www.csdata.org

ISSN 2096-2223
CN 11-6035/N

文献 DOI：
10.11922/csdata.2020.0062.zh
数据 DOI：
10.11922/sciencedb.j00001.00087
文献分类：地球科学

收稿日期：2020-07-11
开放同评：2020-08-17
录用日期：2020-09-21
发表日期：2020-09-28

* 论文通信作者
侯明才：houmc@cdut.edu.cn

专题 岩石显微图像

鄂尔多斯盆地东北缘中侏罗世碎屑岩显微图像数据集

晁晖[1,2]，侯明才[1,2*]，曹海洋[1,2]，梁子若[1,2]

1. 成都理工大学沉积地质研究院，成都 610059
2. 成都理工大学油气藏地质及开发工程国家重点实验室，成都 610059

摘要： 薄片显微图像是地质研究工作中不可缺少的基础数据，而大数据正在成为科学发现的新引擎，显微图像数据集则是地质大数据的一种有力的表现形式。基于此，对鄂尔多斯盆地东北缘中侏罗世地层进行样品采集、薄片磨制、镜下鉴定及拍照，整理出该区域的显微图像数据集。包括 78 个碎屑岩的 208 张单偏光和正交偏光照片，放大倍数多为 4 倍和 10 倍。岩性主要为岩屑砂岩，少量含砾，以及部分细粒岩，磨圆、分选一般，基底胶结，点–线接触是主要接触方式。石英多为单晶石英，长石以微斜长石和斜长石为主，岩屑以粘土质和硅质岩屑居多，变质岩屑其次，能够反映出研究区地层的基本岩相特征和矿物特征，对于本地区中侏罗世研究提供一定的沉积学参考价值，更有利于查找和揭示隐藏在地质大数据中的多种地质要素关联关系，并在此基础上更好地探究成因关系。

关键词： 地质数据；图像数据；信息化；矿物学；沉积学

数据库（集）基本信息简介

数据库（集）名称	鄂尔多斯盆地东北缘中侏罗世碎屑岩显微图像数据集
数据作者	晁晖，侯明才，曹海洋，梁子若
数据通信作者	侯明才（houmc@cdut.edu.cn）
数据时间范围	岩石样品采集地层的时代为中侏罗世（延安组–直罗组，约174–163 Ma）;岩石样品采集时间为2017年；岩石薄片偏光显微照片拍摄于2018年。
地理区域	鄂尔多斯盆地东北缘神山沟剖面（起点坐标为39°45′89″N，110°11′03″E）和高头窑剖面（起点坐标为40°2′6.8″N，109°38′3.7″E）。
偏光显微镜分辨率	1280*960像素
数据量	5.01 GB
数据格式	*.tif，*.xls，*.jpg
数据服务系统网址	https://dx.doi.org/10.11922/sciencedb.j00001.00087
基金项目	国家重点基础研究发展“973”计划（2015CB453001）
数据集组成	数据集共包括3个数据文件，它们分别为：（1）薄片照片.zip是岩石薄片偏光显微照片（tif格式）的图片集，数据量4.91GB；（2）剖面野外宏观照片.zip展示了各实测剖面的野外宏观照片（jpg格式），数据量107 MB；（3）薄片照片信息表.xls是岩石薄片的基本信息以及鉴定数据表，数据量33KB。

引　言

随着互联网技术的发展，现代社会逐渐趋向于数字化和网络化，人们的信息需求也随之从耗时繁琐的文字转变为更为直观省时的图像，因而，获取海量图像信息成为了重中之重[1]。目前，大数据正在被引入地质学，其使地质学从观察学科转变为数据科学[2]。最早的地质依靠野外踏勘，显微镜的出现使地质科学首次从宏观进入微观领域。岩石薄片镜下观察能够全面真实地反映出岩石微观结构构造、矿物成分等特征，是开展地质研究工作最基础的工具和方法之一。因此对岩石薄片显微图像进行建库管理和分类检索显得尤为重要。鄂尔多斯盆地位于华北地台，是中国中西部中生代能源矿产最丰富、保存最完整的内陆含油气盆地[3]。在中侏罗世，盆地东北部位于阴山构造带和太行山构造带的交汇处，是研究蒙古–鄂霍茨克洋盆关闭、古太平洋板块俯冲及其所引发的陆内造山活动的理想之地[4]。同时，盆地东北部为主要的煤富集区及铀矿富集区。据上，分析了解鄂尔多斯盆地东北缘中侏罗世地层的岩石学特征，对于明确该地区的碎屑物质来源、沉积构造演化以及探讨成煤成矿规律均具有十分重要的指导意义。

