construction of standardized rock thin-section datasets, the incorporation of geological prior knowledge, cross-domain model adaptation, and interpretability evaluation. A sedimentary rock intelligent analysis system with high generalization ability and geological

interpretive consistency is expected to provide efficient and reliable technical support for sedimentological research.

Key words: sedimentary rock intelligent analysis; rock thin sections; convolutional neural network; foundation model; artificial intelligence; deep learning

