

青岛市大气颗粒物 SEM 图像数据集

尹唱唱¹, 赵猛¹, 王晓涵², 程学珍^{1*}

ISSN 2096-2223

CN 11-6035/N



文献 CSTR:

32001.14.11-6035.csd.2022.0101.zh



文献 DOI:

10.11922/11-6035.csd.2022.0101.zh

数据 DOI:

10.11922/sciencedb.01170

文献分类: 地球科学

收稿日期: 2022-10-29

开放同评: 2023-01-30

录用日期: 2023-12-01

发表日期: 2024-06-26

1. 山东科技大学电气与自动化工程学院, 青岛 266590

2. 山东科技大学电子信息工程学院, 青岛 266590

摘要: 本文选取青岛市 7 个具有代表性的采样点, 分时段进行大气颗粒物样本的采集, 然后通过扫描电镜, 进一步得到颗粒物的 SEM (Scanning Electron Microscope) 图像。通过对得到 SEM 图像的形貌特征进行观察分析, 进而挑选出含有形貌特征明显的颗粒物的 SEM 图像, 共计 334 张图片。基于得到的大气颗粒物的 SEM 图像进行分析研究, 根据其形貌特性的不同将颗粒物分为 7 类: 链条状颗粒物、絮状颗粒物、纤维状颗粒物、球状颗粒物、类球状颗粒物、不规则矿物颗粒物、规则矿物颗粒物。将得到的 SEM 图片采用水平翻转、色彩平衡、亮度变换、模糊处理等相应的图像处理方法进行颗粒物 SEM 图像样本的扩展, 生成新的图片样本, 并将其添加到数据集中。从而建立起含有 7 类大气颗粒的 SEM 图像数据集, 共含有 2672 张 SEM 图片。本文所建立的数据集可以用于采用机器学习等相关方法对大气颗粒物进行识别分类、分割等相关研究, 也可为大气颗粒物相关研究提供基础数据资料。

关键词: 大气颗粒物; SEM 图像; 形貌特性; 青岛市

数据库(集)基本信息简介

数据库(集)名称	青岛大气颗粒物 SEM 图像数据集
数据通信作者	程学珍 (chengxuezheng@sdust.edu.cn)
数据作者	尹唱唱, 赵猛, 王晓涵, 程学珍
数据时间范围	2016–2018年
地理区域	地理范围为35°93′–36°07′N, 120°12′–120°25′E, 其中包括青岛市的青岛港、金沙滩、中德生态园、恒源热电有限公司、琴岛之眼、新街口北、山东科技大学
空间分辨率	1536×1096像素
数据量	967.5 MB
数据格式	*.jpg, *.xml
数据服务系统网址	http://dx.doi.org/10.11922/sciencedb.01170
基金项目	国家自然科学基金 (62073198)
数据库(集)组成	本数据集包括2672张7类大气颗粒物的SEM图像, 数据量共计966 MB; 每张SEM图像所对应的.XML文件, 包含SEM图像中颗粒物的位置信息以及对应的种类, 数据量共计1.5 MB。

* 论文通信作者

程学珍: chengxuezheng@sdust.edu.cn

引言

大气颗粒物是大气中存在的各种固态和液态颗粒状物质的总称^[1]，其危害人体健康并且影响环境质量，同时在全球气候变化中扮演着重要角色^[2]。在不同的大气污染物中，颗粒物的性质、组成和粒径分布等表现得很复杂。其中大气颗粒物中的可吸入颗粒物 PM_{10} 、细颗粒物 $PM_{2.5}$ 为特征污染物的防治工作受到广泛关注^[3]。特别是细颗粒物 $PM_{2.5}$ ，其空气动力学当量直径小于等于 $2.5\ \mu m$ ，颗粒物粒径小，比表面积大，因此能够携带大量有毒有害物质，并且能够长时间漂浮在大气中和长距离漂移^[4-5]，能通过呼吸系统进入血液，对人类健康产生很大影响^[6]。此外，大气中的 $PM_{2.5}$ 影响气候和环境，已经引起社会各界的重视^[7]。