本文研究是基于鄂尔多斯盆地东北缘中侏罗世碎屑岩薄片图像，以薄片内碎屑岩微观沉积学特征为研究目标，构建对应显微图像数据集，为进一步在相关地区进行地质研究提供图像信息基础，提高地质科学研究对象的空间性和非空间性，推动多学科融合的科学新范式。

1　数据采集和处理方法

数据采集和处理分为野外数据和显微数据两部分。

1.1　野外数据

78 个岩石样品来自鄂尔多斯盆地东北缘高头窑剖面和神山沟剖面。高头窑剖面样品 34 个，神山沟剖面样品 44 个。采样时均去除表面风化物质，取新鲜面样品，然后送至实验室磨制加工成薄片。

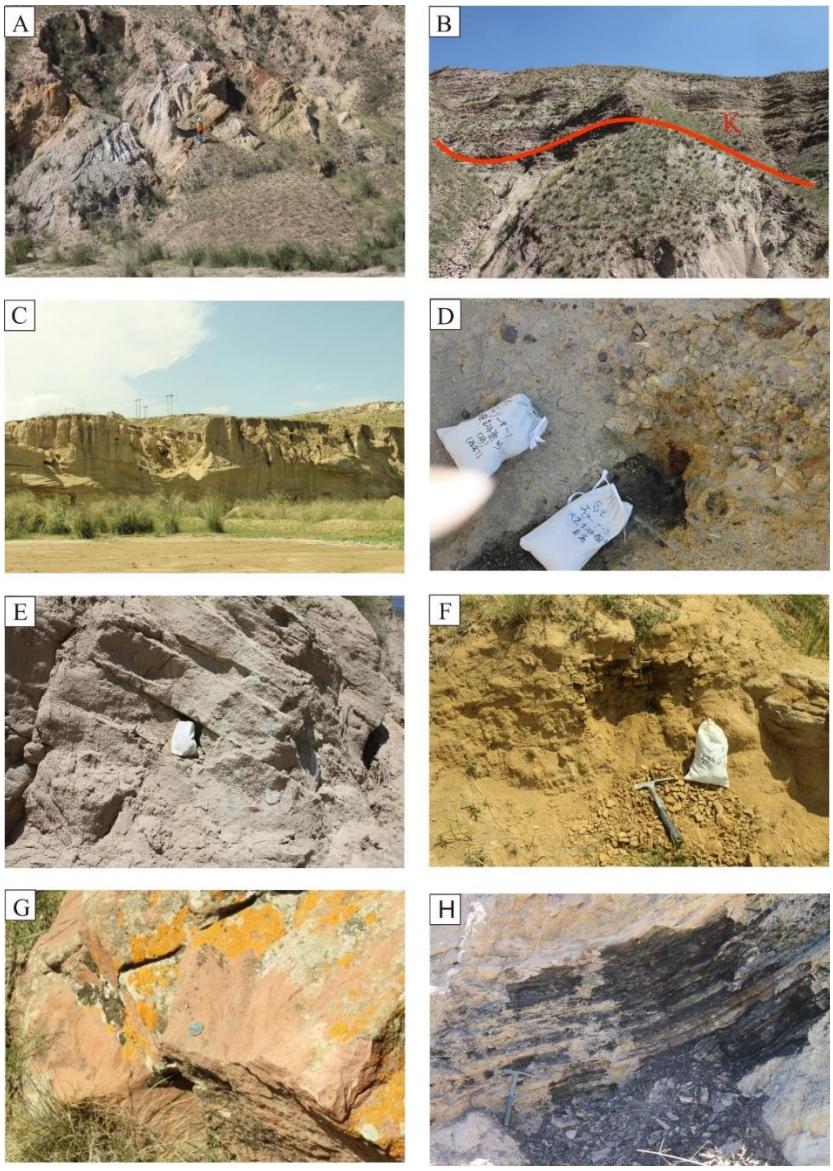
1.2　显微数据

薄片在偏光显微镜下观察，描述矿物的结晶特点，测定其光学性质，确定岩石的矿物成分，研究它的结构、构造，分析矿物的生成顺序，确定岩石类型及其成因特征，最后定出岩石的名称，并根据不同放大倍数和研究需求进行针对性拍照，进而得到显微图像。

