



文献 CSTR:

32001.14.11-6035.csd.2023.0087.zh



文献 DOI:

10.11922/11-6035.csd.2023.0087.zh

数据 DOI:

10.57760/sciencedb.j00001.00793

文献分类: 地球科学

收稿日期: 2023-04-06

开放同评: 2023-04-27

录用日期: 2024-09-20

出版日期: 2024-11-26

* 论文通信作者

韩晓天: hxt730716@163.com

2020–2021 年陕晋冀地区五种破坏类型井下构造煤 的深度学习图像数据集

冯永安^{1,2}, 韩晓天^{2*}, 刘铁³, 吕伏^{1,2}

1. 辽宁工程技术大学, 鄂尔多斯研究院, 内蒙古鄂尔多斯 017000

2. 辽宁工程技术大学, 软件学院, 辽宁葫芦岛 125105

3. 中煤科工集团沈阳研究院有限公司, 辽宁抚顺 113122

摘要: 井下煤体破坏类型的机器视觉识别可以为煤与瓦斯突出事故的预测、井下安全的评估和智慧矿山工程的建设提供信息支撑和技术支持。目前煤矿中构造煤的煤体破坏类型判别仍以经验性地宏观物理观察辨识为主, 但该方法受主观经验和环境因素影响较大。构建构造煤的图像分类机器学习模型需要高质量、均衡、真实的井下构造煤样本图像。本数据集通过人为携带高清防爆相机在多个不同矿井环境下进行拍摄采集, 提供了可用于深度学习的井下原生状态破坏煤图像集。数据集大小约为 1.55 GB, 包含不同矿井环境、光照、角度下的非破坏煤、破坏煤、强烈破坏煤、粉碎煤和全粉煤, 共计 1031 张图像, 平均每类 206 张, 能够满足深度学习模型的基础训练需要。在数据采集过程中, 采取了严格的数据质量控制措施, 对图像的清晰度、真实性等进行了评估, 确保了数据的可靠性。本数据集针对构造煤破坏类型进行构建, 为下一阶段基于机器学习的煤体破坏类型识别研究提供必要的数据基础和研究素材。

关键词: 构造煤; 煤体破坏类型; 井下环境; 深度学习; 识别; 训练

数据库(集)基本信息简介

数据库(集)名称	2020–2021 年中国煤矿五种破坏类型井下构造煤的深度学习图像数据集
数据通信作者	韩晓天 (hxt730716@163.com)
数据作者	吕伏、韩晓天、刘铁
数据时间范围	2020–2021 年
地理区域	孟村煤矿(陕)、神东煤矿(晋)、阜生煤矿(冀)
数据量	1.55 GB
数据格式	*.jpg
数据服务系统网址	https://doi.org/10.57760/sciencedb.j00001.00793
基金项目	国家自然科学基金青年基金(51904144); 国家自然科学基金面上项目(52274206)
数据库(集)组成	Tectonic Coal 数据集共包括 5 个数据文件, 1031 张图片, 其中 (1) 01.Non-destructive coal 文件存储非破坏煤图像; (2) 02.Destructive coal 文件存储破坏煤图像; (3) 03.Strongly destructive coal 文件存储强烈破坏煤图像; (4) 04.Pulverized coal 文件存储粉碎煤图像; (5) 05.Fully pulverized coal 文件存储全粉煤图像。

