



文献 CSTR:

32001.14.11-6035.nasdc.2023.0015.zh



文献 DOI:

10.11922/11-6035.nasdc.2023.0015.zh

数据 DOI:

10.57760/sciencedb.07727

文献分类: 农学

收稿日期: 2023-02-24

开放同评: 2023-02-24

录用日期: 2023-03-30

发表日期: 2024-11-12

2020–2021 年新疆天然草地深度学习识别图像集

李伟¹, 李全胜¹, 孔繁涛², 曹姗姗^{3,4*}, 孙伟^{3,4*}

1. 新疆农业大学计算机与信息工程学院, 乌鲁木齐 830052

2. 中国农业科学院特产研究所, 长春 130112

3. 中国农业科学院农业信息研究所, 北京 100081

4. 国家农业科学数据中心, 北京 100081

摘要: 荒漠草地是天然草地的一种, 新疆是国内荒漠草地分布最广的地区。在该地区采集到的植物图像数据集可使专家学者直观了解当地植物状况, 也可以作为物种鉴定、植物分类等相关科学研究的依据。2020–2021 年对新疆北疆地区天然草地植物进行野外调查、图像采集, 将野外采集的植物照片进行整理和分类, 得到实地拍摄植物图像 2369 张, 涉及木地肤、沙生针茅、冷蒿、北疆风毛菊、白皮锦鸡儿、骆驼蓬、麻黄、盐爪爪、马蔺、沙葱 10 种常见天然草地植物。本数据集经植物学家分类鉴定, 可作为通过采用深度学习的方法开发天然草地植物图像智能识别系统的基础数据, 为该地区植被调查与研究提供技术支持。

关键词: 天然草地植物; 深度学习

数据库(集)基本信息简介

数据库(集)名称	2020–2021 年新疆天然草地深度学习识别图像集
数据作者	孙伟
数据通信作者	孙伟 (sunwei02@caas.cn); 曹姗姗 (caoshanshan@caas.cn)
数据时间范围	2020–2021 年
地理区域	新疆北部地区
数据量	1.2 GB
数据格式	*.jpg
数据服务系统网址	https://doi.org/10.57760/sciencedb.07727
基金项目	国家自然科学基金 (31860180, 32060321, 32271880)
数据库(集)组成	数据集包含新疆北疆地区荒漠植物的图像 2369 张, 涵盖新疆北疆地区木地肤、沙生针茅、冷蒿、北疆风毛菊、白皮锦鸡儿、骆驼蓬、麻黄、盐爪爪、马蔺、沙葱 10 种荒漠植物。

引言

天然草地是地球上广泛分布的生态系统之一, 它由各种草本植物组成, 包括草、花和灌木等^[1–2]。荒漠草地是天然草地的一种, 新疆是国内荒漠草地分布最广的地区^[3]。随着人类活动的增加, 许多荒漠草地已经退化或受到破坏, 这对生态系统的平衡和生物多样性产生了负面影响。因此, 准确识别和保护荒漠草地中的植物变得越来越重要。为了更好地保护荒漠草地, 专家学者对草地进行植被调查, 调查内容包括草地植被类型、覆盖度、组成、多样性等。植物识别是植被调

* 论文通信作者

曹姗姗: caoshanshan@caas.cn

孙伟: sunwei02@caas.cn

查的重要基础，小型化、智能化的识别系统对于草地植被调查、生态保护等提供有力支持。

传统且常用的植物种类的识别方式是由经验丰富的植物专家判断，通过观察植物的外部形态、根、茎、叶等特征进行识别。识别准确度往往取决于专家的专业知识和实践经验。这导致识别过程需要耗费大量时间和精力，并且效率低下。随着数码照相机、具有高清摄像头的手机等移动终端设备走入千家万户，人们可以获取大量高清植物图像，为通过植物图像快速识别植物种类奠定了研究基础。

