Energy-Efficient Plant Monitoring via Knowledge Distillation

Ilyass Moummad[1] [0009-0003-2925-2500], Reda Bensaid[2] [0009-0008-5369-0857], Kawtar Zaher[1,3] [0009-0008-9920-9483], Hervé Goëau[4] [0000-0003-3296-3795], Jean-Christophe Lombardo[1] [0000-0002-9656-4219], Joseph Salmon[5] [0000-0002-3181-0634], Pierre Bonnet[4] [0000-0002-2828-4389], and Alexis Joly[1] [0000-0002-2161-9940]

[1] LIRMM, Univ Montpellier, CNRS, Inria, Montpellier, France
[2] IMT Atlantique, Brest, France. Polytechnique Montréal, Canada
[3] Institut National de l'Audiovisuel, Paris, France
[4] AMAP, Université de Montpellier, CIRAD, INRAE, IRD, Montpellier, France
[5] IMAG, Univ Montpellier, CNRS, Inria, Montpellier, France
{ilyass.moummad,alexis.joly}@inria.fr

Abstract. Recent advances in large-scale visual representation learning have significantly improved performance in plant species and plant disease recognition tasks. However, state-of-the-art models, often based on high-capacity vision transformers or multimodal foundation models, remain computationally expensive and difficult to deploy in resource-constrained environments such as mobile or edge devices. This limitation hinders the scalability of automated biodiversity monitoring and precision agriculture systems, where efficiency is as critical as accuracy. In this work, we investigate knowledge distillation as an effective approach to transfer the representational capacity of large pretrained models into smaller, more efficient architectures. We focus on plant species and disease recognition, and conduct an empirical study on two challenging benchmarks: Pl@ntNet300K-v2 and Deep-Plant-Disease. We evaluate two representative architectures, a ConvNeXt model and a vision transformer, with pretrained initialization, each with and without distillation. Our results show that knowledge distillation consistently improves performance. Distilled models are able to closely match the performance of significantly larger models while maintaining substantially lower computational cost. These findings demonstrate the potential of knowledge distillation techniques to enable efficient and scalable deployment of plant recognition systems in real-world environmental applications.[6]

Keywords: Automatic plant monitoring, transfer learning, knowledge distillation, green pattern recognition

1 Introduction

The rapid decline of biodiversity and the increasing prevalence of plant diseases pose major challenges for ecosystems, agriculture, and global food security [17].

[6] Code and models: https://github.com/ilyassmoummad/distillplant

Scalable monitoring solutions are therefore essential to support biodiversity conservation, ecological research, and precision agriculture [9]. In this context, automated visual recognition systems have emerged as a key enabling technology, allowing non-experts and large-scale platforms to identify plant species and diagnose diseases directly from images [3].

Recent progress in computer vision has significantly advanced the capabilities of such systems. In particular, large-scale pretrained models, often based on transformer architectures, have demonstrated strong performance in fine-grained visual recognition tasks [6]. By leveraging massive datasets and high-capacity architectures, these models are able to capture subtle visual differences between species and disease patterns [23]. As a result, modern systems such as large vision encoders [15] and deployed applications like Pl@ntNet [3] achieve high accuracy across diverse environmental conditions.

However, these gains come at a substantial computational cost [18]. High-capacity models require significant memory and processing resources, making them difficult to deploy in practical scenarios such as mobile devices, embedded platforms, or in-field sensors. This limitation is especially critical for environmental monitoring, where scalability, energy efficiency, and accessibility are key constraints [21]. Consequently, there is a growing need for approaches that retain the accuracy of large models while enabling efficient inference.

Model compression techniques provide a natural direction to address this challenge [13]. Among them, knowledge distillation has emerged as a flexible framework for transferring information from a large teacher model to a smaller student model [7]. Rather than training compact models solely from labeled data, distillation leverages the richer supervisory signal provided by the teacher, enabling improved generalization and performance. While this paradigm has been successfully applied in several computer vision domains, its application in fine-grained plant recognition and disease classification settings remains unexplored.

To address this gap, we present a comprehensive empirical study of knowledge distillation for plant species and disease recognition. We evaluate distillation on the Pl@ntNet300K-v2 and Deep-Plant-Disease benchmarks using a convolutional network and a vision transformer as student architectures. Our analysis compares standard fine-tuning and distillation-based training under pretrained settings across multiple teacher–student configurations. The results show that knowledge distillation consistently improves the performance of compact models, enabling them to better approximate the discriminative capabilities of larger architectures while maintaining lower computational requirements. These findings highlight the potential of distillation as an effective model compression strategy for efficient biodiversity monitoring and plant health assessment systems.