引言

中国是世界上耗煤量最大的国家之一。2021 年,煤炭能源消费量约占中国能源消费总量的 56%。在未来很长一段时间内,它仍将是能源消耗的主要来源^[1]。因此,研究智能高效的煤矿安全生产设备至关重要。中国地质在主要成煤期后经历了多期构造运动叠加,在构造应力的作用下形成不同变形程度的构造煤^[2]。构造煤(Tectonic coal),又称构造变形煤(Tectonically deformed coal)、变形煤(Deformed coal)、破坏煤(Broken coal)等^[3]。苏联矿业研究所在 1958 年依据裂隙密度和煤质特征等,将构造煤分为非破坏煤、破坏煤、强烈破坏煤、粉碎煤和全粉煤 5 种类型^[4]。不同破坏类型的煤体对瓦斯的吸附能力不同,并且物理特性和力学性质相差较大。构造煤发育区是瓦斯突出事故的高发区,这已是采矿工程中的基本认知,特别是韧性变形煤,由于其特殊的物理和化学结构决定了其高瓦斯含量和低渗透的特性,该类构造变形煤一般处于矿井瓦斯突出的危险地带^[5]。而变形程度较低的脆性构造变形煤,如非破坏煤和轻度破坏煤,其裂隙发育、连通性好,渗透率高、分布范围大,具有优良的煤层气储特性,且瓦斯突出危险性低^[6]。因此,构造煤类型的研究能够为瓦斯突出、煤层气勘探和开发提供重要的理论支撑和科学基础。

经验性地宏观物理观察辨识是目前井下采煤现场判断煤体破坏类型的常用方法,但该方法受主观经验和环境因素影响较大。其他常规的构造煤识别方法还有井下采样手标本描述法、钻孔取芯法和测井曲线识别法^[7]。前两种方法存在效率低、取芯率低,依靠经验识别准确率低等缺点。测井曲线识别法根据不同结构类型煤体测井响应特征不同的特点,利用小波分析方法提取测井曲线中的分量信息;再通过聚类分析划分不同结构类型的煤体。这种方法提高了准确率,但现场即时反应度不够,难以在开采现场得到大规模的常规应用。

近年来,随着深度学习的迅速发展,根据图像特征分类的方法凭借其低成本、高效率和模型可移植性强等特点成为当前热点研究方向之一。在计算机辅助识别煤岩领域,图像识别方法逐渐成为热门研究领域。由于煤岩产生过程和地质条件的差异,不同类型的煤岩在纹理、形状和反射面面积方面存在较大差异,图像分类法即依靠图像的特征属性进行分类。伴随各种大型标记数据集的出现、图形处理工具的进步,以及图像深度学习模型的提升,基于图像的识别方法被广泛应用于工业领域,这种基于图像的识别方法,不仅能够实现高准确率、高效率,同时能够解放人力需求,提高井下安全性。关于煤岩的深度学习图像分类数据集较少,且主要集中于煤矸识别^[8],以及不同类型的矿石识别^[9]。同时,数据集图像多来自实验室环境下的拍摄取样、显微样本或井下皮带堆煤图像,井下原生状态的矿壁煤岩样本尚且不足。本数据集致力于提供煤岩中的构造煤分类图像样本。由于构造煤分类的主要价值是提供井下煤岩状态信息,对瓦斯突出事故的预测等提供信息支持,因此构造煤的分类需要在井下环境完成。本数据集拍摄的照片是井下原生状态煤的实际拍摄,拍摄环境包括煤壁,矿井散落煤堆等,具有井下环境的真实性,有利于井下实际环境的构造煤分类应用。

本研究基于煤体破坏类型井下识别模型的构建任务,人工携带高清防爆相机,在满足井下安全标准的情况下到多处煤矿井下采集真实图像。考虑到深度学习模型训练任务对数据量和图像质量的高要求,在进行预处理和图像扩增后,建立了一个可以满足深度学习基本要求的构造煤破坏类型图像数据集,弥补了当前煤岩研究领域相关数据资源稀少的缺陷。同时,数据集中构造煤均处于井下的实际环境中,具有足够的真实性和环境多样性,能够适用于真实井下破坏煤识别的模型构建。本数据集能够为煤体识别和计算机视觉的跨学科研究提供数据基础,具有较高的价值和实用性。

1 数据采集和处理方法

1.1 数据采集方法

项目组于 2020 年 11 月–2021 年 3 月分别在孟村煤矿、神东煤矿、阜生煤矿及其下属煤矿进行不同破坏类型构造煤的图像采集。图像采集工作由矿上专业技术人员按以下流程进行操作：

