

## 中国粉蝶标本图像特征数据集

王江宁<sup>1</sup>, 韩艳<sup>1</sup>, 纪力强<sup>1\*</sup>

1. 中国科学院动物研究所, 北京 100101

**摘要:** 中国粉蝶标本图像特征数据集是在对蝴蝶标本图像自动识别的研究中产生的。本数据集收集了从457幅经过处理后的中国粉蝶标本图像中提取的3个最常用特征的数值数据, 即颜色、形状和纹理特征。本数据集的每条记录都包含了蝴蝶的分类信息、图像编号以及特征信息, 为模式识别、昆虫分类等研究提供了基础数据。本数据集是中国蝴蝶标本图像特征数据集的重要组成部分。

**关键词:** 粉蝶科; 标本图像; 图像特征; 模式识别

### 数据库(集)基本信息简介

数据库(集)名称	中国粉蝶标本图像特征数据集
数据作者	王江宁, 韩艳, 纪力强
数据通信作者	纪力强 (ji@ioz.ac.cn)
数据时间范围	1994年
地理区域	中国
数据量	1.62 MB
数据格式	*.xlsx
数据服务系统网址	<a href="http://www.sciencedb.cn/dataSet/handle/483">http://www.sciencedb.cn/dataSet/handle/483</a>
基金项目	国家自然科学基金青年基金(2016), 基于在线学习的昆虫图像识别方法的研究及其在蝴蝶识别中的应用(31501841); 国家科技部国家科技基础条件平台项目“国家基础科学数据共享服务平台”(DKA2017-12-02-20)
数据库(集)组成	本数据集收集了457幅中国粉蝶标本图像的分类信息, 从中提取的颜色、形状、纹理特征数据, 以及原始标本缩略图的索引。excel的sheet1表由4个字段组成, 其数据样本描述可以参考《中国凤蝶标本图像特征数据集》一文的数据样本描述。
关联数据集	《中国凤蝶标本图像特征数据集》, 发表于《中国科学数据》2016年第1卷第3期, 网址: <a href="http://www.csdata.org/p/34/">http://www.csdata.org/p/34/</a> 。
关联说明	数据集加工方法、质控方法的相同, 标本数据源的生物类群不同。

动物的分类系统相对于其它生命的分类系统较为复杂、种类繁多, 因此动物图像采集与其它生物图像采集相比较为困难。对于蝴蝶来说, 从标本采集到图像采集, 再至特征提取, 能够使用的数据非常少。为了便于蝴蝶模式识别研究和应用, 2016年我们发布了《中国凤蝶标本图像特征数据集》<sup>[1]</sup>。

文献 DOI:

11922/csdata.2017.15.zh

数据 DOI:

11922/sciencedb.483

文献分类: 生物学

收稿日期: 2017-09-21

开放同评: 2017-12-11

发表日期: 2018-01-26

\* 论文通信作者

纪力强: [ji@ioz.ac.cn](mailto:ji@ioz.ac.cn)



粉蝶和凤蝶是蝴蝶的两大常见类群。继收集整理《中国蝶类志》(1994 版)<sup>[2]</sup>中的凤蝶标本图片后,我们采用类似的方法<sup>[1,3]</sup>,从 457 幅经过处理后的中国粉蝶标本图像中提取颜色、形状、纹理 3 个最常用特征的数值数据,得到粉蝶图像特征数据集,依然配有准确的分类学信息、模式识别方法。

由于本数据集与《中国凤蝶标本图像特征数据集》的采集方法、结构组成、质量评估和控制、使用方法等都相同,因此这些部分请参考关联文献<sup>[1]</sup>,此处不再赘述。两者的比较如表 1 所示。两者共同构成了更完整的蝴蝶标本图像特征数据集。但是因为数据获取需要时间,而且面向的蝴蝶类群不同,因此也是相对独立的数据集,是对蝴蝶整体数据的重要补充。昆虫分类学研究人员可以将其与凤蝶标本图像特征数据集结合使用。