长期以来，针对大气颗粒物的研究主要集中在质量浓度、化学成分、源解析、时空分布等方面。基于颗粒物的形貌特性，采用图像处理的方法对大气颗粒物进行相关研究分析较少，并且缺少相关数据集。颗粒物的形貌特性不同对人体的危害程度也就不同，因此基于形貌特征对大气颗粒物进行研究具有重要意义^[8]。通过电子扫描显微镜可以对采集得到的颗粒物样本进行放大，从而得到大气颗粒物的微观形貌特征图像。通过对大气颗粒物的形貌特性进行观察和分析，可以研究颗粒物的成分及来源，从而更好地对大气颗粒物进行源解析^[9]，进而更好地对大气颗粒物的污染进行防治。

采用机器学习的方法对大气颗粒物 SEM 图像进行研究，需要相应的数据集。而 SEM 图像获取较为复杂，并且对大气颗粒物的 SEM 图像分类和标记也需要专业的知识，现在这方面的数据集还处于空白阶段。需要针对大气颗粒物的 SEM 图像进行颗粒物的相关研究的学者缺少相应的公开的数据集。基于以上原因，作者花费大量时间进行颗粒物的采集、制备、生成 SEM 图像，并进行相应的分类标记，建立了含有 7 类大气颗粒的 SEM 图像数据集。本数据集包含了根据形貌特征分类的 7 类大气颗粒物 SEM 图像，颗粒物的标注文件。其中可以观察到大气颗粒物的形貌特征，希望为采用图像处理的方法对大气颗粒物进行相关研究，为大气污染防治的研究提供新的研究基础，实现人机共用，并与需要的学者共享。

1 数据采集和处理方法

选取青岛市不同功能区的采样点进行大气颗粒物的采集，分别是青岛港、金沙滩、中德生态园、恒源热电有限公司、琴岛之眼、新街口北、山东科技大学，共计 7 个采样点。各个采样点的信息如表 1 所示^[10]，采样点位置如图 1 所示。采样分别选择不同的功能区，保证了采集得到的颗粒物样本的多样性，进而保证了数据集内含有的样本数据的多样性。

表 1 青岛市采样点信息

Table 1 Details of sampling points in Qingdao

序号	采样点	采样高度/m	功能区	坐标纬度	坐标经度
1	青岛港	5	仓库码头区	36°01'N	120°20'E
2	金沙滩	5	风景区	35°95'N	120°25'E
3	中德生态园	10	居民区	36°07'N	120°13'E
4	恒源热电有限公司	5	工业区	36°01'N	120°13'E

序号	采样点	采样高度/m	功能区	坐标纬度	坐标经度
5	琴岛之眼	5	商业区	35°93'N	120°18'E
6	新街口北	10	城乡结合区	36°03'N	120°16'E
7	山东科技大学	10	文教区	36°00'N	120°12'E