(1) 根据构造煤在煤体的光泽度、结构特点、断裂特性、强度等呈不同状态^[4]。依据其物理特性的不同将构造煤分为五类：非破坏煤（Non-destructive coal），破坏煤（Destructive coal），强烈破坏煤（Strongly destructive coal），粉碎煤（Pulverized coal）和全粉煤（Fully pulverized coal）。其具体分类标准如表 1 所示。

表 1 不同煤体破坏类型的特点

Table 1 Characteristics of different destruction types of tectonic coal

煤体破坏类型	光泽度	结构特征	强度
非破坏煤 Non-destructive coal	有光泽和半光泽	大块岩石，可数的岩石裂隙	很难打碎
破坏煤 Destructive coal	有光泽和半光泽	岩石裂隙较多	较难碎裂
强烈破坏煤 Strongly destructive coal	有光泽和半光泽	碎片结构	较容易碎裂成块
粉碎煤 Pulverized coal	无光泽	颗粒状碎石	相对较软，可压碎
全粉煤 Fully pulverized coal	无光泽	粉末状	软、松碎

(2) 携带电池进行替换后的符合井下安全标准的防爆型手机，经过安全审查后到井下进行采集。采集设备分别采用两种不同的智能手机 vivo 和 HUAWEI Mate 30，相机型号分别为 vivo X27 和 TAS-AN00，拍摄原始图像采用 3000 × 4000 分辨率进行拍摄。

(3) 由于井下环境的复杂性，为保证图像采样环境的丰富多样，满足模型训练要求，需要在多角度、多光线环境、多尺寸下进行拍摄。所获取图像具有多旋转角度和姿态特征，包含多尺度变化和明暗变化。

(4) 对原始采集图像进行煤体破坏类型的鉴定、分类和筛选。过程中根据通用分类标准，经专家讨论和整合意见后进行分类，保证破坏类型分类的准确性。同时，筛选优质图像，删减质量差、光影杂乱、杂物遮挡过大的图像。

1.2 数据处理

筛选完成的原始图像存在不同破坏类型混合、边角杂物遮挡、数据分布不平衡等情况，因此需要进行进一步的数据处理。

(1) 对混合类型的图像进行裁剪、分割、再分类；对有小型遮挡物、杂物图像进行裁剪，去除杂物；出现数据不平衡的情况，对所包含数据进行数据增强，包含旋转、裁剪、对称等操作，最终使得五类图像所包含数据数量达到基本平衡。

(2) 对分类完成的图像再次筛选完成后按照数据标签存储在不同文件夹下，以破坏类别英文名称作为文件夹标签，防止模型训练过程中出现乱码问题。

(3) 图像名称以“类别名_序号”的格式进行存储。

(4) 将数据集按小型数据集的常规训练标准进行划分，训练集与测试集比例为 4: 1，训练集

占全部图像的 80%，测试集占 20%。采用的划分方式为 python 的 `train_test_split` 函数，保证每类图像按照比例进行随机分配。

2 数据样本描述

煤体图像数据集以 Tectonic Coal 命名，包含 5 个文件夹，分别以破坏类型英文名称命名，每张图像代表一个煤体样本。数据集中样本示例如图 1 所示，分别是非破坏煤、破坏煤、强烈破坏煤、粉碎煤和全粉煤的示例图。



