

RESEARCH

Open Access



# Orchid2024: A cultivar-level dataset and methodology for fine-grained classification of Chinese *Cymbidium* Orchids

Yingshu Peng<sup>1,2</sup>, Yuxia Zhou<sup>1,2</sup>, Li Zhang<sup>1,2</sup>, Hongyan Fu<sup>1,2</sup>, Guimei Tang<sup>1,2</sup>, Guolin Huang<sup>1,2\*</sup> and Weidong Li<sup>1,2\*</sup>

## Abstract

**Background** Chinese *Cymbidium* orchids, cherished for their deep-rooted cultural significance and significant economic value in China, have spawned a rich tapestry of cultivars. However, these orchid cultivars are facing challenges from insufficient cultivation practices and antiquated techniques, including cultivar misclassification, complex identification, and the proliferation of counterfeit products. Current commercial techniques and academic research primarily emphasize species identification of orchids, rather than delving into that of orchid cultivars within species.

**Results** To bridge this gap, the authors dedicated over a year to collecting a cultivar image dataset for Chinese *Cymbidium* orchids named Orchid2024. This dataset contains over 150,000 images spanning 1,275 different categories, involving visits to 20 cities across 12 provincial administrative regions in China to gather pertinent data. Subsequently, we introduced various visual parameter-efficient fine-tuning (PEFT) methods to expedite model development, achieving the highest top-1 accuracy of 86.14% and top-5 accuracy of 95.44%.

**Conclusion** Experimental results demonstrate the complexity of the dataset while highlighting the considerable promise of PEFT methods within flower image classification. We believe that our work not only provides a practical tool for orchid researchers, growers and market participants, but also provides a unique and valuable resource for further exploring fine-grained image classification tasks. The dataset and code are available at <https://github.com/pengyingshu/Orchid2024>.

**Keywords** Chinese *Cymbidium* orchids, Orchid image dataset, Fine-grained image classification, Visual parameter-efficient fine-tuning

\*Correspondence:

Guolin Huang  
huangguolin2002@126.com

Weidong Li  
liweidong86@126.com

<sup>1</sup>Horticultural Research Institute, Hunan Academy of Agricultural Sciences, Changsha 410125, PR China

<sup>2</sup>Yuelu Mountain Laboratory, Changsha, PR China



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

## Background

Chinese *Cymbidium* orchids are a collective term for several terrestrial orchid species native to China, which, according to plant taxonomy, belong to the genus *Cymbidium* within the family *Orchidaceae* [1]. Due to Confucius' deep fondness for orchids and his unparalleled influence, along with the numerous poems related to Chinese *Cymbidium* orchids that have sprung up throughout history, as well as the importance of these flowers in terms of their beautiful appearance, diverse applications and medicinal properties, these factors collectively contribute to the Chinese "orchid culture" being a source of immense pride and cultural identity for many people in China [2]. In this historical context, Chinese *Cymbidium* orchids have been artificially cultivated for centuries due to their exceptional ornamental and cultural values, resulting in the development of numerous cultivars [3]. Concurrently, this cultivation has led to a flourishing orchid trade. The prices of different cultivars exhibit substantial variation, with commonplace cultivars generally priced below US 1,000 dollars, whereas rare cultivars or natural mutations cultivars can fetch tens or even hundreds of thousands of dollars in the market [4].

However, the increasing popularity and commercialization of Chinese *Cymbidium* orchids has raised various issues that pose significant challenges to their long-term survival and conservation. On the one hand, long-term inadequate cultivation management and outdated technical methods, Chinese *Cymbidium* orchids are facing problems including mixed varieties and difficulties in cultivar identification [5]. These limitations restrict the development of orchid cultivation, improvement, scientific research and other related work in China, and seriously hinder the stable supply of orchids to the market. On the other hand, urbanization and changing market dynamics have fueled price speculation and rapid over-collection of many Chinese *Cymbidium* orchids, resulting in population collapse and local extinctions of many cultivars [6]. Additionally, the immense commercial worth of Chinese *Cymbidium* orchids has also sparked a rise in counterfeiting and forgery of rare cultivars, leading to frequent losses for orchid enthusiasts.

As a professional team dedicated to the comprehensive utilization of orchids, we fully recognize proactively addressing the above challenges faced by the Chinese *Cymbidium* orchids industry. Therefore, authors require an efficient, accurate and directly applicable cultivar identification system urgently to support the management, cultivation, research, marketing, and protection of Chinese *Cymbidium* orchids. Firstly, Chinese *Cymbidium* orchid has more than 1000 known cultivars, each exhibiting subtle discrepancies in characteristics such as flower shape, color, and patterns. Manual identification by professionals can be time-consuming and prone to

inaccuracies. Hence, an automated system can streamline the identification process, guaranteeing both accuracy and efficiency. Secondly, accurate identification and cataloging of endangered cultivars empower researchers to track population trends, monitor distribution patterns, and formulate effective conservation strategies. Additionally, this system can facilitate the discovery of new cultivars. Lastly, a comprehensive cultivar recognition system can assist retailers and wholesalers in labeling and categorizing orchids appropriately, thereby ensuring transparency and fostering trust within the market.

The advent of deep learning technology has brought significant improvements in the accuracy of image classification tasks, including fine-grained image classification [7]. However, to the author's knowledge, there are no cases involving the cultivar classification of Chinese *Cymbidium* orchids in either commercial applications or scientific research fields. The few studies on orchid identification focus solely on the image classification of orchid species, without delving into the specialized classification of cultivars [8].

To this end, the authors have meticulously developed a Chinese *Cymbidium* cultivars classification system from scratch by utilizing the recent advances in the field of deep learning. Specifically, the authors spent more than a year creating a new high-quality fine-grained image classification dataset, Orchid2024, with an emphasis on cultivars classification of Chinese *Cymbidium* orchids. The Orchid2024 dataset contains 1,269 cultivars from 8 species of Chinese *Cymbidium* orchids and 6 additional categories, comprising a total of 156,630 images, with each image featuring at least one flowering orchid. The dataset covers nearly all common Chinese *Cymbidium* cultivars currently found in China. Each cultivar contains at least 10 images, up to a maximum of 1,387 images, essentially in proportion to their prevalence in China.

Considering the similarity of images across different orchid cultivars, we explored a different route in developing an efficient classification model, deviating from the methodologies employed in previous research on flower classification models. In relevant research, it was common practice to meticulously design a model specifically for a given dataset [9], or fine-tune existing models on the target dataset [10], or utilize neural architecture search (NAS) techniques to automatically design models for specific tasks [11]. While these methods have all been found to be effective, their practical application is often limited due to inherent limitations, particularly when employing particularly large model architectures [12]. Designing a suitable model for a dataset manually demands significant time and effort [13], and its applicability may diminish as the dataset undergoes changes. Fine-tuning existing models on the target dataset, whose performance heavily relies on the quality and relevance

of the pre-trained model to the target dataset [14]. NAS techniques are computationally expensive and time-consuming since they involve searching through a large space of possible architecture [15].

In contrast to these traditional approaches, the remarkable success of Transformer architecture and PEFT technology in natural language processing (NLP) has spurred their adoption in computer vision. The Vision Transformer (ViT) model [16] and Swin Transformer (Swin) model [17] are two notable examples that have effectively utilized the Transformer architecture, surpassing different convolution neural network (CNN) models [18] in various computer vision tasks. At the same time, researchers in computer vision have drawn inspiration from the remarkable success of PEFT technology in the field of NLP, resulting in a plethora of noteworthy accomplishments in incorporating PEFT into the realm of computer vision [19].

Therefore, this paper introduces visual parameter-efficient fine-tuning (PEFT) for the orchid cultivar accurately classification. Inspired by recent advancements in language models, PEFT offers a powerful alternative to full-parameter fine-tuning for adapting pre-trained vision models, particularly in resource-constrained scenarios [19, 20]. By selectively training a narrow subset of model parameters while freezing the rest, PEFT reduces storage requirements, accelerates training, and mitigates overfitting on limited datasets, often achieving comparable or even exceeding the performance of full-parameter fine-tuning [21, 22]. To thoroughly evaluate PEFT's effectiveness, we performed comprehensive experiments on the Orchid2024 dataset. Our analysis involved comparing the performance of various PEFT methods against full-parameter fine-tuning and established baseline performance for existing state-of-the-art classification models. The results reveal the challenge presented by the Orchid2024 dataset and unveil the potential of PEFT for further exploration in this domain.

In summary, constructing a comprehensive Chinese *Cymbidium* orchid image dataset and classification system can serve as a valuable tool for the management of orchid germplasm resources. This will provide researchers in horticulture and botany with a valuable resource for studying orchid morphology, physiology, and ecology, while also aiding in the identification, cultivation, preservation, and marketing of Chinese *Cymbidium* orchids.

## Dataset and methods

We have built the Orchid2024 dataset through the following four stages: (1) develop general principles for constructing the dataset, (2) images collection, (3) preliminary pre-processing, and (4) professional data annotation.