2　数据样本描述

数据集包括野外剖面宏观照片（图 1），薄片照片（图 2）和薄片照片信息表。

野外宏观照片主要表现为高头窑剖面和神山沟剖面中侏罗世地层的宏观岩性、界限接触及沉积构造等内容，其中高头窑剖面 11 张，神山沟剖面 7 张，保存格式为 jpg。



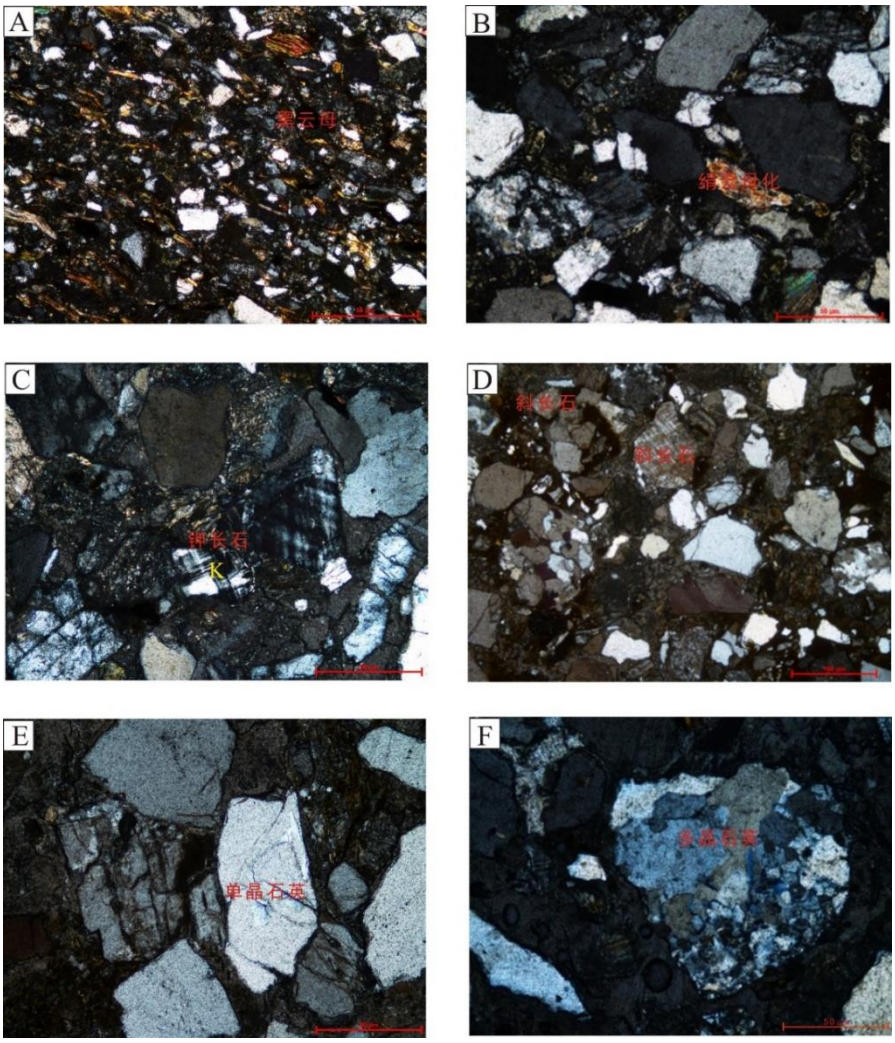
A：剖面宏观；B：白垩系与中侏罗世延安组不整合接触，白垩系紫红色含砾粗砂岩；C：姜黄色粗砂岩；D：杂色砾岩；E：灰白色粗砂岩；F：土黄色细砂岩夹杂泥岩发育；G：肉红色粗砂岩；H：深灰色含碳质泥页岩

图 1　鄂尔多斯盆地东北缘中侏罗世碎屑岩野外照片

薄片照片为来自鄂尔多斯盆地东北缘中侏罗世（延安组 J2y 和直罗组 J2z）地层 78 个碎屑岩样品的 516 张镜下照片。高头窑剖面样品 34 个，包括 27 个砂岩，1 个粉砂岩和 6 个泥岩样品，镜下照片共 222 张；神山沟剖面样品 44 个，包括 33 个砂岩，7 个粉砂岩和 4 个泥岩，镜下照片共 294 张（表 1）。显微图像的命名格式采用“样品名”+“镜下放大倍数”+“（正交/单偏光符号）”，保存格式为 tif。

表 1　高头窑剖面、神山沟剖面薄片基本信息

剖面	时代	坐标	显微照片数量	薄片数量
高头窑剖面	中侏罗世延安组（GtJ2y） 中侏罗世直罗组（GtJ2z）	40°2′6.8″N， 109°38′3.7″E	222	砂岩（27），粉砂岩（1），泥岩（6）
神山沟剖面	中侏罗世延安组（ShJ2y） 中侏罗世直罗组（ShJ2z）	39°45′89″N， 110°11′03″E	294	砂岩（33），粉砂岩（7），泥岩（4）