(a) 非破坏煤 Non-destructive coal



(b) 破坏煤 Destructive coal



(c) 强烈破坏煤 Strongly destructive coal



(d) 粉碎煤 Pulverized coal



(e) 全粉煤 Fully pulverized coal

图 1 五种破坏类型的构造煤示例图

Figure 1 Samples of five destruction types of tectonic coal

数据集的各类数据量分布如图 2 所示。

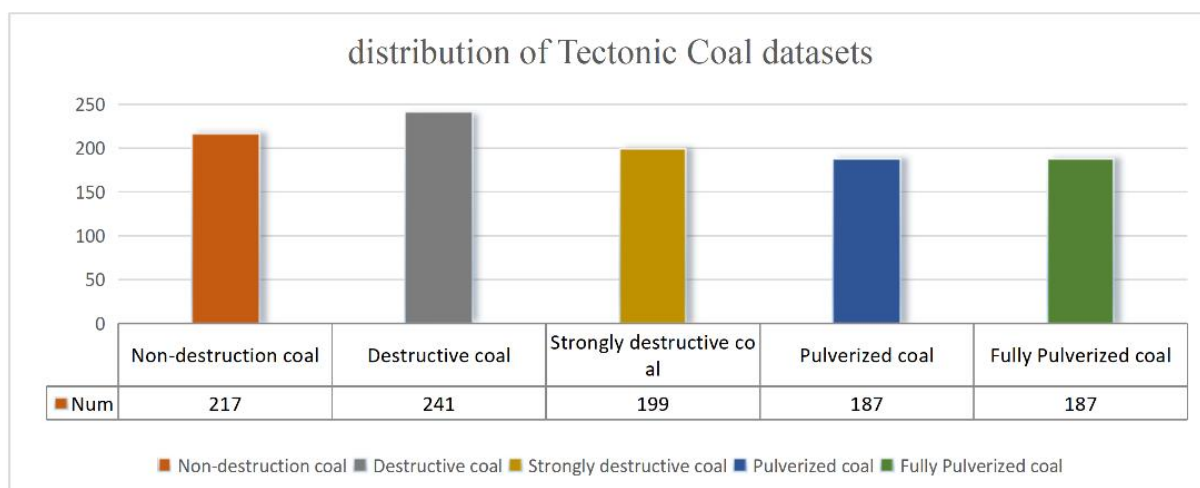


图 2 五种破坏类型煤体图像数据分布

Figure 2 Distribution of images for five destruction types of tectonic coal

3 数据质量控制和评估

本数据集的构建过程中，数据采集和数据集处理均在专业标准要求下，由专业技术人员按照标准操作流程进行操作。为保证数据的质量，采取了以下措施：（1）从图像采集标准、采集数量、采集人员、采集设备等方面严格要求操作流程的准确性；（2）现场拍摄图像并预先记录拍摄场所的主要煤体破坏类型，确保后续分类的有据可依；（3）确保图像采集时多角度、多光照环境下、多尺度进行拍摄，保证真实环境的多样性；（4）数据标注阶段注意混合型 and 中间型煤体图像。对混合型图像进行图像切割或删除；对中间型图像标签分类向专业技术人员请教、核实，确保所有类型图像标签正确。（5）对所有图像进行多轮人工筛查，确保剔除模糊图像、反射严重等不合格图像，保证本数据集数据的质量和可靠性。

4 数据价值

本数据集是基于构造煤煤体破坏类型构建的图像数据集，由专业技术人员在三大煤矿及其下属煤矿的井下环境进行采集，在数据采集、筛选和加工处理过程中，完全体现了数据集构造的真实性、标准性和可靠性准则。本数据根据标准的五类煤体破坏类型的分类准则，为领域相关研究人员提供了可用于深度学习的可靠图像数据。随着未来的不断更新和数据扩充，本数据集将为煤体破坏类型的图像识别构建坚实的数据资源基础，并随着深度学习算法的发展，共同推动构造煤图像识别研究的发展，为“智慧矿山”的发展添砖加瓦。