## The proposed Orchid2024 dataset

### General principles for building the dataset

**Naming conventions** Until the present day, the classification of plants, fungi, and animals in scientific research and education across the globe has consistently adhered to the fundamental principles (Linnaean system) established by Linnaeus [23]. The Linnaean system obeys two fundamental rules, the binomial as basic format for species names, including a genus-level name and a specific epithet, and rank-based higher classifications, with the main ranks encompassing kingdom, phylum, class, order, family, genus, species. Chinese *Cymbidium* orchids are a generic term for 8 terrestrial species that originated in China, which belong to the genus *Cymbidium* within the family *Orchidaceae*. These 8 species botanically include *Cymbidium goeringii*, *C. faberi*, *C. ensifolium*, *C. sinense*, *C. kanran*, and three variants of *Cymbidium goeringii*, namely, *C. var. longibracteatum*, *C. var. tortisepalum*, and *C. var. serratum*. In botanical terms, a variant refers to a distinct form or variety of a particular plant species that exhibits specific characteristics or traits to differentiate them from the standard form of the species [24].

Our dataset focuses on different cultivars within these 8 species. Wherein, cultivars are varieties or cultivated forms of a particular plant species. They are a subset of the species that have been specifically bred or selected for desirable traits by humans, such as unique flower colors, or strong disease resistance. In the present study, the authors followed the general principle of Linnaean classification that cultivars are given names by their Chinese phonetic(pinyin), indicated by single quotes (") following the genus and species names. For example, *Cymbidium goeringii* 'Cai Yun', where *Cymbidium* is a genus of family *Orchidaceae*, *goeringii* is a species of genus *Cymbidium*, 'Cai Yun' is a cultivar of specie *goeringii*.

**Classification criteria** Typically, botanists or orchid experts classify orchid species and cultivars by carefully observing various characteristics, including the shape, color, seeds, and roots of the plants [25]. However, when it comes to different cultivars under the same Chinese *Cymbidium* orchids, the organs such as leaves, stems, and roots appear strikingly similar. This similarity arises from the fact that the primary objective of cultivating these orchids is to appreciate their exquisite flowers. Hence, in our work, we categorize orchids based on the appearance of their blooming flowers, as these floral characteristics often serve as the most prominent differentiating factor among various orchid species and cultivars.

**Additional categories setting** In order to enhance the model's classification accuracy, broaden its application range, and meet the practical demand of recogniz-

ing unfamiliar species or flowers that are not related to Chinese *Cymbidium* orchids, we have introduced six additional categories into our dataset. These categories include images of orchids outside the genus *Cymbidium*, which belong to the family *Orchidaceae*, as well as flower images from outside the *Orchidaceae* family. Additionally, the categories contain images of four other species within the *Cymbidium* genus, namely *Cymbidium lancifolium*, *C.* hybrid, *C. floribundum*, and *C. szechuanicum*.

As far as our current understanding extends, there exists no mature and large-scale cultivar system for these four species within the *Cymbidium* genus. For a long time, horticulturists and some botanists have categorized these species as Chinese *Cymbidium* orchids. However, mainstream botanists consider them to belong to a distinct species separate from Chinese *Cymbidium* orchids [26]. In this study, we have adopted the viewpoint of mainstream botanists.

**Images collection and pre-processing**

Between October 2022 and January 2024, we systematically collected images of Chinese *Cymbidium* orchids in accordance with the blooming periods of different species. Considering the complexity and diversity of these orchids, the authors collected relevant images through various methods, as summarized below (Fig. 1).

1. A large number of images were collected from the Hunan Academy of Agricultural Sciences in Changsha, China. This organization boasts numerous professional *Cymbidium* orchid cultivation bases and greenhouses featuring standardized planting patterns, as well as the Chinese National *Cymbidium faberi* Germplasm *Cymbidium* Resource Center managed by the author’s team.
2. By conducting scientific surveys of orchids and regularly attending orchid exhibitions and academic conferences, we are able to capture close-up photos of different orchid species, especially those that are rare or endangered. During the collection of orchid pictures, the authors traveled to 20 cities in 12 provincial administrative regions in China, including Hunan, Jiangsu, Zhejiang, Guangdong, Fujian, Yunnan, Sichuan, Shaanxi, Shandong, Guizhou, Hubei and Shanghai.
3. The author’s institution, which houses a national entrepreneurship training base for technical envoys, provides technical support and cultivar identification to many companies and flower growers through its platform. This platform allows researchers to collect image data directly from planting sites.
4. This work extensively searches online platforms, botanical websites, and horticultural forums to amass a substantial collection of images, including critically endangered or extinct cultivars.



**Fig. 1** The source of image acquisition, photographed in 2023

Through the above steps, we initially collected more than 500,000 images, which required a substantial amount of manual image annotation work. For the purpose of improving annotation efficiency, we utilized computer programs to preprocess these images. The preprocessing process includes the following steps in chronological order: (1) Employing a no-reference image quality assessment model proposed by [27] to filter out low-quality images, including the removal of severely blurred or distorted images. (2) Removing duplicate images from all images by the perceptual hashing algorithm [28]. (3) We utilized the existing data within our team to build a binary classification model, which is capable of accurately distinguishing whether an image belongs to Chinese *Cymbidium* orchids. Through preliminary classification, we were able to significantly enhance data purity and reduce the workload of annotation. The binary classification model was trained by fine-tuning the ViT/B-16 model on our proprietary dataset consisting of 30,000 Chinese *Cymbidium* orchid images and 30,000 images of other flowers. After data processing, we end up with a preliminary image dataset with a hierarchical structure.

#### **Professional data annotation**

Chinese *Cymbidium* orchids involve 8 different species. For image annotations of cultivars under each species, we hired a corresponding senior grower with extensive experience in the cultivation of that species. Therefore, in total, we invited 8 senior growers to annotate the images filtered in the previous procedure. During the annotation process, for each common orchid cultivar, the authors provided a corresponding example image for these growers. Then the grower compared other orchid images from the same cultivar to determine if they share similar characteristics with the reference picture. These characteristics include petals, sepals, and color patterns. These images could only be identified as belonging to the same cultivars if the grower confirms that they have similar characteristics to the example image. Any images dissimilar to the example image are deleted directly. For the uncommon cultivars, we provided text description information related to taking or obtaining images of the current cultivar to help the growers check the cultivar annotation. This information includes flower labels at exhibitions or descriptive text accompanying the pictures.

After completing the image annotation work, the author team conducted three rounds of quality evaluation on all images. In the first round, the author team selected five orchid experts with more than three years of research experience to quickly browse through all images to screen out clearly unqualified ones. In the second round, five experts in the previous round used their professional knowledge and experience to scrutinize

each image in the dataset to confirm whether it meets the annotation requirements, and correct possible errors or omissions. The final round of review involves the authors carefully examining each image to ensure that the annotation quality of the images meets the expected standards.

#### **Dataset structure**

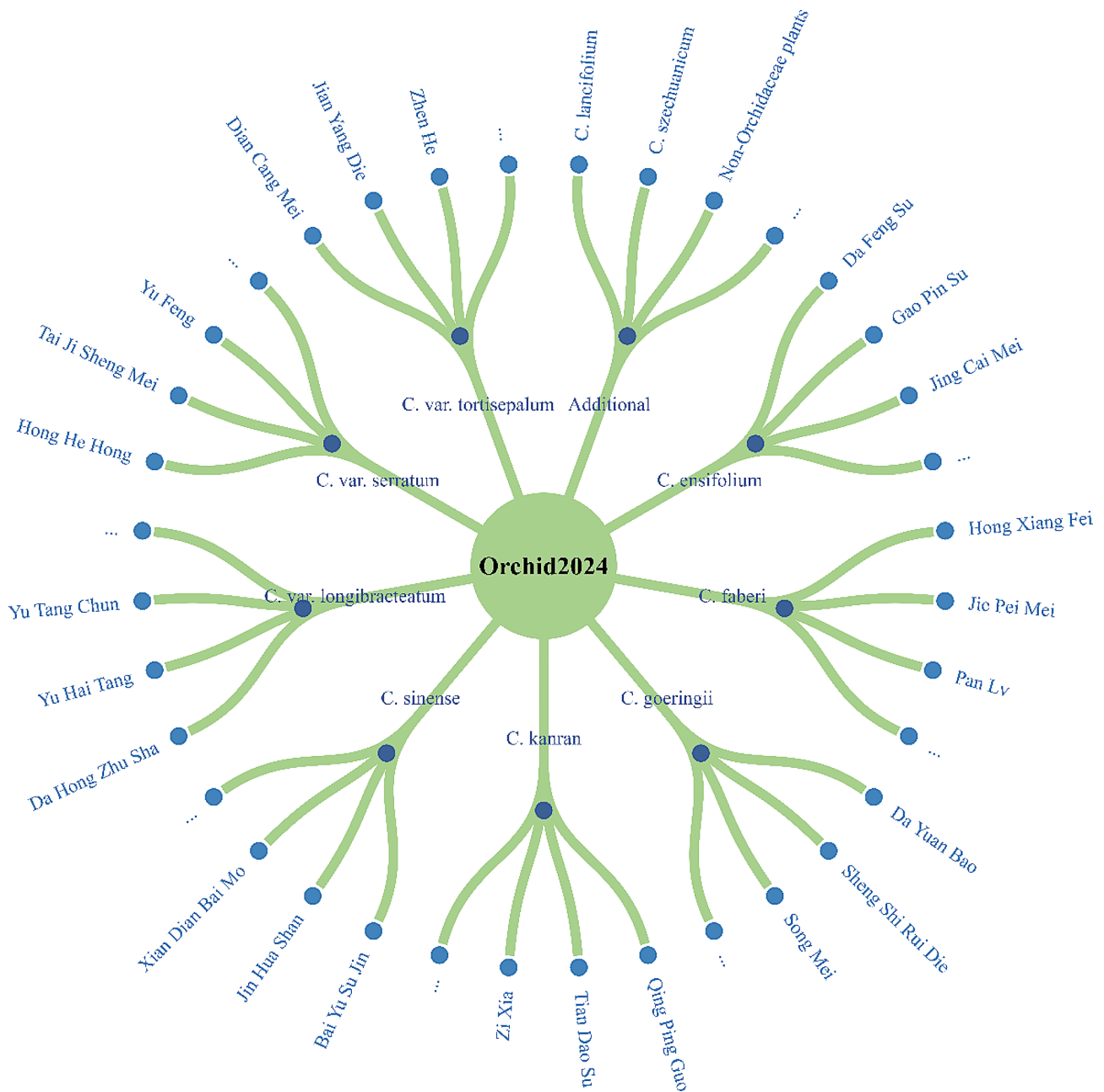
The Orchid2024 dataset follows a hierarchical structure based on the botanical classification system. Figure 2 shows its detailed taxonomy structure. After data cleaning and annotation, the Orchid2024 dataset contains 156,630 images and 1,275 classes. The dataset employs a fine-grained hierarchical structure with two distinct levels. The coarse-grained level corresponds to Chinese *Cymbidium* orchid species, containing 8 species and an additional class. The fine-grained level, corresponds to Chinese *Cymbidium* orchid cultivars, containing 1,269 cultivars and 6 additional subclasses.