2 Related Work

Automatic plant monitoring. The development of automated plant monitoring systems has accelerated in recent years [3], driven by the need for scalable biodiversity assessment and accessible tools for species identification. Applications such as ObsIdentify [14], Flora Incognita [11], iNaturalist [8] and Pl@ntNet [3] have demonstrated the potential of large-scale visual recognition systems to support both citizen science and expert-driven ecological studies. These systems typically rely on large deep learning models trained on massive datasets, enabling robust performance across diverse environmental conditions and species distributions [2]. However, these approaches depend on computationally intensive architectures, limiting their deployment in real-world scenarios where low-latency and energy-efficient inference are required. In contrast to prior work that prioritizes accuracy through increasingly large models, our work focuses on improving the efficiency of plant recognition systems by transferring the capabilities of such models into smaller architectures.

Plant recognition datasets. The progress of plant recognition methods has been closely tied to the availability of large and diverse annotated datasets. General-purpose biodiversity datasets such as iNaturalist [23] provide large-scale collections of species observations in natural conditions, supporting the development of robust fine-grained classifiers. More specialized datasets have been introduced to address domain-specific challenges: Pl@ntNet300K-v2 [2] focuses on plant species recognition from user-contributed images with high variability in viewpoint and quality; Deep-Plant-Disease [1] targets disease classification with controlled and field conditions; PlantDoc [20] and PlantWild [24] introduce challenging real-world disease and species recognition scenarios with complex backgrounds; and the Pl@ntCLEF benchmark series provides standardized evaluation protocols for plant identification at scale [4].

Model compression and knowledge distillation. Model compression techniques aim to reduce the computational and memory footprint of deep neural networks while maintaining predictive performance [13]. Common approaches include pruning [10], which removes redundant parameters, quantization [5], which lowers numerical precision to improve efficiency, and knowledge distillation [7], which transfers knowledge from a large teacher model to a smaller student model. Among these methods, knowledge distillation has demonstrated strong empirical effectiveness across a wide range of computer vision tasks [16, 12]. However, its application to fine-grained plant recognition, including plant species and disease classification, remains underexplored. To address this gap, we investigate the effectiveness of knowledge distillation across multiple teacher–student architectures for plant species and disease recognition.

3 Knowledge Distillation for Plant Monitoring

3.1 Method

We aim to train efficient models for plant species and disease recognition by transferring knowledge from large pretrained visual models. To this end, we adopt a task-specific knowledge distillation framework from Marrie et al. [12], in which a large teacher model is first adapted to the downstream task, and a compact student model is subsequently trained to reproduce its predictions. The overall pipeline is illustrated in Figure 1.

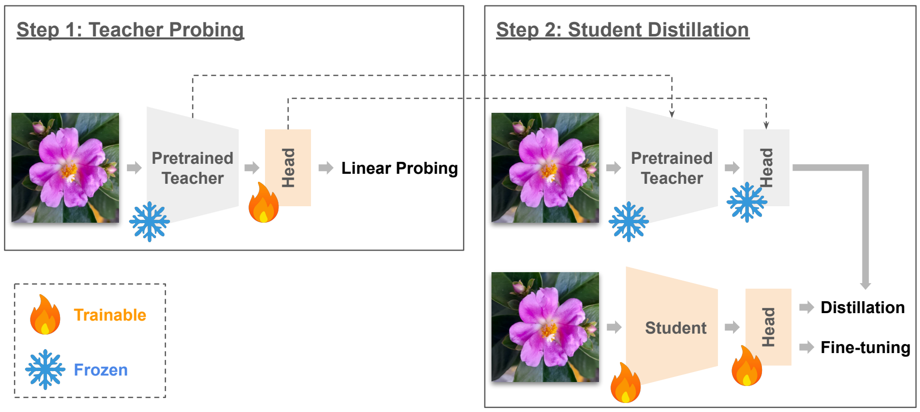

Fig. 1. Two-step knowledge distillation pipeline. **Step 1:** A pretrained teacher model is adapted to the downstream task via linear probing. **Step 2:** A student model is trained to solve the task while aligning with the teacher's predictions.

To evaluate the impact of distillation across realistic scenarios, we consider multiple training configurations. Specifically, we vary (i) the training objective (standard supervised learning or distillation), (ii) the teacher model (supervised, self-supervised, or multimodal), and (iii) the student architecture, including both a convolutional network and a vision transformer.

3.2 Experimental Setup

Datasets. We conduct experiments on two large-scale and complementary benchmarks. Pl@ntNet300K-v2 [2] is a fine-grained plant species recognition dataset composed of images collected in real-world conditions, exhibiting high intra-class variability and long-tailed distributions. Deep-Plant-Disease [1] focuses on plant disease classification and includes diverse disease categories across multiple plant species. Together, these datasets capture both biodiversity monitoring and agricultural use cases.

Table 1. Summary of evaluation datasets.