目前比较主流的植物图像快速识别方法是基于植物叶片特征的机器识别算法。杜吉祥等^[4]、王丽君等^[5]通过构建图像数据集，先计算并提取植物图像的特征，再将这些特征作为不同的机器学习分类器的输入，经过分析预测后得到植物分类识别的结果。在这个过程中，植物识别研究的一个关键是图像特征提取，工作量比较大。随着计算机视觉技术的不断发展，利用深度学习技术实现自动植物识别已成为一个热门研究领域。研究人员只需要通过构建大规模的植物图像数据集，使用深度神经网络来训练分类器，从而实现高精度的植物识别。高宏元等^[6]、黄方亮等^[7]均采用该方法取得了良好的植物种类识别效果。植物图像快速识别对图像数据集的要求包括多样性、高质量、丰富的样本等。已发布的数据集包括 Flavia Dataset、Leafsnap Dataset 等，提供了大量的植物图像用于科学研究。然而，由于植物种类繁多，已有数据集仍无法涵盖所有植物。

本数据集基于 2020 - 2021 年对新疆北疆地区天然草地植物野外调查结果构建，弥补了现有数据集的不足。本数据集中包含植物图像 2369 张，涉及木地肤、沙生针茅、冷蒿、北疆风毛菊、白皮锦鸡儿、骆驼蓬、麻黄、盐爪爪、马蔺、沙葱 10 种常见天然草地植物。图像均收集于野外环境，背景环境复杂多变，具有高度的多样性，有助于模型学习到植物在不同环境下的形态变化和特征表现，从而提高模型的泛化能力。

1 数据采集和处理方法

2020 年 5 月至 2021 年 9 月，调查团队在新疆北疆地区的天然草地进行荒漠草地植物样本采集。调查队员通常用数码相机对植物进行多点位的拍照，既能展示植物的整体形态，又能包含茎、叶片等局部特征。拍摄过程尽量做到对于同一物种植物要从不同的拍摄距离、不同的拍摄角度进行拍摄，图像主体要突出、清晰。野外调查返回后，将野外采集到的图像数据集交由植物专家检查判别，人工记录植物图像数据集的种类。最后由专业技术人员将野外采集的植物照片进行整理和分类，记录植物物种名称。

2 数据样本描述

实地拍摄植物图像 2369 张，图像采集设备为 Canon EOS 90D 型数码单反相机，配备佳能 EF-S 18-200mmf/3.5-5.6 IS 镜头。将采集得到的新疆北部地区天然草地的植物图像数据集按照物种名称进行分类，可分为 10 种不同的天然草地植物物种。每个物种一个压缩包文件，共有 10 个压缩包文件。每个压缩包中的每张植物图像表示一个数据样本。数据集中的部分植物样本图像如图 1 所示。



(a) 麻黄 *Ephedra sinica* Stapf (b) 白皮锦鸡儿 *Caragana leucophloea* Pojark

图 1 天然草地植物图像数据集中的样本示例

Figure 1 Sample examples in the dataset of natural grassland plants

不同植物图像的数量统计如表 1 所示。

表 1 天然草地植物样本数据分布

Table 1 Distribution of plant samples in natural grassland plants

序号	中文名称	学名	数量
1	木地肤	<i>Kochia prostrata</i> (L.) Schrad	201
2	沙生针茅	<i>Stipa glareosa</i> P.A. Smirn	213
3	冷蒿	<i>Artemisia frigida</i> Willd	261
4	北疆风毛菊	<i>Saussurea xinjiangensis</i> Y. S. Chen	253
5	白皮锦鸡儿	<i>Caragana leucophloea</i> Pojark	247
6	骆驼蓬	<i>Peganum harmala</i> L	226
7	麻黄	<i>Ephedra sinica</i> Stapf	269
8	盐爪爪	<i>Kalidium foliatum</i> (Pall.) Moq	264
9	马蔺	<i>Iris lactea</i> var. <i>chinensis</i>	213
10	沙葱	<i>Allium mongolicum</i> Regel	222

3 数据质量控制和评估

为了得到高质量的野外植物照片数据集，需要采取以下几种措施：（1）在采集植物图片数据的同时，对采集植物图片数据记录物种名称，确保种类名称有据可查。对于物种名称不明确的可后期由植物专家商讨确定。（2）对采集的植物图像数据进行清洗，删除模糊、重复、不清晰等质量较差的数据，以提高数据集的质量。（3）对于野外植物照片数据集，按照《新疆植物志》^[8]、《中国沙漠植物志》^[9]、《中国植物志》^[10]等文献进行鉴定，以确保数据集的质量和准确性。