The 6 additional sub-classes are part of the additional class, including orchid images from genera other than *Cymbidium* in the family *Orchidaceae*, flower images from families outside the family *Orchidaceae*, and images of 4 other species within the genus *Cymbidium*.

To construct a reliable and accurate model, we divided the dataset into training, validation, and testing sets using a 6:2:2 split at the sub-class level. More specifically, we allocated 94,036 images for training, 31,297 images for validation, and 31,297 images for testing in the classification task. Table 1 shows the number of training/validation/test (denoted as training/validation/test) set images, number of classes and imbalance ratio (IR) of the Orchid2024 dataset at different class levels. In classification tasks, IR quantifies the skew between classes in a dataset [29]. It's calculated by dividing the number of instances in the majority class by the number in the minority class. A higher IR indicates a more imbalanced dataset, where the majority class significantly outnumbers the minority class. The Orchid2024 dataset employs cultivar-level labels as the foundation for image classification tasks.

#### **Dataset characteristics**

**High diversity** Our dataset is characterized by the inclusion of image data from different species and cultivars with varying magnitudes of difference between these species, as well as varying magnitudes of difference between images of different cultivars under the same species (Fig. 3). For instance, the interspecific distinctions between the specie *C. goeringii* and *C. ensifolium* are quite pronounced, whereas the disparities between *C. goeringii* and the specie *C. var. longibracteatum* or *C. var. tortisepalum* are comparatively subtle. This is because *C.*



**Fig. 2** Taxonomy of the Orchid2024 dataset

var. longibracteatum and *C. var. tortisepalum* are variants of species *C. goeringii*.

Here, working with such a dataset also poses challenges for building a classification model. The varying degrees of differences between species and cultivars add complexity to the task of training deep learning models. Models need to effectively handle the variations in appearance, size, shape, and other visual attributes across different species and cultivars.

**Imbalanced distribution** As illustrated in Fig. 4 and detailed in Table 1, our proposed dataset demonstrates a noticeable imbalance in distribution at both the specie and cultivar levels. This disparity is predominantly attributed to factors such as market demand, breeding complexities, culminating in a long-tail distribution. The quantity of images per category spans from a minimum of 10 to a maximum of 1,387. Despite our efforts to balance the image count across various cultivars, certain cultivars face endangerment or are exceedingly rare, posing challenges in acquiring a sufficient number of images for these

**Table 1** Detailed structure of the Orchid2024 dataset

Super-Class	Class	Train	Val	Test	IR
<i>C. goeringii</i>	626	39,023	12,991	12,991	109.4
<i>C. faberi</i>	74	7,403	2,463	2,463	63.1
<i>C. ensifolium</i>	254	26,570	8,850	8,850	138.7
<i>C. sinense</i>	49	2,692	896	896	36.8
<i>C. kanran</i>	96	2,904	962	962	50.2
<i>C. var. longibracteatum</i>	72	5,359	1,782	1,782	75.1
<i>C. var. tortisepalum</i>	86	6,897	2,293	2,293	55.2
<i>C. var. serratum</i>	12	570	187	187	33.2
Additional class	6	2,618	873	873	2.6
Orchid2024	1,275	94,036	31,297	31,297	138.7

specific cultivars. Consequently, the skewed nature of the data may lead the classification model to exhibit biases toward classes with a relatively higher quantity of training samples.

**Comparison with other datasets** In Table 2, we compare the Orchid2024 dataset with several popular fine-grained image datasets. It can be observed that currently known famous fine-grained datasets mainly focus on common object classification within the general domain, such as food, animals and daily necessities. In contrast, our dataset offers a deep dive into a specific domain, providing fine-grained detail for orchid cultivars classification. Give that orchids exhibit unique characteristics and complex morphology, their distinction and recognition require more meticulous observation and understanding. Therefore, our dataset has a greater level of granularity. Furthermore, Orchid2024 dataset is comparable to the widely used fine-grained datasets currently available in terms of both the size and diversity of the images.

## Methods

### Exploring PEFT strategies with Orchid2024 data

This study explores how to efficiently build a classification model based on the Orchid2024 dataset. In order to make the model more accurate, versatile, and less expensive to train, we adopt the research paradigm of pre-training and fine-tuning. Specifically, the commonly used Vision Transformer model was selected as the pre-trained model, and then PEFT methods were employed to train the pre-trained model in order to build an image classification model for Chinese *Cymbidium* orchids. The core of PEFT technology lies in introducing a minimal number of new trainable parameters during fine-tuning. This is achieved by keeping the pre-trained model's parameters frozen. Alternatively, a small portion of the original parameters can be fine-tuned while minimizing the introduction of new ones. Inherent redundancy within the model parameters allows this technique to effectively utilize limited data resources and reduce computational demands.

In this study, we employed the ViT/B-16 model as the pre-trained model to find an equilibrium between model complexity and experimental conditions. Based on this foundation, our research assesses the effectiveness of different PEFT methods on the Orchid2024 dataset. As illustrated in Fig. 5, some representative PEFT methods include Adapter [38], LoRA (Low-Rank Adaptation) [39], and VPT (Visual Prompt Tuning) [20]. These methods utilize a range of techniques to minimize the number of fine-tuning parameters needed, achieved by adjusting the input, backbone, or classification head of ViT, while keeping the other parameters of the pretrained model fixed (the parameters are not adjusted during the training process). A detailed examination of all PEFT methods employed in this paper is outlined below:

#### A. Full-parameter fine-tuning:

(1) Full-parameter fine-tuning: update all model parameters of the pre-trained model, including the backbone and classification heads.

B. Methods that leverage the pre-trained model for feature extraction and concentrate on tailoring the classification head for the specific task:

(2) Linear probe: use a single linear layer as the classification head. Throughout training, only update the parameters of this layer to learn the class differentiations.

(3) Partial-k [40]: update the parameters of the last k backbone layers and the linear classification header, while fixing the weights of all other layers. We set k to 1 in our experiments.

C. Methods that allow for targeted updates to a subset of backbone parameters or the incorporation of additional trainable parameters within the backbone architecture:

(4) Adapter [38]: enhance the ViT architecture by integrating additional lightweight adapter modules within each Transformer module. An adapter module linearly down-projects the data to reduce its dimensionality, then applies a nonlinear activation function to learn more complex relationships in the data, and finally linearly up-projects the data back to its original size, with a residual connection added to the output. In our work, we used an adapter compression factor (reduction factor) of 128. This factor means that the output feature dimensions of the linear projection layer within the adapter module are compressed by a factor of 128.