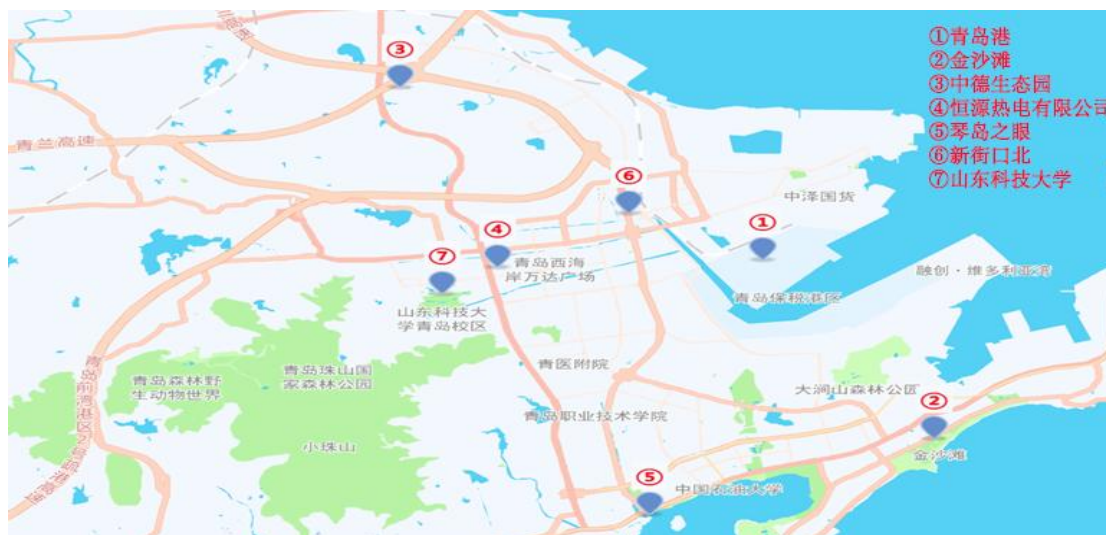


图 1 采样点位置

Figure 1 Location of sampling points

针对大气颗粒物的采样器为环境级精细颗粒物采样器 BGI-PQ200 和崂应 2050 型空气/智能 TSP 综合采样器。采样前对采样器进行流量校正。为了保证数据集内颗粒物背景的多样性，采样滤膜选取 3 种滤膜，分别为石英纤维滤膜，醋酸纤维素滤膜和聚碳酸酯滤膜。石英纤维滤膜呈现出复杂的纤维状，因此其对气流的阻力较小，进而粒径较小的颗粒物容易嵌在纤维缝隙中，从而容易捕获大气颗粒物。醋酸纤维素滤膜相比石英纤维滤膜薄更透气，形态更为粗糙，颗粒物容易嵌在孔洞或者缝隙中，但颗粒物容易被滤膜纤维遮挡^[1]。聚碳酸酯滤膜的微观形态清晰，呈现为表面光滑的微孔形态。由于颗粒物采样采用的聚碳酸酯滤膜的孔径为 0.1 μm 的，孔洞比较小，颗粒物附着在滤膜表面，不易嵌入孔洞中，从而得到的图像中的颗粒物不易被遮挡。数据集中对大气颗粒物进行采集时采用不同的滤膜进行采样，从而得到的大气颗粒物的图像具有不同的背景，更有利于后续使用数据集过程中网络模型的通用性。

为保证滤膜上采集的颗粒物数量足够并且分布均匀，在进行大气颗粒物的样本采集的时候采取分时段的方式。考虑气候对颗粒物样本的影响，选取的采样时间为 2016 年夏季、2016 年秋季、2017 年春季、2017 年秋季及 2018 年春季、2018 年冬季，每次样本采集时间为 2–5 小时。每次采样之前都留有空白滤膜，方便与采有颗粒物样本的滤膜进行对比分析，用来对比分析滤膜是否受到污染，保证采样样本的准确性。

得到采集到大气颗粒物的滤膜后，接着通过扫描电镜完成对大气颗粒物样本的制作。本文采用的扫描电子显微镜是美国 FEI Nova Nano SEM 450。首先在采样得到的滤膜上随机裁剪为大小为 5 mm×10 mm 的样本，通过导电胶将裁剪得到的滤膜粘贴在扫描电镜的样本台上，并在真空条件下通过离子溅射仪在得到的样本表面喷一层铂，从而使样本具有导电性，符合观测要求。将样本台放入扫描电子显微镜，在高真空模式下，设置合适的加速电压和放大倍数，获得大气颗粒物的 SEM

图像。通过扫描电镜得到的 SEM 图像有的含有单个颗粒物，有的含有多个颗粒物，而且图片底部包含有相关的文字标注信息。

2 数据样本描述

基于大气颗粒物的形貌特性，从颗粒物的 SEM 图像中挑选出含有特征明显的颗粒物的作为数据集的样本。根据不同颗粒物的 SEM 图像形貌特性的不同，将数据集中的颗粒物分为 7 类，分别为：链条状颗粒物、絮状颗粒物、纤维状颗粒物、球状颗粒物、类球状颗粒物、不规则矿物颗粒物、规则矿物颗粒物。7 类颗粒物的形貌如图 2 所示。