致 谢

感谢煤矿技术人员对本研究的支持。

作者分工职责

冯永安（1978—），男，广东信宜人，硕士，副教授，研究方向为智能数据处理、大数据技术及图像识别研究等。主要承担工作：总体方案涉及与组织实施。

韩晓天（1998—），女，河北保定人，硕士，研究生，研究方向为人工智能及计算机视觉。主要承担工作：数据整理、汇总整理及论文撰写。

刘铁（1996—），男，辽宁营口人，硕士，工程师，研究方向为图像与视觉信息处理、煤矿应急救援技术。主要承担工作：图像采集。

吕伏（1980—），女，辽宁鞍山人，博士，副教授，研究方向为智能数据处理、图像与视觉信息计算。主要承担工作：方案设计与图像资源汇总。

参考文献

- [1] 张吉雄, 张强, 周楠, 等. 煤基固废充填开采技术研究进展与展望[J]. 煤炭学报, 2022, 47(12): 4167–4181. DOI: 10.13225/j.cnki.jccs.2022.1053. [ZHANG J X, ZHANG Q, ZHOU N, et al. Research progress and prospect of coal based solid waste backfilling mining technology[J]. Journal of China Coal Society, 2022, 47(12): 4167–4181. DOI: 10.13225/j.cnki.jccs.2022.1053.]
- [2] 琚宜文, 林红, 李小诗, 等. 煤岩构造变形与动力变质作用[J]. 地学前缘, 2009, 16(1): 156–166. DOI: 10.3321/j.issn: 1005-2321.2009.01.018. [JU Y W, LIN H, LI X S, et al. Tectonic deformation and dynamic metamorphism of coal[J]. Earth Science Frontiers, 2009, 16(1): 156–166. DOI: 10.3321/j.issn: 1005-2321.2009.01.018.]
- [3] 姜波, 李明, 屈争辉, 等. 构造煤研究现状及展望[J]. 地球科学进展, 2016, 31(4): 335–346. DOI: 10.11867/j.issn.1001-8166.2016.04.0335. [JIANG B, LI M, QU Z H, et al. Current research status and prospect of tectonically deformed coal[J]. Advances in Earth Science, 2016, 31(4): 335–346. DOI: 10.11867/j.issn.1001-8166.2016.04.0335.]
- [4] YUE J W, WANG Z F, CHEN J S, et al. Investigation of pore structure characteristics and adsorption characteristics of coals with different destruction types[J]. Adsorption Science & Technology, 2019, 37(7/8): 623–648. DOI: 10.1177/0263617419868076.
- [5] 姜波, 秦勇, 琚宜文, 等. 构造煤化学结构演化与瓦斯特性耦合机理[J]. 地学前缘, 2009, 16(2): 262–271. DOI: 10.3321/j.issn: 1005-2321.2009.02.020. [JIANG B, QIN Y, JU Y W, et al. The coupling mechanism of the evolution of chemical structure with the characteristics of gas of tectonic coals[J]. Earth Science Frontiers, 2009, 16(2): 262–271. DOI: 10.3321/j.issn: 1005-2321.2009.02.020.]
- [6] 姜波, 秦勇, 琚宜文, 等. 煤层气成藏的构造应力场研究[J]. 中国矿业大学学报, 2005, 34(5): 564–569. DOI: 10.3321/j.issn: 1000-1964.2005.05.005. [JIANG B, QIN Y, JU Y W, et al. Research on tectonic stress field of generate and reservoir of coalbed methane[J]. Journal of China University of Mining & Technology, 2005, 34(5): 564–569. DOI: 10.3321/j.issn: 1000-1964.2005.05.005.]
- [7] 陈萍, 张荣飞, 唐修义. 对利用测井曲线判识构造煤方法的认识[J]. 煤田地质与勘探, 2014, 42(3): 78–81. DOI: 10.3969/j.issn.1001-1986.2014.03.018. [CHEN P, ZHANG R F, TANG X Y. Some understanding of identifying tectonic coal by logging curve[J]. Coal Geology & Exploration, 2014, 42(3):

78–81. DOI: 10.3969/j.issn.1001-1986.2014.03.018.]

[8] 薛光辉, 李秀莹, 钱孝玲, 等. 基于随机森林的综放工作面煤矸图像识别[J]. 工矿自动化, 2020, 46(5): 57–62. DOI: 10.13272/j.issn.1671-251x.2019110064. [XUE G H, LI X Y, QIAN X L, et al. Coal-gangue image recognition in fully-mechanized caving face based on random forest[J]. Industry and Mine Automation, 2020, 46(5): 57–62. DOI: 10.13272/j.issn.1671-251x.2019110064.]