(5) LoRA [39]: streamlines model fine-tuning by leveraging low-rank matrix decomposition. It focuses on the attention layer, specifically the query (Q), key (K), value (V) and output projection (O) layer within a ViT model. During training, LoRA efficiently adapts the model by optimizing a smaller set of

**(a) *Cymbidium goeringii***



Song Mei      Da Yuan Bao      Sheng Shi Rui Die

**(b) *Cymbidium faberi***



Hong Xiang Fei      Jie Pei Mei      Pan Lv

**(c) *Cymbidium ensifolium***



Jing Cai Mei      Da Feng Su      Gao Pin Su

**(d) *Cymbidium sinense***



Jin Hua Shan      Bai Yu Su Jin      Xian Dian Bai Mo

**(e) *Cymbidium kanran***



Zi Xia      Tian Dao Su      Qing Ping Guo

**(f) *Cymbidium* var. *longibracteatum***



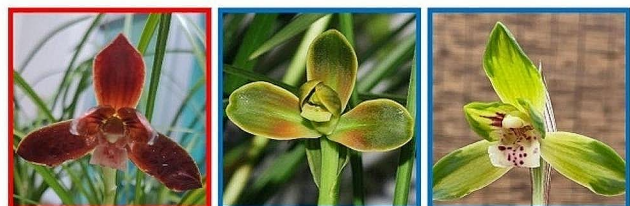
Da Hong Zhu Sha      Yu Hai Tang      Yu Tang Chun

**(g) *Cymbidium* var. *tortisepalum***



Jian Yang Die      Zhen He      Dian Cang Mei

**(h) *Cymbidium* var. *serratum***



Hong He Hong      Tai Ji Sheng Mei      Yu Feng

**Fig. 3** Examples of different species and cultivars. Due to differences in plant breeding methods and goals, each species exhibits highly similar characteristics among the cultivars denoted by the blue border, whereas the cultivar outlined in red demonstrates notable differences from those marked with the blue border

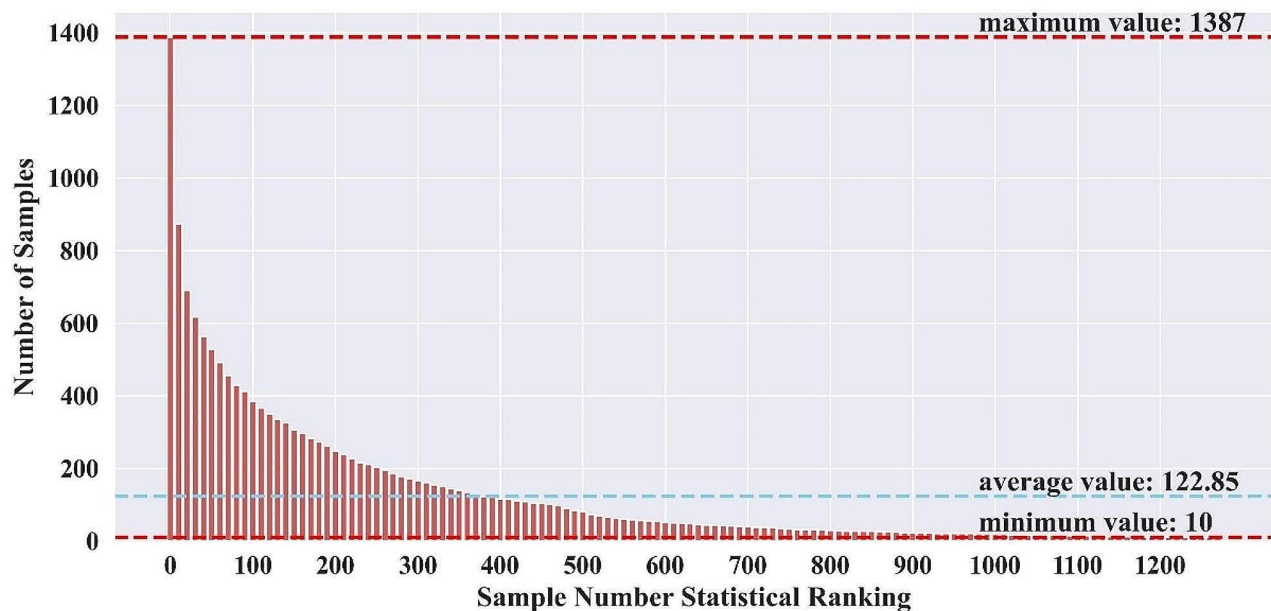
low-rank matrix parameters instead of modifying all the weight parameters directly. we focus on tuning the Q and V parameters within each attention block, as inspired in [41].

- (6) Bias tuning [42]: update only the bias terms within the pre-trained backbone and the parameters of the linear classification head.
- (7) Side tuning [43]: train a “side” network (such as AlexNet [44]) and linear interpolate between pre-trained features and side-tuned features before being fed into the classification head.

D. Methods that incorporate a limited number of trainable parameters into the input space of the ViT.

- (8) VPT-deep: transform an input image into a series of patches and consider them as tokens, add a collection of trainable parameters called prompts, which are then placed before the input space of the ViT. In VPT-deep, prompts are added to the input space of each Transformer layer. In addition, the output of the prompt from the previous Transformer layer is eliminated, and a fresh prompt is inserted





**Fig. 4** Sample number distribution of the Orchid2024 dataset. Sort according to the number of samples, and then display the data in such a way that every fifth value is sampled

**Table 2** Comparison with other datasets

Dataset	Year	Meta-Class	Images	Categories
Oxford Flowers102 [30]	2008	Flowers	8,189	102
CUB200-2011 [31]	2011	Birds	11,788	200
Stanford Dogs [32]	2011	Dogs	20,580	120
Food101 [33]	2014	Food dishes	101,000	101
DeepFashion [34]	2016	Clothes	800,000	1,050
Veg200 [35]	2017	Vegetable	91,117	200
Products-10 K [36]	2020	Retail products	150,000	10,000
iNat2021 [37]	2021	Plants & Animals	3,286,843	10,000
Orchid2024	2024	Orchid cultivars	156,630	1,275

into the current Transformer layer. The VPT method uses 10 token prompts by default in our work.

### Experiment and evaluation

Within the Orchid2024 dataset, we conduct a systematic evaluation of various PEFT methods to identify the configuration that achieves optimal classification performance. Subsequently, we delve into a rigorous analysis to understand how design choices within PEFT methods influence their overall effectiveness.

#### Experiment setup

For our experiments, the ViT/B-16 model trained on the ImageNet- 21k dataset [45] is used as the pre-trained model. Each PEFT method underwent training for a total of 100 epochs with a batch size of 32. During training, we employed an initial learning rate of 0.001 with 10 epoch

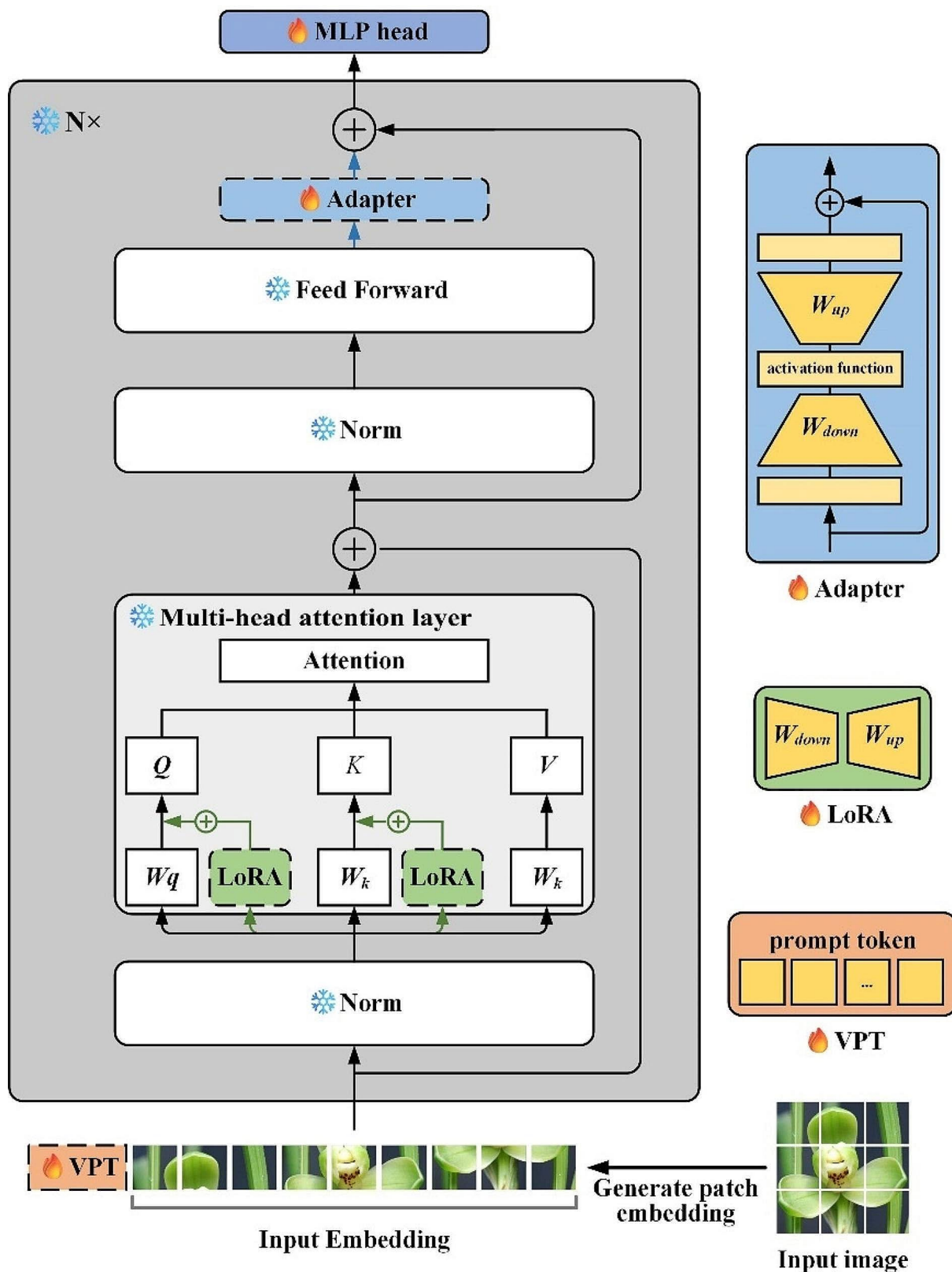
warmup to mitigate potential learning rate instability. To further prevent overfitting, weight decay of 0.00001 was applied. The model received a  $224 \times 224$  image cropped from the center of a pre-scaled  $256 \times 256$  image. For all PEFT methods, LoRA was used with settings from [41], while other methods employed settings from [20]. The experiment was conducted on 8 NVIDIA Titan X GPUs, each equipped with 12GB of onboard memory.