(a) 链条状颗粒物：属于大气中的烟尘集合体，主要来自燃烧源排放，包括燃煤、机动车尾气以及生物质的燃烧等^[12]。

(b) 絮状颗粒物：由大量微小的圆形颗粒物聚集在一起形成“蓬松状”集合体。

(c) 纤维状颗粒物：一般来自纤维状的矿物、植物纤维和空气中飘散的人工纤维。

(d) 球状颗粒物：主要来源于燃煤飞灰，形态呈现出光滑的球形，特征比较明显，主要是由固体飞灰在高温条件熔融形成^[13]。

(e) 类球状颗粒物：是飞灰由于各种原因发生了变形，从而呈椭圆形、局部呈现出缺陷等^[14]。

(f) 不规则矿物颗粒物：来源主要有风沙、道路扬尘、建筑施工等^[15]。

(g) 规则矿物颗粒物：一般来自扬尘和二次大气化学反应的产物。

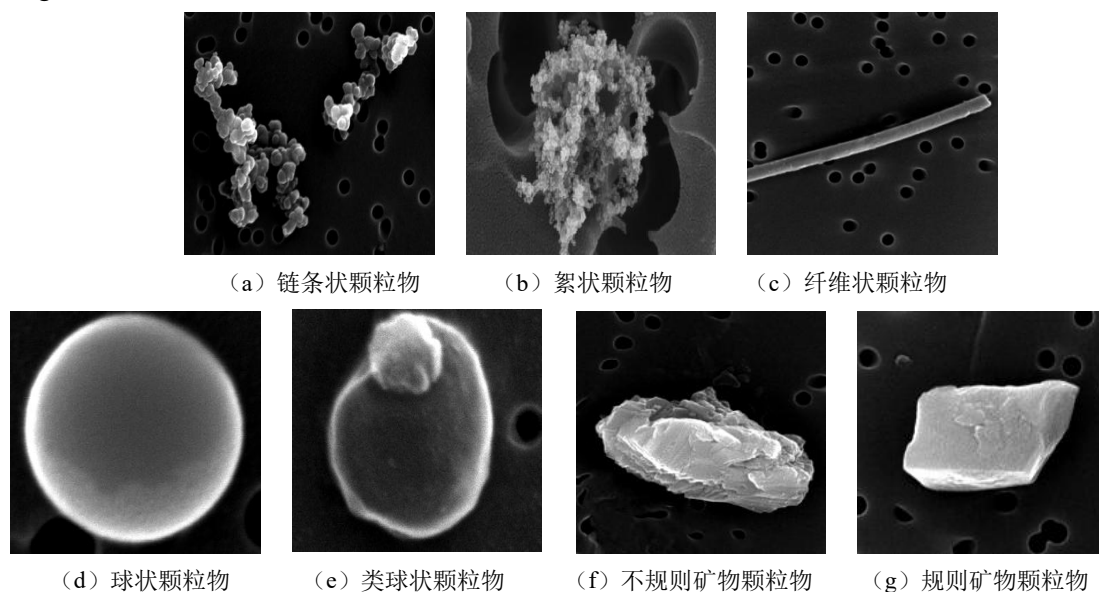


图 2 7 类大气颗粒物的 SEM 图像

Figure 2 SEM images of seven types of atmospheric particulates

数据集中含有 7 类颗粒物的原始 SEM 图像一共 334 张，随后将得到的 SEM 图像通过 LabelImg 软件进行手动标注，标注得到的数据集内的样本图像统一保存为 jpg 格式。由于 SEM 图像样本获得比较费时费力，为了扩展数据集的样本数，进一步将数据集的原始图像通过旋转 90° 、 180° 、 270° 、水平翻转、亮度变换、色彩平衡、模糊处理^[16]的方法进行数据的扩展，生成 7 组新的图像样本，从而建立起含有 7 类大气颗粒的 SEM 图像数据集。

数据集共含有 2672 张 SEM 图像，其中编号为 000001-002400 的图像采用的是聚碳酸酯滤膜，采集时间为 2018 年春季和冬季；编号为 002401-002568 的图像采用的是醋酸纤维素滤膜，采集时

间为 2017 年春季和秋季；编号为 002569-002672 的图像采用的是石英纤维滤膜，采集时间为 2016 年夏季和秋季。数据集中链条状颗粒物占 14.22%、絮状颗粒物占 14.52%、纤维状颗粒物占 36.58%、球状颗粒物占 6.83%、类球状颗粒物占 4.11%、不规则矿物颗粒物占 20%、规则矿物颗粒物占 3.72%。数据集中还含有相对应的标注文件，XML 标注文件中有相应的颗粒物的位置信息、种类信息等相关信息。