A：粉砂质泥岩，泥质中含黑云母，单偏光，直罗组；B：中细粒级长石石英砂岩，部分云母出现绢云母化，正交偏光，延安组；C：中粗岩屑砂岩，钾长石具有卡式双晶，正交偏光，延安组；D：中细粒岩屑石英砂岩，斜长石具有聚片双晶，单偏光，延安组；E：中粗石英砂岩，单晶石英颗粒线状接触，正交偏光，延安组；F：粗粒岩屑砂岩，多晶石英接触边界多呈缝合接触，正交偏光，直罗组。

图 2　鄂尔多斯盆地东北缘中侏罗世碎屑岩显微照片

薄片信息表内容主要为样品采集处的行政区划的地理位置或 GPS 坐标位置信息，岩石样品所属时代与层位，以及颗粒组分（颗粒类型、相对含量等）、填隙物特征（陆源杂基、灰泥、亮晶方解石等）、特殊的结构构造，成岩作用等，保存格式为 xls。

3　数据质量控制和评估

本文薄片归属样品均由论文编写者采自野外露头剖面，且与前人研究进行相关对比，保证地层资料真实可信；镜下观察及显微照片拍摄在成都理工大学显微实验室由相关专业学生完成，对于每个涉及视域，分别拍摄了对应的单偏光或正交偏光显微照片，照片质量清晰可靠。

地质现象错综复杂，微观世界千变万化，本图像数据集相较于文中研究区两条剖面，信息准确可信，对于鄂尔多斯盆地中侏罗世地层的综合研究能够在定量上具备一定的参考值。

4 数据价值

前人对于岩石显微图像数据开展过大量工作，多偏向于某种单一的应用，如岩石孔隙图像检索[5-6]、岩石矿物识别[7-9]、岩石粒度计算[10-11]等，以定量化研究居多。但是迄今为止，地质仍属于描述性科学范畴，多数地质作用和地质过程，例如构造作用、岩浆作用、沉积作用、变质作用、成矿作用、成煤作用、油气成藏作用等，都无法用单纯的确定性或随机性数学模型来确定[12]。特此，本文的数据集相较以上而言，更具定性化。如根据矿物颗粒接触关系去判断成岩作用，通过胶结物去讨论沉积环境，统计重矿物组分来分析沉积物源。

另一方面，地质学的基础理论是地质年代表，不同年代和时代的地质情况存在什么变化，不同时代的岩石属性也是不同的，如中侏罗世在鄂尔多斯盆地东北缘为碎屑岩陆相沉积，但同时期在西藏地区则为碳酸盐岩海相沉积，因此本图像数据集更具有非主体性，即任何看到本数据集的个人均可以自行处理数据，并能够结合多种地质数据，探寻之间隐藏的地质要素关系。

数据作者分工职责

晁晖（1991—），女，新疆哈密人，博士研究生，研究方向为古气候学。主要承担工作：样品采集、薄片鉴定、拍照、资料收集、数据集撰写。

侯明才（1968—），男，四川南充人，博士生导师，研究方向为沉积学。主要承担工作：样品采集、总体方案涉及、数据质量控制。

曹海洋（1988—），男，河南南阳人，博士，研究方向为构造地质学。主要承担工作：样品采集、薄片鉴定、资料整理。

梁子若（1995—），女，吉林松原人，硕士研究生，研究方向为沉积学。主要承担工作：薄片鉴定、拍照。

参考文献

[1] 张旗，周永章. 大数据正在引发地球科学领域一场深刻的革命——《地质科学》2017 年大数据专题代序[J]. 地质科学，2017, 52 (3): 637-648.

[2] 罗建民, 张旗. 大数据开创地学研究新途径:查明相关关系, 增强研究可行性[J]. 地学前缘, 2019, 26(04): 6-12.

[3] 谭聪. 鄂尔多斯盆地上二叠统—中上三叠统沉积特征及古气候演化[D]. 北京: 中国地质大学, 2017.

[4] 柳永清, 旷红伟, 彭楠, 等. 中生代陆相沉积古地理、古构造与脊椎动物古生态环境[C]// 2015 年全国沉积学大会沉积学与非常规资源论文摘要集, 2015.

[5] 岳清清. 深度学习在岩石薄片图像检索中的应用研究[D]. 西安石油大学, 2019.

[6] 王康. 基于 Hadoop/Spark 的岩石薄片图像拼接系统[D]. 西安石油大学, 2019.