The Orchid2024 dataset demonstrates a significant imbalance in class distribution. To address this challenge, we adopted a range of comprehensive metrics for the classification task, including top-1 accuracy and top-5 accuracy. Top-1 accuracy reflects the percentage of samples where the predicted class label matches the true class label. In contrast, top-5 accuracy provides some flexibility in predictions by considering the 5 most probable classes, deeming a prediction correct if the true label falls within these 5 classes. Of these metrics, the authors place the highest priority on top-5 accuracy to roughly determine the range of cultivars. This decision is rooted in the exceptionally challenging nature of classifying orchid cultivars, a task that even the most proficient experts in this field struggle with, as they may not be able to accurately classify all categories.

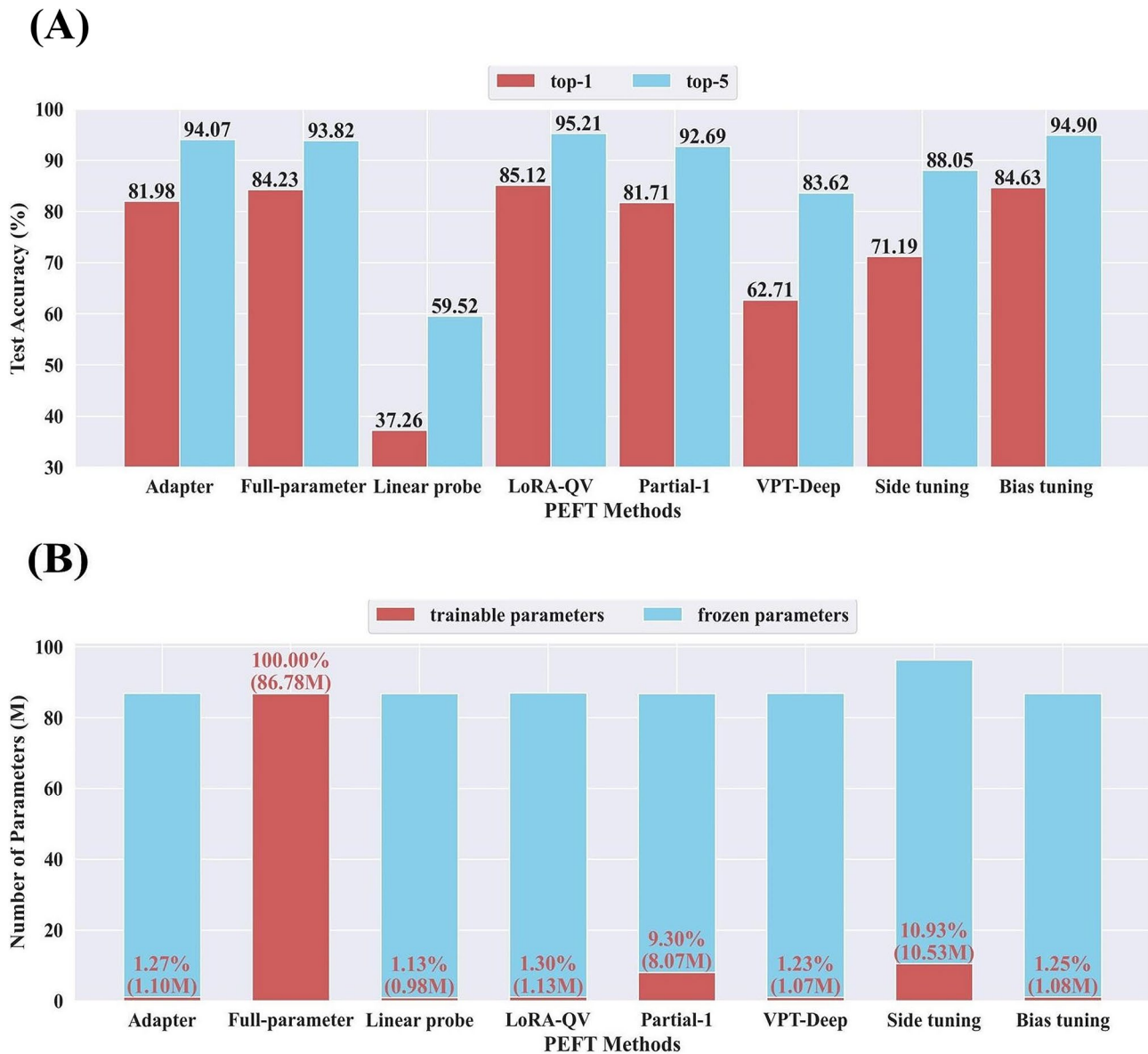
### Results

#### Performance comparison of PEFT methods

Figure 6 compares the performance of various PEFT methods using default configuration on the Orchid2024 dataset. It presents the top-1 and top-5 accuracy achieved by each method, along with the number of parameters



**Fig. 5** Common visual parameter-efficient fine-tuning methods. During model training, pre-trained modules depicted with white or gray backgrounds have their parameters frozen. In contrast, the parameters associated with colored sections, representing other methods and the classification head, are updated to adapt to the specific task



**Fig. 6** Performance comparison of the PEFT methods using default configuration on the Orchid2024 dataset. **(A)** the test accuracy of different PEFT methods on the Orchid2024 dataset, and **(B)** shows the proportion of trainable parameters of these PEFT methods

updated during training. Trainable parameters are those parts of the model updated during fine-tuning, while frozen parameters are those parts that remain unchanged. The ensuing observations gleaned from the results are as follows:

1. LoRA and Bias tuning methods surpass full-parameter fine-tuning in top-1 and top-5 test accuracy, while Adapter exhibits comparable performance. It is worth noting that these three PEFT methods achieve this level of accuracy using trainable parameters that do not exceed 1.5% of the model total parameters. This highlights the potential of PEFT technology in image classification tasks, as
2. Our findings on the Orchid2024 dataset largely corroborate those reported in [20] and [41], with the exception of VPT and Bias tuning. VPT demonstrates significantly lower performance compared to other methods, while Bias tuning stands out as an effective approach. Possible reasons for this phenomenon include the close resemblance of flowers within the Orchid2024 dataset in terms of their appearance. Additionally, the Orchid2024 dataset contains fewer data points for individual

categories and exhibits a more uneven distribution of categories. These factors pose challenges for the VPT in learning generalizable visual prompts. Consequently, VPT may struggle to capture the nuanced features of orchids. Bias tuning only adjusts the bias of the pre-trained model without adjusting other parameters, so it can effectively avoid learning general features that are irrelevant to the fine-grained features of flowers.

3. While the effectiveness of PETF methods undeniably hinges on the chosen benchmark task and dataset, their relative performance can also shift based on the specific task and data distribution. However, LoRA, a leading PETF method especially successful in NLP, achieves superior accuracy on the Orchid2024 dataset compared to other approaches. This result shows LoRA's powerful generalization ability, suggesting its effectiveness can be broadly applied across diverse tasks.

## Ablation results

### **Ablation on different PEFT methods parameter configuration**

This study investigates the performance of various configurations for PEFT methods (Table 3), including Linear probe, LoRA, Partial-1, and VPT-Deep. Specifically, we compared the performance of Linear probe with MLP-3. MLP-3 employs a multilayer perceptron (MLP) consisting of three linear layers as the classification head, in contrast to Linear probe which employs a single linear layer. Additionally, we examined the impact of LoRA on different modules of the attention layer. Partial-1 and Partial-2 entail updating the parameters of the last 1 and

**Table 3** Effect of different PEFT methods parameter configuration on model performance. "Params" refers to the total number of parameters in the model. "Trainable params" indicates the proportion of parameters that can be adjusted during training. The comparison item of each method is the default configuration of the method

Methods	Top-1 (%)	Top-5 (%)	Params (M)	Trainable Params
Linear probe	37.26	59.52	86.78	1.13%
MLP-3	75.76	91.30	87.96	2.46%
LoRA-QV	85.12	95.21	86.93	1.30%
LoRA-QVK	85.71	95.34	87.00	1.38%
LoRA-QVKO	86.14	95.44	87.07	1.46%
Partial-1	81.71	92.69	86.78	9.30%
Partial-2	84.76	94.06	86.78	17.47%
VPT-Deep-tokens1	58.53	79.57	86.79	1.14%
VPT-Deep-tokens5	61.93	82.72	86.83	1.18%
VPT-Deep-tokens10	62.71	83.62	86.87	1.23%
VPT-Deep-tokens100	65.15	85.10	87.70	2.17%
VPT-Shallow-tokens10	45.11	68.11	86.79	1.14%
Full-parameter	84.23	93.82	86.78	100.00%

2 backbone layers and the linear classification header, respectively. VPT-Deep utilizes varying prompt token lengths, whereas VPT-shallow insert prompts solely into the first Transformer layer of the ViT model, thereby adding them only at the initial input to the model.