3 数据质量控制和评估

本数据集提供 7 类大气颗粒物 SEM 图片，具有典型的代表性，并且根据其形貌特性已标记出颗粒物的类型和相应的坐标位置。数据集中的 SEM 图片的分辨率为 1536×1096 ，保存格式为 JPG，相应的标注文件为 XML 格式。本数据集的 SEM 图像清晰可见，具有较高的专业性。

4 数据价值

本数据集选取青岛市 7 个采样点，采样点分别为不同的功能区，具有代表性。数据集提供了 7 种典型的大气颗粒物 SEM 图片。本数据集可以作为采用机器学习的方法对大气颗粒物的 SEM 图像进行分类、识别、分割的参考样本基础。可以为基于图像处理的方法对大气颗粒物进行相关研究提供相应的图像，也可以作为基于大气颗粒物的 SEM 图像对大气颗粒物进行相关分析研究的参考数据集。

5 数据使用方法和建议

(1) 本数据集里具有相应的标注好类别的大气颗粒物 SEM 图片和相应的颗粒物坐标文件。现有的分类可以为将来颗粒物的分类识别打下基础，并为其提供相应的分类参考标准。

(2) 由于本数据集已经发表了一些相关研究成果^[8,10,13,16]，可以进一步阅读，从而了解更多的研究情况。

(3) 如果单纯使用数据集，可以直接下载使用。如果需要进一步解决颗粒物相关问题，需要进一步结合数据集做具体分析。

数据作者分工职责

尹唱唱（1983—），女，安徽固镇人，博士生，研究方向为图像处理、模式识别以及机器学习。主要承担工作：颗粒物样本采集、SEM 图像样本分类、论文撰写。

赵猛（1978—），男，辽宁营口人，博士，副教授，研究方向为图像处理、模式识别以及机器学习。主要承担工作：颗粒物采集计划设计、数据集的设计、论文修改。

王晓涵（1997—），女，山东青州人，硕士生，研究方向为图像处理、模式识别。主要承担工作：颗粒物的采集、SEM 图像的制备。

程学珍（1964—），女，山东沂水人，博士，教授，研究方向检测技术与系统集成、信息与图像处理。主要承担工作：颗粒物采集计划设计、数据集的设计、论文撰写。

参考文献

- [1] 王永红, 刘豪, 彭景权, 等. 贵阳郊区空气中PM₁₀和PM_{2.5}质量浓度变化特征分析[J]. 绿色科技, 2015(8): 249–251. [WANG Y H, LIU H, PENG J Q, et al. Analysis of the variations character of mass concentration of PM₁₀ and PM_{2.5} in the suburbs of Guiyang[J]. Journal of Green Science and Technology, 2015(8): 249–251.]
- [2] 朱淑贞, 佟洁, 鲍丰, 等. 廊坊市秋冬季大气细颗粒物污染特征及来源解析[J]. 环境科学, 2023, 44(1): 20–29. DOI: 10.13227/j.hjlx.202112128. [ZHU S Z, TONG J, BAO F, et al. Characteristics and source apportionment of atmospheric fine particles in Langfang in autumn and winter[J]. Environmental Science, 2023, 44(1): 20–29. DOI: 10.13227/j.hjlx.202112128.]
- [3] 赵紫微, 童梦雪, 李源遽, 等. 深圳市餐饮源排放颗粒物的特征[J]. 环境化学, 2020, 39(7): 1763 – 1773. DOI:10.7524/j.issn.0254-6108.2019050804. [ZHAO Z W, TONG M X, LI Y J, et al. Characteristics of particulate matters emitted from cooking in Shenzhen[J]. Environmental Chemistry, 2020, 39(7): 1763–1773. DOI:10.7524/j.issn.0254-6108.2019050804.]
- [4] 李建生, 刘炳光. 大气中二次无机气溶胶的形成反应和清除方法[J]. 无机盐工业, 2018, 50(10): 1–6. [LI J S, LIU B G. Formation reaction and removal methods of inorganic secondary aerosols in atmosphere[J]. Inorganic Chemicals Industry, 2018, 50(10): 1–6.]
- [5] 赵顺征, 易红宏, 唐晓龙, 等. 空气细颗粒物污染的来源、危害及控制对策[J]. 科技导报, 2014, 32(33): 61–66. DOI: 10.3981/j.issn.1000-7857.2014.33.008. [ZHAO S Z, YI H H, TANG X L, et al. Source and harm of air fine particulate matter pollution and control strategies[J]. Science & Technology Review, 2014, 32(33): 61–66. DOI: 10.3981/j.issn.1000-7857.2014.33.008.]
- [6] 刘作云, 刘衡林. 衡阳市PM_{2.5}污染特征及影响因素相关性分析[J]. 湖南生态科学学报, 2021, 8(3): 69–75. DOI: 10.3969/j.issn.2095-7300.2021.03.011. [LIU Z Y, LIU H L. Analysis on the characteristics of pollution and correlation of influencing factors of PM_{2.5} in Hengyang, China[J]. Journal of Hunan Ecological Science, 2021, 8(3): 69–75. DOI: 10.3969/j.issn.2095-7300.2021.03.011.]