4 数据价值

本数据集由野外天然草地环境采集得到，共采集了 10 种常见的荒漠草地植物图像，有助于专家学者直观认识和理解天然草地植物的性状，也可以作为物种鉴定、植物分类等科学研究的重要资料，具有较高的实用价值。后期将会继续完善荒漠植物图像数据集，并增加更多种类的植物图像。

此外，本数据集经植物学家分类鉴定，可作为计算机视觉技术研究的基础数据，通过采用深度学习的方法开发天然草地植物图像智能识别系统。

作者分工职责

李伟（1986—），男，重庆潼南人，硕士，讲师，研究方向农业信息化。主要承担工作：数据汇总整理，论文撰写。

李全胜（1979—），男，新疆库尔勒人，硕士，副教授，研究方向为农业信息化。主要承担工作：数据汇总整理。

孔繁涛（1968—），男，山东滕州人，博士，研究员，研究方向为农业信息技术。主要承担工作：数据校对。

曹姗姗（1984—），女，黑龙江哈尔滨人，博士，副研究员，研究方向为农林时空信息智能分析。主要承担工作：总体方案设计与组织实施。

孙伟（1978—），男，山东海阳人，博士，副研究员，研究方向为农林时空信息智能分析。主要承担工作：组织实施、论文指导与修改。

参考文献

- [1] 范月君, 侯向阳, 石红霄, 等. 气候变暖对草地生态系统碳循环的影响[J]. 草业学报, 2012, 21(03): 294-302. [FAN Y J, HOU X Y, SHI H X, et al. Effects of global warming on carbon cycling in grassland ecosystems [J]. Acta Prataculturae Sinica, 2012, 21(03): 294-302.]
- [2] 杨红飞, 刚成诚, 穆少杰, 等. 近 10 年新疆草地生态系统净初级生产力及其时空格局变化研究[J]. 草业学报, 2014, 23(03): 39-50. [YANG H F, GANG C C, MU S J, et al. Spatio-temporal pattern of net primary productivity of grassland ecosystem in Xinjiang over the past decade [J]. Acta Prataculturae Sinica, 2014, 23(03): 39-50.]
- [3] 许鹏. 新疆草地资源及其利用[M]. 乌鲁木齐: 新疆科技卫生出版社 1993:167-168. [XU P. Grassland Resources and their Utilization in Xinjiang [M]. Urumqi: Xinjiang Science and Technology Health Publishing House, 1993: 167-168.]
- [4] 杜吉祥, 汪增福. 基于径向基概率神经网络的植物叶片自动识别方法[J]. 模式识别与人工智能, 2008, 21(02): 206-213. [DU J X, WANG Z F. Automatic recognition method of plant leaves based on radial basis probabilistic neural network[J]. Pattern Recognition and Artificial Intelligence, 2008, 21(02): 206-213.]
- [5] 王丽君, 淮永建, 彭月橙. 基于叶片图像多特征融合的观叶植物种类识别[J]. 北京林业大学学报, 2015, 37(01): 55-61. [WANG L J, HUAI Y J, PENG Y C. Observation-based plant species recognition using multi-feature fusion of leaf images[J]. Journal of Beijing Forestry University, 2015, 37(01): 55-61.]
- [6] 高宏元, 高新华, 冯琦胜, 等. 基于深度学习的天然草地植物物种识别方法[J]. 草业科学, 2020, 37(09): 1931-1939. [GAO H Y, GAO X H, FENG Q S, et al. Plant species recognition method for