The results indicate that augmenting the necessary trainable parameters in Linear probe, LoRA, Partial-1, and VPT-Deep leads to an enhancement in accuracy. Methods employing a higher number of parameters tend to yield superior results, primarily due to their increased capacity to comprehend the intricate correlations between input data and output labels. While Partial-2 outperforms Partial-1 and even surpasses full-parameter fine-tuning in terms of performance, it comes at a cost. Partial-2 necessitates a significantly higher number of trainable parameters (17.47%) compared to other PETF methods. Although increasing the prompt length in VPT offers some improvement, it falls short of the performance achieved through full-parameter fine-tuning. Therefore, for fine-grained flower data sets such as Orchid2024, training the model itself is necessary.

In contrast, different parameter configurations of LoRA have achieved better performance than other methods by fine-tuning the query (Q), key (K), value (V), and output (O) modules of the ViT self-attention layer. Among these configurations, the LoRA-QVKO method achieves the best performance on both metrics, with a top-1 accuracy of 86.14% and a top-5 accuracy of 95.44% (Table 3).

### **Ablation on different pre-trained models**

We enhanced the capabilities of PETF methods by leveraging a broader toolkit of pre-trained models. In addition to the pre-trained backbone trained on the large ImageNet-21k dataset, we explored the potential of masked autoencoders (MAE) [46] trained with self-supervised learning on the ImageNet-1 K dataset. All of these pre-trained models are built upon ViT/B-16 architecture. The corresponding results are presented in Table 4.

Supervised pre-training often leads to superior performance in PEFT methods. Models leveraging this approach consistently outperform those using MAE for pre-training. This is likely because MAE primarily focuses on reconstructing input images, prioritizing coarse-grained features essential for general object recognition. In contrast, fine-grained orchid datasets necessitate a deeper grasp of subtle morphological variations and intricate patterns, which MAE might not capture adequately. In many instances, the model pre-trained on the ImageNet-21k dataset tends to outperform that pre-trained on the ImageNet-1k dataset. This is largely attributed to the richer training data available in the ImageNet-21k dataset, enabling the construction of more robust pre-trained models.

**Table 4** Effect of pre-trained models in different PEFT methods on model performance. “Pre-trained Model” column delineates the training techniques and dataset employed in training the respective pre-trained model. “Params” refers to the total number of parameters in the model. “Trainable params” indicates the proportion of parameters that can be adjusted during training. All PEFT methods use the pre-trained model trained on the imagenet-21k dataset by default

Methods	Pre-trained Model	Top-1 (%)	Top-5 (%)	Params (M)	Trainable Params
Adapter	Supervised/ ImageNet-21k	81.98	94.07	86.90	1.27%
	Supervised/ ImageNet-1k	80.88	93.15	86.90	1.27%
	MAE/ImageNet-1k	77.31	91.61	86.90	1.27%
Full-parameter	Supervised/ ImageNet-21k	84.23	93.82	86.78	100.00%
	Supervised/ ImageNet-1k	84.25	93.81	86.78	100.00%
	MAE/ImageNet-1k	79.56	90.86	86.78	100.00%
Linear probe	Supervised/ ImageNet-21k	37.26	59.52	86.78	1.13%
	Supervised/ ImageNet-1k	67.07	86.48	86.78	1.13%
	MAE/ImageNet-1k	12.28	26.98	86.78	1.13%
LoRA-QV	Supervised/ ImageNet-21k	85.12	95.21	86.93	1.30%
	Supervised/ ImageNet-1k	80.61	93.34	86.93	1.30%
	MAE/ImageNet-1k	74.29	89.71	86.93	1.30%
Partial-1	Supervised/ ImageNet-21k	81.71	92.69	86.78	9.30%
	Supervised/ ImageNet-1k	79.04	91.36	86.78	9.30%
	MAE/ImageNet-1k	77.72	89.73	86.78	9.30%
VPT-Deep	Supervised/ ImageNet-21k	62.71	83.62	86.87	1.23%
	Supervised/ ImageNet-1k	81.06	94.33	86.87	1.23%
	MAE/ImageNet-1k	77.22	92.58	86.87	1.23%
Side tuning	Supervised/ ImageNet-21k	71.19	88.05	96.33	10.93%
	MAE/ImageNet-1k	71.14	88.20	96.33	10.93%
	Supervised/ ImageNet-1k	69.34	87.43	96.33	10.93%
Bias tuning	Supervised/ ImageNet-21k	84.63	94.90	86.78	1.25%
	Supervised/ ImageNet-1k	83.11	94.09	86.78	1.25%
	MAE/ImageNet-1k	81.05	93.31	86.78	1.25%

Additionally, the effectiveness of PEFT methods hinges heavily on the underlying pre-trained model’s performance. This is demonstrably true by the significant accuracy variations observed across PEFT methods when applied to different pre-trained models. Notably, LoRA achieves superior results when the pre-trained model is

**Table 5** Effect of learning rate and weight decay in different PEFT methods on model performance. The learning rate of 0.001 and weight attenuation of 0.00001 are the default values of all training methods in our work

Methods	Learning Rate	Weight Decay	Top-1(%)	Top-5(%)
Adapter	0.01	0.00001	82.50	94.09
	0.005	0.00001	82.30	93.96
	0.001	0.00001	81.98	94.07
	0.001	0.0001	81.98	94.07
Full-parameter	0.01	0.00001	40.90	70.03
	0.005	0.00001	83.93	93.59
	0.001	0.00001	84.23	93.82
	0.001	0.0001	84.18	93.99
Linear probe	0.01	0.00001	58.10	79.71
	0.005	0.00001	57.03	78.67
	0.001	0.00001	37.26	59.52
	0.001	0.0001	36.34	58.53
LoRA-QV	0.01	0.00001	77.01	93.22
	0.005	0.00001	85.30	95.12
	0.001	0.00001	85.12	95.21
	0.001	0.0001	85.12	95.21
Partial-1	0.01	0.00001	80.71	93.27
	0.005	0.00001	81.63	93.37
	0.001	0.00001	81.71	92.69
	0.001	0.0001	81.71	92.69
VPT-Deep	0.01	0.00001	79.41	94.28
	0.005	0.00001	77.95	93.14
	0.001	0.00001	62.71	83.62
	0.001	0.0001	62.29	83.15
Side tuning	0.01	0.00001	72.48	88.40
	0.005	0.00001	72.52	89.01
	0.001	0.00001	71.19	88.05
	0.001	0.0001	71.57	88.39
Bias tuning	0.01	0.00001	84.07	94.72
	0.005	0.00001	84.00	94.73
	0.001	0.00001	84.63	94.90
	0.001	0.0001	84.63	94.90

trained on a massive dataset like ImageNet-21k. However, this advantage lessens when using a smaller dataset like ImageNet-1k. Nevertheless, for building an effective model on the Orchid2024 dataset, the combination of LoRA and a pre-trained model on ImageNet-21k remains an advantageous approach.

**Ablation on initial learning rate and weight decay**

We explored the impact of different initial learning rates and weight decay on model accuracy. Overall, the results of this study suggest that there is no one-size-fits-all training method for training models with PEFT methods.

On the whole, using lower learning rates and weight decay tends to result in better performance for PEFT methods. The experimental outcomes outlined in Table 5 indicate that, given PEFT methods involve fine-tuning only specific parameters of the model, they exhibit

heightened sensitivity to learning rates, necessitating meticulous adjustments to attain optimal results. A high initial learning rate may precipitate excessive parameter updates, potentially compromising the established feature representation within the pre-trained model. Consequently, it is advisable to employ a smaller initial learning rate for PEFT methods compared to full-parameter fine-tuning. Moreover, the influence of learning rates on model performance surpasses that of weight decay.

In most cases, LoRA-QV achieves higher top-1 and top-5 accuracies than other PETF methods across various combinations of initial learning rates and weight decay. This suggests that LoRA-QV exhibits greater robustness to learning rate and weight decay hyperparameter tuning compared to other PETF methods.

## Discussion

Currently, there is no research or dataset available on the image classification of Chinese *Cymbidium* orchids and cultivars. For other species which belong to family *Orchidaceae* with Chinese *Cymbidium* orchids, some related public datasets are released. For example, [47] collected a dataset consisting of only 1,500 samples across 15 classes to classify orchid species in genus *Paphiopedilum* from Thailand. Subsequently, [8] proposed a hybrid model architecture for better image classification of orchid species based on an orchid species dataset containing 3,559 samples and 52 categories. [48] also proposed a dataset containing 7,156 orchid images, including 156 different orchid species, most of which were obtained through network search. It is evident that the current orchid species datasets solely emphasize the identification of orchid species without providing any additional categorization for distinguishing cultivars within these species. Furthermore, it is worth noting that these datasets are relatively limited in scale, as each one comprises less than 10,000 samples.