- [7] 邵龙义, 王文华, 幸娇萍, 等. 大气颗粒物理化特征和影响效应的研究进展及展望[J]. 地球科学, 2018, 43(5): 1691–1708. [SHAO L Y, WANG W H, XING J P, et al. Physicochemical characteristics and effects of airborne particles: research progress and prospects[J]. Earth Science, 2018, 43(5): 1691–1708.]
- [8] 牛利, 尹唱唱, 刘喜禄, 等. 一种基于迭代直径圆检测的大气颗粒物图像分割方法[J]. 中国科技论文, 2018, 13(20): 2330–2336. DOI: 10.3969/j.issn.2095-2783.2018.20.008. [NIU L, YIN C C, LIU X L, et al. An image segmentation method of atmospheric particulates based on iterative diameter circle detection[J]. China Sciencepaper, 2018, 13(20): 2330–2336. DOI: 10.3969/j.issn.2095-2783.2018.20.008.]
- [9] 王哲, 董树屏, 梁汉东, 等. 内蒙古乌达-乌斯太工业园环境单颗粒研究[J]. 中国环境科学, 2018, 38(2): 478–489. DOI:10.19674/j.cnki.issn1000-6923.2018.0055. [WANG Z, DONG S P, LIANG H D, et al. Study on environmental individual particles in Wuda-Wusitai Industrial Park, Inner Mongolia[J]. China Environmental Science, 2018, 38(2): 478–489. DOI:10.19674/j.cnki.issn1000-6923.2018.0055.]
- [10] 刘喜禄. 基于卷积神经网络的大气颗粒物SEM图像识别[D]. 山东科技大学, 2019. DOI:10.27275/d.cnki.gsdku.2019.001444. [LIU X L. Recognition of SEM Images of Atmospheric Particulates Based on Convolution Neural Networks[D], Qingdao, 2019.]

DOI:10.27275/d.cnki.gsdku.2019.001444.]

[11] 蒋锦晓, 郑昱晴, 陆香君, 等. 大气细颗粒物监测技术研究进展[J]. 杭州师范大学学报(自然科学版), 2019, 18(1): 58–65. DOI: 10.3969/j.issn.1674-232X.2019.01.012. [JIANG J X, ZHENG Y Q, LU X J, et al. Research progress on monitoring technology for atmospheric particles[J]. Journal of Hangzhou Normal University (Natural Science Edition), 2019, 18(1): 58–65. DOI: 10.3969/j.issn.1674-232X.2019.01.012.]

[12] 冯茜丹, 明彩兵, 刘晖, 等. 2011年秋季广州城区大气PM_{2.5}微观形貌和粒度分布[J]. 中国环境科学, 2015, 35(4): 1013–1018. [FENG X D, MING C B, LIU H, et al. Microscopic morphology and size distribution of PM_{2.5} in Guangzhou urban area in fall 2011[J]. China Environmental Science, 2015, 35(4): 1013–1018.]

[13] YIN C C, CHENG X Z, LIU X L, et al. Identification and classification of atmospheric particles based on SEM images using convolutional neural network with attention mechanism[J]. Complexity, 2020, 2020: 9673724. DOI:10.1155/2020/9673724.

[14] PIPAL A S, JAN R, SATSANGI P G, et al. Study of surface morphology, elemental composition and origin of atmospheric aerosols (PM_{2.5} and PM₁₀) over Agra, india[J]. Aerosol and Air Quality Research, 2014, 14(6): 1685–1700. DOI:10.4209/aaqr.2014.01.0017.