表 1 新旧数据集比较

比较项目	说明
数据采集和处理方法	相同
数据结构组成	涉及生物类群不同
数据质量控制和评估	相同
数据价值	与不同生物类群的应用价值有关联,因此数据价值相似但不完全相同
数据使用方法	相同

## 致 谢

感谢研究组张荣在原始图像处理中所做的工作。

## 数据作者分工职责

王江宁(1982—),男,博士,助研,研究方向:昆虫图像识别。主要承担工作:本数据集的规划、建设和维护。

韩艳(1972—),女,学士,工程师,研究方向:生物多样性信息学。主要承担工作:本数据集原始数据的采集和整理。

纪力强(1961—),男,博士,研究员,研究方向:生物多样性信息学。主要承担工作:本数据集的设计。

## 参考文献

- [1] 王江宁,韩艳,纪力强.中国凤蝶标本图像特征数据集[J].中国科学数据,2016,1(3). DOI: 10.11922/csdata.180.2015.0008.
- [2] 周尧.中国蝶类志[M].郑州:河南科学技术出版社,1994.
- [3] WANG J N, JI L Q, LIANG A P, et al. The identification of butterfly families using content-based image retrieval [J]. Biosystems Engineering, 2012, 111(1): 24–32.

## 论文引用格式

王江宁, 韩艳, 纪力强. 中国粉蝶标本图像特征数据集[J/OL]. 中国科学数据, 2018, 3(1). (2018-01-24). DOI: 10.11922/csdata.2017.15.zh.

## 数据引用格式

王江宁, 韩艳, 纪力强. 中国粉蝶标本图像特征数据集[DB/OL]. Science Data Bank, 2017. (2017-09-21). DOI: 10.11922/sciencedb.483.

# A dataset of image feature for Chinese Pieridae specimen

Wang Jiangning<sup>1</sup>, Han Yan<sup>1</sup>, Ji Liqiang\*

1. Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, P. R. China

\*Email: ji@ioz.ac.cn

**Abstract:** The dataset of image feature for Chinese Pieridae specimen is created from researches of butterfly image recognition. This dataset collects the color, texture and shape features from 457 pre-processed specimen images of Chinese Pieridae by standardized feature extraction methods. Each record contains the classification information, image identifier, feature ID and feature value of an image. The dataset could support research on pattern recognition and entomology, and supplement the dataset of image feature for Chinese butterflies specimens.

**Keywords:** Pieridae; specimen image; image feature; pattern recognition

### Dataset Profile

<b>Title</b>	A dataset of image feature for Chinese Pieridae specimen
<b>Data authors</b>	Wang Jiangning, Han Yan, Ji Liqiang
<b>Data Corresponding author</b>	Ji Liqiang (ji@ioz.ac.cn)
<b>Time range</b>	1994
<b>Geographical scope</b>	China
<b>Data volume</b>	1.62 MB
<b>Data format</b>	*.xlsx
<b>Data service system</b>	<a href="http://www.sciencedb.cn/dataSet/handle/483">http://www.sciencedb.cn/dataSet/handle/483</a>
<b>Sources of funding</b>	“Insect Image Recognition Based on Online Learning and Its Application in Butterfly Recognition (Grant No.31501841), National Natural Science Youth Fund (2016); “Fundamental Science Data Sharing Platform” (DKA2017-12-02-20), National R&D Infrastructure and Facility Development Program of China

	(2017)
<b>Dataset composition</b>	This dataset collected feature data on 390 Chinese Papilionidae specimen images including the color, shape, texture and index of original specimen images. "Sheet1" of the Excel file consists of four data fields. For sample description of each field, readers can refer to an article published earlier, titled "A dataset of image feature for Chinese papilionidae specimen".
<b>Associated dataset</b>	"A dataset of image feature for Chinese papilionidae specimen", published in <i>China Scientific Data</i> (Vol.1, No.3, 2017), available at: < <a href="http://www.csdata.org/en/p/34/">http://www.csdata.org/en/p/34/</a> >.
<b>Association description</b>	This study uses the same methods for data collection, processing, quality control and assessment as the associated study, but generates different taxon of Chinese Pieridae.