Several renowned flower image datasets have also encountered similar issues. For instance, the Oxford 102 Flower Dataset [30] and HFD100 dataset [10] present limited scale and only oriented to general domains. The Oxford 102 flower dataset is an image classification dataset consisting of 8,189 images, encompassing 102 distinct flower categories from the United Kingdom. The HFD100 dataset contains more than 10,700 hyperspectral flower images which belong to 100 categories.

In comparison, our proposed dataset covers 1,275 classes and consists of 156,630 images. It is specifically created for professional fields and provides cultivar-level labeling beneath each species, characterized by fine granularity and a long-tail distribution. To the author's knowledge, current commercial identification software focuses solely on classifying Chinese *Cymbidium* orchid species and does not include the classification of their cultivars.

In this study, we build the Orchid2024 dataset and evaluate the performance of various PEFT methods on this dataset for classifying Chinese *Cymbidium* orchids. Our findings reveal that LoRA emerges as the most effective overall PEFT method for a building classification model on the Orchid2024 dataset. Bias tuning, Partial-2, and Adapter also demonstrate strong performance, achieving comparable or even surpassing results obtained with full-parameter fine-tuning. Conversely, Linear probe, VPT-Deep, and Side tuning exhibited lower effectiveness.

Outperforming other LoRA configurations, LoRA-QKVO obtains state-of-the-art results on the Orchid2024 dataset. By integrating the LoRA module into the query (Q), key (K), value (V), and output (O) components of the ViT self-attention layer, it delivers exceptional top-1 accuracy of 86.14% and top-5 accuracy of 95.44%, all while using only 1.46% of the model parameters. This remarkably efficient approach meets our requirements for model classification accuracy on Orchid2024.

However, this study just provides a foundational understanding, and exploring more advanced PEFT techniques could significantly boost accuracy on the Orchid2024 dataset. The current choice of a ViT/B-16 pre-trained model is a starting point. There are superior pre-trained models available, with better performance and larger parameter sizes. Utilizing these models with the PEFT methods is expected to improve Orchid2024 accuracy within a reasonable range, but it will also demand greater computational resources. Furthermore, a more extensive exploration of PEFT parameter configurations and a diversification of PEFT methodologies warrant further investigation.

Despite careful collection and annotation efforts to create the Orchid2024 dataset, some known limitations exist. First, inherent similarities among many Chinese *Cymbidium* orchid cultivars can lead to mislabeling, even by experts. Secondly, the dataset might not include many rare orchid species due to their inherent scarcity and the difficulty of collection, especially for wild orchids. Additionally, other countries like South Korea and Japan have a rich history of cultivating Chinese *Cymbidium* orchids, and including cultivars from these regions would be beneficial. Finally, while various orchid cultivars can theoretically flower annually, limitations in cultivation technology and climate can cause some to bloom only every few years in specific regions. This cyclical blooming pattern might unintentionally introduce bias towards more frequently flowering cultivars within the dataset. The author team will attempt to address these limitations in the future and expand the Orchid 2024 dataset to include more diverse orchid cultivars and geographical origins.

## Conclusion

In this work, we collect a large-scale fine-grained dataset, named Orchid2024, for Chinese *Cymbidium* orchid cultivars classification. It comprises a total of 156,630 images, encompassing 1,275 distinct cultivars across 8 species. Compared to previous datasets, the Orchid2024 dataset exemplifies an elevated degree of granularity and specialization within fine-grained classification, characterized by its uniqueness and practicality. Meanwhile, we also compared the performance of various PEFT methods against full-parameter fine-tuning methods on the proposed dataset. The results demonstrate that the LoRA-QVKO method achieves the best performance with a top-1 accuracy of 86.14% and a top-5 accuracy of 95.44%. Our work provides strong support for the development and research in the field of Chinese *Cymbidium* orchid cultivation, while also offering a new data source for existing fine-grained image classification studies. The outcomes from this work are intended for direct application in germplasm resource centers, laboratories, and planting bases managed by the authors.

## Author contributions

Yingshu Peng: Conceptualization, Data collection and annotation, Methodology, Model construction and training, Data analysis, Project administration, Writing—original draft, Writing—review & editing. Yuxia Zhou: Data collection and annotation. Li Zhang: Data collection and annotation. Hongyan Fu: Data collection and annotation. Guimei Tang: Data collection and annotation. Guolin Huang: Resources, Data collection and annotation, Validation, Writing—review & editing, Supervision. Weidong Li: Data annotation, Validation, Resources, Writing—review & editing, Supervision.

## Funding

This work was supported by the National Key Research and Development Project of China (Grant No.2019YFD1100400), Hunan Provincial Science and Technology Innovation Fund (Grant No.2023CX95), Hunan Key Laboratory of Germplasm Innovation and Comprehensive Utilization of Ornamental Plant (Grant No.2022YLHH001).

## Data availability

No datasets were generated or analysed during the current study.

## Declarations

### Ethics approval and consent to participate

This work doesn't involve human participants, human data or human tissue.

### Competing interests

The authors declare no competing interests.

### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Received: 11 June 2024 / Accepted: 29 July 2024

Published online: 13 August 2024

## References

- Ning H, Ao S, Fan Y, Fu J, Xu C. Correlation analysis between the karyotypes and phenotypic traits of Chinese cymbidium cultivars. *Hortic Environ Biotechnol*. 2018;59:93–103.

- Hew CS. Ancient Chinese orchid cultivation: a fresh look at an age-old practice. *Sci Hort*. 2001;87:1–10.
- Su S, Shao X, Zhu C, Xu J, Tang Y, Luo D, et al. An AGAMOUS-like factor is associated with the origin of two domesticated varieties in *Cymbidium sinense* (Orchidaceae). *Hortic Res*. 2018;5:48.
- Seyler BC, Gaoue OG, Tang Y, Duffy DC. Understanding knowledge threatened by declining wild orchid populations in an urbanizing China (Sichuan). *Environ Conserv*. 2019;46:318–25.
- Zhu G, Yang F, Shi S, Li D, Wang Z, Liu H, et al. Transcriptome characterization of *Cymbidium sinense* Dharma using 454 pyrosequencing and its application in the identification of genes Associated with Leaf Color Variation. *PLoS ONE*. 2015;10:e0128592.
- Seyler BC, Gaoue OG, Tang Y, Duffy DC, Aba E. Collapse of orchid populations altered traditional knowledge and cultural valuation in Sichuan, China. *Anthropocene*. 2020;29:100236.
- Wei X-S, Song Y-Z, Aodha OM, Wu J, Peng Y, Tang J, et al. Fine-grained image analysis with Deep Learning: a Survey. *IEEE Trans Pattern Anal Mach Intell*. 2022;44:8927–48.
- Sarachai W, Bootkrajang J, Chaijaruwanich J, Somhom S. Orchid classification using homogeneous ensemble of small deep convolutional neural network. *Mach Vis Appl*. 2022;33:17.
- Peng Y, He X, Zhao J. Object-part attention model for fine-grained image classification. *IEEE Trans Image Process*. 2018;27:1487–500.
- Zheng Y, Zhang T, Fu Y. A large-scale hyperspectral dataset for flower classification. *Knowl Based Syst*. 2022;236:107647.
- Xue C, Wang X, Yan J, Li C-GA, Max-Flow. Based Approach for neural Architecture search. In: Avidan S, Brostow G, Cissé M, Farinella GM, Hassner T, editors. *Computer vision – ECCV 2022*. Cham: Springer Nature Switzerland; 2022. pp. 685–701.
- Dutt R, Ericsson L, Sanchez P, Tsafaris SA, Hospedales T. Parameter-Efficient Fine-Tuning for Medical Image Analysis: The Missed Opportunity [Internet]. *arXiv*; 2023 [cited 2024 Mar 13]. <http://arxiv.org/abs/2305.08252>
- Negrinho R, Gordon G, DeepArchitect. Automatically Designing and Training Deep Architectures [Internet]. *arXiv*; 2017 [cited 2024 Mar 13]. <http://arxiv.org/abs/1704.08792>
- Ridnik T, Ben-Baruch E, Noy A, Zelnik-Manor L. ImageNet-21K Pretraining for the Masses [Internet]. *arXiv*; 2021 [cited 2023 Jun 29]. <http://arxiv.org/abs/2104.10972>
- Baymurzina D, Golikov E, Burtsev M. A review of neural architecture search. *Neurocomputing*. 2022;474:82–93.
- Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *arXiv:2010.11929 [cs]* [Internet]. 2021 [cited 2022 Mar 20]; <http://arxiv.org/abs/2010.11929>
- Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z et al. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. 2021 IEEE/CVF International Conference on Computer Vision (ICCV). 2021. pp. 9992–10002.
- Zeiler MD, Fergus R. Visualizing and understanding Convolutional Networks. In: Fleet D, Pajdla T, Schiele B, Tuytelaars T, editors. *Computer vision – ECCV 2014*. Cham: Springer International Publishing; 2014. pp. 818–33.
- Xin Y, Luo S, Zhou H, Du J, Liu X, Fan Y et al. Parameter-Efficient Fine-Tuning for Pre-Trained Vision Models: A Survey [Internet]. *arXiv*; 2024 [cited 2024 Apr 15]. <http://arxiv.org/abs/2402.02242>
- Jia M, Tang L, Chen B-C, Cardie C, Belongie S, Hariharan B, et al. Visual prompt tuning. In: Avidan S, Brostow G, Cissé M, Farinella GM, Hassner T, editors. *Computer vision – ECCV 2022*. Cham: Springer Nature Switzerland; 2022. pp. 709–27.
- He H, Cai J, Zhang J, Tao D, Zhuang B. Sensitivity-Aware Visual Parameter-Efficient Fine-Tuning. 2023 [cited 2024 Apr 15]. pp. 11825–35. [https://openaccess.thecvf.com/content/ICCV2023/html/He\\_Sensitivity-Aware\\_Visual\\_Parameter-Efficient\\_Fine-Tuning\\_ICCV\\_2023\\_paper.html](https://openaccess.thecvf.com/content/ICCV2023/html/He_Sensitivity-Aware_Visual_Parameter-Efficient_Fine-Tuning_ICCV_2023_paper.html)
- Liu P, Yuan W, Fu J, Jiang Z, Hayashi H, Neubig G. Pre-train, prompt, and predict: a systematic survey of prompting methods in Natural Language Processing. *ACM Comput Surv*. 2023;55:195:1–195.
- Lücking R. Stop the abuse of Time! Strict temporal banding is not the future of Rank-based classifications in Fungi (including Lichens) and other organisms. *CRC Crit Rev Plant Sci*. 2019;38:199–253.
- Briggs D, Walters SM. *Plant Variation and Evolution*. Cambridge University Press; 2016.
- Apriyanti DH, Spreuwers LJ, Lucas PJF. Deep neural networks for explainable feature extraction in orchid identification. *Appl Intell*. 2023;53:26270–85.