[15] 刘童, 王晓军, 陈倩, 等. 烟台市环境受体PM_{2.5}四季污染特征与来源解析[J]. 环境科学, 2019, 40(3): 1082–1090. DOI:10.13227/j.hjlx.201807252. [LIU T, WANG X J, CHEN Q, et al. Pollution characteristics and source apportionment of ambient PM_{2.5} during four seasons in Yantai City[J]. Environmental Science, 2019, 40(3): 1082–1090. DOI:10.13227/j.hjlx.201807252.

[16] 李昕. 基于YOLOV3算法的大气颗粒物SEM图像识别[D]. 青岛: 山东科技大学, 2020. DOI:10.27275/d.cnki.gsdku.2020.000713. [LI X. Atmospheric Particulates Matter SEM Image Recognition Based on YOLOV3 Algorithm[D]. Qingdao: Shandong University of Science and Technology, 2020. DOI:10.27275/d.cnki.gsdku.2020.000713]

论文引用格式

尹唱唱, 赵猛, 王晓涵, 等. 青岛市大气颗粒物 SEM 图像数据集[J/OL]. 中国科学数据, 2024, 9(2). (2024-06-26). DOI: 10.11922/11-6035.csd.2022.0101.zh.

数据引用格式

尹唱唱, 赵猛, 王晓涵, 等. 青岛市大气颗粒物 SEM 图像数据集[DS/OL]. V2. Science Data Bank, 2023. (2023-10-08). DOI: 10.11922/sciencedb.01170.

A dataset of SEM images of atmospheric particles in Qingdao

YIN Changchang¹, ZHAO Meng¹, WANG Xiaohan², CHENG Xuezheng^{1*}

1. College of Electrical Engineering and Automation, Shandong University of Science and Technology, Qingdao 266590, P.R. China

2. College of Electronic and Information Engineering, Shandong University of Science and Technology, Qingdao 266590, P.R. China

*Email: chengxuezheng@sdust.edu.cn

Abstract: In this paper, we selected seven representative sampling points in Qingdao to collect atmospheric particulate samples in different periods, and then obtained the SEM (Scanning Electron Microscope) images of particulate matters through scanning electron microscopy. By observing and analyzing the morphological characteristics of the images, we identified and selected 334 SEM images of particles with distinct morphological characteristics. Based on the obtained SEM images of atmospheric particles, the particles are divided into seven categories according to their morphological characteristics: chain particles, flocculent particles, fibrous particles, spherical particles, quasi spherical particles, irregular mineral particles, and regular mineral particles. These obtained SEM images are further enhanced using corresponding image processing methods such as horizontal inversion, color balance, brightness transformation, fuzzy processing, etc. New image samples are generated and added to the dataset, presenting a dataset of SEM images for seven types of atmospheric particles, totaling 2,672 SEM images in total. The dataset can be used for identifying, classifying, segmenting research on atmospheric particles using machine learning and other related methods. Moreover, it can provide fundamental data for other related research on atmospheric particles.

Keywords: atmospheric particulate matter; SEM images; morphological characteristics; Qingdao

Dataset Profile

Title	A dataset of SEM images of atmospheric particles in Qingdao
Data corresponding author	CHENG Xuezheng (chengxuezheng@sdust.edu.cn)
Data author(s)	YIN Changchang, ZHAO Meng, WANG Xiaohan, CHENG Xuezheng
Time range	2016–2018
Geographical scope	Qingdao (35°93′–36°07′N , 120°12′–120°25′E), including Qingdao Port, Golden Beach, Sino-German Ecopark, Hengyuan Thermal Power Co., Ltd., Qindao Zhiyan, Xijiekou North, Shandong University of Science and Technology
Spatial resolution	1536×1096 pixels
Data volume	967.5 MB
Data format	*.jpg, *.xml
Data service system	< http://dx.doi.org/10.11922/sciencedb.01170 >
Source of funding	National Natural Science Foundation of China (Grant No. 62073198)
Dataset composition	This dataset includes 2,672 SEM images of atmospheric particles categorized into seven

	categories, with a total data volume of 966 MB. Each XML file corresponding to each SEM image contains the location information and corresponding particle types of particles in the SEM image, with a total data volume of 1.5 MB.
--	---