- natural grasslands based on deep learning[J]. Pratacultural Science, 2020, 37(09): 1931-1939.]
- [7] 黄方亮, 俞磊, 沈同平, 等. 基于 AlexNet 深度学习模型的中草药植物图像分类研究与实现[J]. 齐鲁工业大学学报, 2020, 34(02): 44-49. [HUANG F L, YU L, SHEN T P, et al. Research and implementation of Chinese herbal medicine plant image classification based on Alex Net deep learning model [J]. Journal of Qilu University of Technology, 2020, 34(2) : 44-49.]
- [8] 新疆植物志编辑委员会. 新疆植物志, 第二卷, 第一分册[M]. 乌鲁木齐: 新疆科技卫生出版社, 1994.84-106. [Xinjiang Flora Editorial Committee. Flora of Xinjiang, Vol. 2, Subvol. 1 [M]. Urumqi: Xinjiang Science and Technology Health Publishing House, 1994.84-106.]
- [9] 刘嫫心, 杨喜林, 姚育英. 中国沙漠植物志[M]. 北京: 科学出版社, 1987.[LIU Y X, YANG X L, YAO Y Y. Chinese Desert Flora [M]. Beijing: Science Press, 1987.]
- [10] 中国科学院中国植物志编辑委员会. 中国植物志 (第 25 卷, 第 2 分册) [M]. 北京: 科学出版社, 1979. [Editorial Committee of Chinese Flora, Chinese Academy of Sciences. Flora of China (Vol. 25, Vol. 2) [M]. Beijing: Science Press, 1979.]

论文引用格式

李伟, 李全胜, 孔繁涛, 等. 2020–2021 年新疆北疆地区主要荒漠植物调查数据集[J/OL]. 中国科学数据, 2024, 9(4). (2024-11-12). DOI: 10.11922/11-6035.nasdc.2023.0015.zh.

数据引用格式

孙伟. 2020–2021 年新疆天然草地深度学习识别图像集[DS/OL]. V1. Science Data Bank, 2023[2023-04-06]. <https://doi.org/10.57760/sciencedb.07727>. DOI:10.57760/sciencedb.07727.

A dataset of natural grassland plants for deep learning image recognition in Xinjiang from 2020 to 2021

LI Wei¹, LI Quansheng¹, KONG Fantao², CAO Shanshan^{3,4*}, SUN Wei^{3,4*}

1. College of Computer and Information Engineering, Xinjiang Agricultural University, Urumqi 830052, P. R. China

2. Institute of Special Animal and Plant Sciences of CAAS, Changchun 130112, P.R. China

3. Agricultural Information Institute of CAAS, Beijing 100081, P.R.China

4. National Agriculture Science Data Center, Beijing 100081, P.R.China

* Email: sunwei02@caas.cn (SUN Wei); caoshanshan@caas.cn (CAO Shanshan)

Abstract: Desert grassland is a type of natural grassland, with Xinjiang boasting the widest distribution of this ecosystem in China. The plant image dataset collected in this region can be used as the most visually informative resource for experts and scholars to understand the plants in this area, as well as the basis for

species identification, plant classification, and scientific research efforts. From 2020 to 2021, we conducted a field survey and image collection of natural grassland plants in Northern Xinjiang, and meticulously sorted and classified the collected plant photos. A total of 2,369 field-shot plant images were obtained, featuring 10 common natural grassland plants like *Artemisia frigida* and *Stipa capillata*. This dataset has been rigorously classified and identified by botanists. It can serve as the basic data for developing an intelligent identification system for natural grassland plant images using deep learning methods, and provide technical support for vegetation investigation and research in the region.

Keywords: Natural grassland plants; deep learning

Dataset Profile

Title	A dataset of natural grassland plants for deep learning image recognition in Xinjiang from 2020 to 2021
Data corresponding author	SUN Wei (sunwei02@caas.cn) ; CAO Shanshan (caoshanshan@caas.cn)
Data authors	SUN Wei
Time range	2020-2021
Geographical scope	Northern Xinjiang
Data volume	1.2 GB
Data format	*.jpg
Data service system	< http://dx.doi.org/10.57760/sciencedb.07727 >
Sources of funding	National Natural Science Foundation of China (31860180, 32060321, 32271880).
Dataset composition	The dataset comprises 2,369 images of desert plants in Northern Xinjiang, covering 10 types of desert plants, namely <i>Kochia prostrata</i> , <i>Stipa glareosa</i> , <i>Artemisia frigida</i> Willd., <i>Saussurea xinjiangensis</i> , <i>Caragana leucophloea</i> , <i>Peganum harmala</i> , <i>Ephedra sinica</i> , <i>Kalidium foliatum</i> , <i>Iris lactea</i> var. <i>chinensis</i> , and <i>Allium mongolicum</i> Regel.