[7] 杨丹丹. 矿岩颗粒图像分割方法研究综述[J]. 盐城工学院学报(自然科学版), 2014, 27(03):36-40, 44.

[8] 叶润青, 牛瑞卿, 张良培. 基于多尺度分割的岩石图像矿物特征提取及分析[J]. 吉林大学报(地

球科学版), 2011, 41(04): 1253-1261.
[9] 白林, 魏昕, 刘禹, 等. 基于 VGG 模型的岩石薄片图像识别[J]. 地质通报, 2019, 38(12): 2053-2058.
[10] 吴拥, 苏桂芬, 滕奇志, 等. 岩石薄片正交偏光图像的颗粒分割方法[J]. 科学技术与工程, 2013, 13(31): 9201-9206.
[11] 袁瑞, 朱锐, 瞿建华, 等. 一种基于岩石薄片图像的粒度分析新方法[J]. 岩性油气藏, 2015, 27(05): 104-107.
[12] 吴冲龙, 刘刚. 大数据与地质学的未来发展[J]. 地质通报, 2019, 38(07): 1081-1088.

论文引用格式

晁晖, 侯明才, 曹海洋, 等. 鄂尔多斯盆地东北缘中侏罗世碎屑岩显微图像数据集[J/OL]. 中国科学数据, 2020, 5(3). (2020-09-26). DOI: 10.11922/csdata.2020.0062.zh.

数据引用格式

晁晖, 侯明才, 曹海洋, 等. 鄂尔多斯盆地东北缘中侏罗世碎屑岩显微图像数据集[DB/OL]. Science Data Bank, 2020. (2020-08-17). DOI: 10.11922/sciencedb.j00001.00087.

A micrograph dataset of Middle Jurassic clastic rocks in northeastern Ordos Basin

Chao Hui[1,2], Hou Mingcai[1,2*], Cao Haiyang[1,2], Liang Ziruo[1,2]

1. Institute of Sedimentary Geology, Chengdu University of Technology, Chengdu 610059, P.R. China
2. State Key Laboratory of Oil and Gas Reservoir Geology and Exploitation, Chengdu University of Technology, Chengdu 610059, P.R. China
* Email: houmc@cdut.edu.cn

Abstract: Micrographs of thin sections have been the basis for geological research, which are even more important now than ever as big data is becoming a new engine of scientific discovery. Micrograph data is a powerful form of geoscience data. In view of this, we collected samples, segmented thin sections, performed identification under microscope and took micrographs for the Middle Jurassic clastic rocks in the northeastern Ordos Basin. The dataset consists of 208 plane- and cross-polarized micrographs of 78 samples in total, under four or ten times magnification. Most of the samples are lithic sandstone, followed by fine grained sandstone. The quartz consists largely of single-crystal quartz, while the feldspar is primarily microcline and plagioclase. Debris is mostly clay and siliceous debris, followed by metamorphic debris. The micrographs can reflect the mineralogy and sedimentary formation of respective strata in the study area. The dataset can be used to reveal the correlation between various geological factors, and on the basis of that, to attribute the causes of varied geological formations .

Keywords: geological data; micrograph data; informatization; mineralogy; sedimentology

Dataset Profile

Title	A dataset of Middle Jurassic clastic rocks in northeastern Ordos Basin
Data authors	Chao Hui, Hou Mingcai, Cao Haiyang and Liang Ziruo
Data corresponding author	Hou Mingcai (houmc@cdut.edu.cn)
Time range	The rock thin sections were collected in 2017, with a stratigraphic age attributable to Middle Jurassic (Yanan Formation-Zhiluo Formation; ~174-163 Ma). The photomicrographs were taken in 2018.
Geographical scope	Both Shenshangou (39°45′89″N, 110°11′03″E) and Gaotouyao (40°2′6.8″N, 109°38′3.7″E) sections are in northeastern Ordos Basin.
Polarized microscope resolution	1280*960 pixels
Data volume	5.01 GB
Data format	*.tif, *.xls, *.jpg
Data service system	<https://dx.doi.org/10.11922/sciencedb.j00001.00087>
Source of funding	National Key Basic Research Program of China (973 Program) (Grant No. 2015CB453001)
Dataset composition	The dataset includes three data files, which are: (1) “Photomicrographs.zip”, which is a dataset of polarized photomicrographs of the rock thin sections in tif format, with a data volume of 4.91 GB; (2) “Field photos of the measured sections.zip”, which contains the field photos of the measured sections and the characteristics of outcrops, with a data volume of 107MB; (3) “Information table.xls”, which contains profile and corresponding identification data of the rock thin sections, with a data volume of 33KB.