26. Sing-Chi C, Zhong-Jian LIU. Critical notes on some taxa of Cymbidium. *J Syst Evol.* 2003;41:79.
27. Wen S, Wang J. A strong baseline for image and video quality assessment [Internet]. arXiv; 2021 [cited 2024 Mar 13]. <http://arxiv.org/abs/2111.07104>
28. Samanta P, Jain S. Analysis of Perceptual Hashing algorithms in Image Manipulation Detection. *Procedia Comput Sci.* 2021;185:203–12.
29. Tarekegn AN, Giacobini M, Michalak K. A review of methods for imbalanced multi-label classification. *Pattern Recogn.* 2021;118:107965.
30. Nilsback M-E, Zisserman A. Automated Flower Classification over a Large Number of Classes. 2008 Sixth Indian Conference on Computer Vision, Graphics & Image Processing [Internet]. 2008 [cited 2024 Mar 13]. pp. 722–9. <https://ieeexplore.ieee.org/abstract/document/4756141>
31. Wah C, Branson S, Welinder P, Perona P, Belongie SJ. The Caltech-UCSD Birds-200-2011 Dataset. 2011 [cited 2024 Apr 23]. <https://www.semantic-scholar.org/paper/The-Caltech-UCSD-Birds-200-2011-Dataset-Wah-Branson/c069629a51f6c1c301eb20ed77bc6b586c24ce32>
32. Khosla A, Jayadevaprakash N, Yao B, Li F-F. Novel Dataset for Fine-Grained Image Categorization: Stanford Dogs.
33. Bossard L, Guillaumin M, Van Gool L. Food-101 – mining discriminative components with Random forests. In: Fleet D, Pajdla T, Schiele B, Tuytelaars T, editors. *Computer vision – ECCV 2014*. Cham: Springer International Publishing; 2014. pp. 446–61.
34. Liu Z, Luo P, Qiu S, Wang X, Tang X, DeepFashion. Powering Robust Clothes Recognition and Retrieval With Rich Annotations. 2016 [cited 2024 Mar 13]. pp. 1096–104. [https://openaccess.thecvf.com/content\\_cvpr\\_2016/html/Liu\\_DeepFashion\\_Powering\\_Robust\\_CVPR\\_2016\\_paper.html](https://openaccess.thecvf.com/content_cvpr_2016/html/Liu_DeepFashion_Powering_Robust_CVPR_2016_paper.html)
35. Hou S, Feng Y, Wang Z, VegFru. A Domain-Specific Dataset for Fine-Grained Visual Categorization. 2017 [cited 2024 Mar 13]. pp. 541–9. [https://openaccess.thecvf.com/content\\_iccv\\_2017/html/Hou\\_VegFru\\_A\\_Domain-Specific\\_ICCV\\_2017\\_paper.html](https://openaccess.thecvf.com/content_iccv_2017/html/Hou_VegFru_A_Domain-Specific_ICCV_2017_paper.html)
36. Bai Y, Chen Y, Yu W, Wang L, Zhang W. Products-10K: A Large-scale Product Recognition Dataset. arXiv:200810545 [cs] [Internet]. 2020 [cited 2022 Mar 13]; <http://arxiv.org/abs/2008.10545>
37. Van Horn G, Cole E, Beery S, Wilber K, Belongie S, Mac Aodha O. Benchmarking Representation Learning for Natural World Image Collections. 2021 [cited 2024 Mar 13]. pp. 12884–93. [https://openaccess.thecvf.com/content/CVPR2021/html/Van\\_Horn\\_Benchmarking\\_Representation\\_Learning\\_for\\_Natural\\_World\\_Image\\_Collections\\_CVPR\\_2021\\_paper.html](https://openaccess.thecvf.com/content/CVPR2021/html/Van_Horn_Benchmarking_Representation_Learning_for_Natural_World_Image_Collections_CVPR_2021_paper.html)
38. Housby N, Giurgiu A, Jastrzebski S, Morrone B, Laroussilhe QD, Gesmundo A et al. Parameter-Efficient Transfer Learning for NLP. *Proceedings of the 36th International Conference on Machine Learning* [Internet]. PMLR; 2019 [cited 2024 Mar 13]. pp. 2790–9. <https://proceedings.mlr.press/v97/housby19a.html>
39. Hu EJ, Shen Y, Wallis P, Allen-Zhu Z, Li Y, Wang S et al. LoRA: Low-Rank Adaptation of Large Language Models [Internet]. arXiv; 2021 [cited 2024 Mar 9]. <http://arxiv.org/abs/2106.09685>
40. Yosinski J, Clune J, Bengio Y, Lipson H. How transferable are features in deep neural networks? *Advances in Neural Information Processing Systems*. Curran Associates, Inc.; 2014 [cited 2024 Mar 13]. [https://proceedings.neurips.cc/paper\\_files/paper/2014/hash/375c71349b295f5be2dcda9206f20a06-Abstract.html](https://proceedings.neurips.cc/paper_files/paper/2014/hash/375c71349b295f5be2dcda9206f20a06-Abstract.html)
41. Zhu Y, Shen Z, Zhao Z, Wang S, Wang X, Zhao X et al. MeLo: Low-rank Adaptation is Better than Fine-tuning for Medical Image Diagnosis [Internet]. arXiv; 2023 [cited 2024 Apr 15]. <http://arxiv.org/abs/2311.08236>
42. Cai H, Gan C, Zhu L, Han S, TinyTL. Reduce Memory, Not Parameters for Efficient On-Device Learning. *Advances in Neural Information Processing Systems* [Internet]. Curran Associates, Inc.; 2020 [cited 2024 Mar 13]. pp. 11285–97. <https://proceedings.neurips.cc/paper/2020/hash/81f7acabd411274fcf65ce2070ed568a-Abstract.html>
43. Zhang JO, Sax A, Zamir A, Guibas L, Malik J, Side-Tuning C. 2020 [cited 2024 Mar 13]. pp. 698–714. [https://link.springer.com/chapter/https://doi.org/10.1007/978-3-030-58580-8\\_41](https://link.springer.com/chapter/https://doi.org/10.1007/978-3-030-58580-8_41)
44. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*. Red Hook, NY, USA: Curran Associates Inc.; 2012. pp. 1097–105.
45. Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L. ImageNet: A large-scale hierarchical image database. 2009 IEEE Conference on Computer Vision and Pattern Recognition. 2009. pp. 248–55.
46. He K, Chen X, Xie S, Li Y, Dollár P, Girshick R. Masked Autoencoders Are Scalable Vision Learners. 2022 [cited 2024 Apr 15]. pp. 16000–9. [https://openaccess.thecvf.com/content/CVPR2022/html/He\\_Masked\\_Autoencoders\\_Are\\_Scalable\\_Vision\\_Learners\\_CVPR\\_2022\\_paper](https://openaccess.thecvf.com/content/CVPR2022/html/He_Masked_Autoencoders_Are_Scalable_Vision_Learners_CVPR_2022_paper)
47. Arwatchananukul S, Kirimasthong K, Aunsri N. A New Paphiopedilum Orchid Database and its Recognition using convolutional neural network. *Wirel Pers Commun.* 2020;115:3275–89.
48. Apriyanti DH, Spreeuwens LJ, Lucas PJF, Veldhuis RNJ. Automated color detection in orchids using color labels and deep learning. *PLoS ONE.* 2021;16:e0259036.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.