[9] 赖文, 蒋璟鑫, 邱检生, 等. 南京大学岩石教学薄片显微图像数据集[J/OL]. 中国科学数据, 2020, 5(3). (2020-08-28). DOI: 10.11922/csdata.2020.0071.zh. [LAI W, JIANG J X, QIU J S, et al. A photomicrograph dataset of rocks for petrology teaching at Nanjing University[J/OL]. China Scientific Data, 2020, 5(3). (2020-08-28). DOI: 10.11922/csdata.2020.0071.zh.]

论文引用格式

冯永安, 韩晓天, 刘铁, 等. 2020 - 2021 年陕晋冀地区五种破坏类型井下构造煤的深度学习图像数据集[J/OL]. 中国科学数据, 2024, 9(4). (2024-11-18). DOI: 10.11922/11-6035.csd.2023.0087.zh.

数据引用格式

吕伏, 韩晓天, 刘铁. 2020–2021 年中国煤矿五种破坏类型井下构造煤的深度学习图像数据集 [DS/OL]. V1. Science Data Bank, 2023. (2023-04-11). DOI: 10.57760/sciencedb.j00001.00793.

A dataset of deep learning images of underground tectonic coal with five destruction types in Shaanxi, Shanxi and Hebei (2020–2021)

FENG Yong'an^{1,2}, HAN Xiaotian^{2*}, LIU Tie³, LÜ Fu^{1,2}

1. Ordos Institute, Liaoning Technical University, Ordos 017000, P. R. China

2. Software College, Liaoning Technical University, Huludao 125105, P. R. China

3. CCTEG Shenyang Engineering Co., Ltd., Fushun 113122, P. R. China

*Email: hxt730716@l63.com

Abstract: Machine vision recognition of underground coal destruction types can provide crucial information and technical support for predicting coal and gas herniation accidents, evaluating underground safety, and advancing the construction of intelligent mine engineering. At present, the discrimination of coal destruction types of tectonic coals in mines still relies on empirical macro-physical observations, which are heavily influenced by subjective experience and environmental conditions. To develop an effective machine learning model for classifying tectonic coal destruction types, it is essential to use high-quality, balanced, and realistic sample images of underground tectonic coal. Based on the photos taken by human-carried high-definition explosion-proof cameras in several different mine environments, this dataset provides a set of downhole native-state destructive coal images suitable for deep learning. The dataset size

is about 1.55 G, including images of Non-destructive coal, Destructive coal, Strongly damaged coal, Pulverized coal and Fully pulverized coal under different mine conditions, illumination and angles. A total of 1,031 images, with an average of 206 images per class, can meet the basic training needs of deep learning models. During the data collection process, strict data quality control measures were taken, and the clarity and authenticity of the images were evaluated to ensure the reliability of the data. Designed for recognizing tectonic coal destruction types, this dataset is expected to provide essential resources for the future machine learning-based research on the recognition of coal destruction types.

Keywords: tectonic coal; destruction types; downhole environment; deep learning; identify; training

Dataset Profile

Title	A dataset of deep learning images of underground tectonic coal with five destruction types in China (2020–2021)
Data corresponding author	HAN Xiaotian (hxt730716@163.com)
Data authors	LÜ Fu, HAN Xiaotian, LIU Tie
Time range	2020–2021
Geographical scope	Mengcun Coal Mine, Shendong Coal Mine, and Fusheng Coal Mine
Data volume	1.55 GB
Data format	*.jpg
Data service system	< https://doi.org/10.57760/sciencedb.j00001.00793 >
Sources of funding	National Natural Science Foundation of China Youth Fund (51904144); National Natural Science Foundation China (General Program) (52274206)
Dataset composition	The dataset consists of 5 data files with 1,031 images: (1) 01.Non-destructive coal files, storing non-destructive coal images, (2) 02.Destructive coal files, storing destructive coal images, (3) 03.Strongly destructive coal file, storing images of strongly destructive coal, (4) 04.Pulverized coal file, storing pulverized coal images, and (5) 05.Fully pulverized coal file, storing fully pulverized coal images.