Dataset	#Train	#Val	#Test	#Classes
Pl@ntNet300K-v2	243,866	31,115	31,106	1,000
Deep-Plant-Disease	198,710	–	49,866	175

Teacher models. We consider three high-capacity pretrained models as teachers, selected for their strong performance on plant-related visual tasks. BioCLIP-2 [6] is a ViT-L model trained with a contrastive language-image pretraining objective on a large-scale dataset structured around biological taxonomy, providing rich multimodal representations. Pl@ntCLEF 2024 [4] is a ViT-B model pretrained specifically for plant species classification on the flora of southwestern Europe, making it suitable for fine-grained plant recognition. Finally, DINOv3 [19] ViT-L is a self-supervised model pretrained on large-scale image data, producing high-quality visual representations that capture general-purpose features without relying on labels, that are potentially useful for plant recognition tasks. All teacher models are adapted to downstream tasks using a linear probing strategy, where a linear classification head is trained on top of frozen features. This approach yields strong yet computationally efficient teacher models for task-specific knowledge distillation.

Student models. We evaluate two compact architectures representative of modern vision models: ConvNeXt-Tiny (CNX-T) and ViT-Small (ViT-S). Models are initialized from DINOv3 pretrained weights obtained via self-supervised learning on the large-scale LVD-1689M [19]. All student models are trained under two regimes: standard supervised fine-tuning, and knowledge distillation combined with supervised fine-tuning.

Computational complexity. Table 2 summarizes the number of parameters and estimated inference cost measured in GFLOPs (Giga Floating-Point Operations per second) for both student and teacher models at a 224×224 resolution. The compact student models require substantially fewer resources than the larger teacher models, underscoring their practical efficiency for deployment in resource-constrained scenarios.

Table 2. Model complexity at 224×224 resolution.

Model	**#Params**	**GFLOPs**
Student models		
ConvNeXt-T	29M	3.8
ViT-S	21M	9.2
Teacher models		
ViT-B	86M	36.0
ViT-L	300M	125.0

Training details. Both teacher and student models are trained for 30 epochs using the AdamW optimizer with a learning rate of 10^{-4}, a minimum learning rate of 10^{-6} following a cosine decay schedule, and a weight decay of 10^{-4}. Teachers are trained via linear probing on frozen features, while students are trained end-to-end. For a fair comparison across all configurations, we adopt this single stable training setup without hyperparameter tuning, which is sufficient to achieve reliable convergence within 30 epochs. For distillation hyperparameters, we use the same setup as Marrie et al. [12].

3.3 Results on Pl@ntNet300K-v2

Table 3 reports Top-1 (micro) accuracy on Pl@ntNet300K-v2. Different teacher models achieve comparable performance, providing strong supervision signals.

Model	Init	Strategy	Teacher	Top-1 (%)
		Teachers		
BioCLIP-2 (300M)	TreeOfLife-200M	Linear Probe	–	86.8
Pl@ntCLEF (86M)	Pl@ntCLEF2024	Linear Probe	–	86.1
DINOv3-L (300M)	LVD-1689M	Linear Probe	–	86.1
		Students		
ConvNeXt-T (29M)	LVD-1689M	Finetune	–	81.9
		Distill	BioCLIP-2	85.8
		Distill	Pl@ntCLEF	85.4
		Distill	DINOv3-L	85.5
ViT-S (21M)	LVD-1689M	Finetune	–	83.1
		Distill	BioCLIP-2	85.5
		Distill	Pl@ntCLEF	85.2
		Distill	DINOv3-L	85.3

Table 3. Task-specific distillation of large pretrained visual models on Pl@ntNet300K-v2. Teacher models are trained via linear probing. Student models are trained either with fine-tuning or distillation from teacher logits.

Effect of distillation. Knowledge distillation consistently improves performance over standard fine-tuning, providing consistent improvements, typically in the range of +2 to +4 points. Notably, ConvNeXt-T improves from 81.9% to 85.8% when distilled from BioCLIP-2, nearly reaching teacher-level performance.

3.4 Feature space analysis

For qualitative analysis, we use a subset of the Pl@ntNet300K dataset from the MetaAlbum benchmark [22], restricted to the 25 most populous classes following BioCLIP [6], to visualize feature representations. We project the learned embeddings using t-SNE on this subset (Figure 2). CNX-T initialized from pretrained DINOv3 features exhibits limited class separability, with overlapping clusters across species. Fine-tuning improves the structure of the embedding space, leading to more compact and better-separated clusters, while distillation further enhances this effect by producing tighter clusters and clearer class boundaries. Interestingly, the distilled student exhibits more structured representations than the teacher in this setting, suggesting that task-specific distillation can refine feature organization beyond what is directly inherited from large pretrained models.

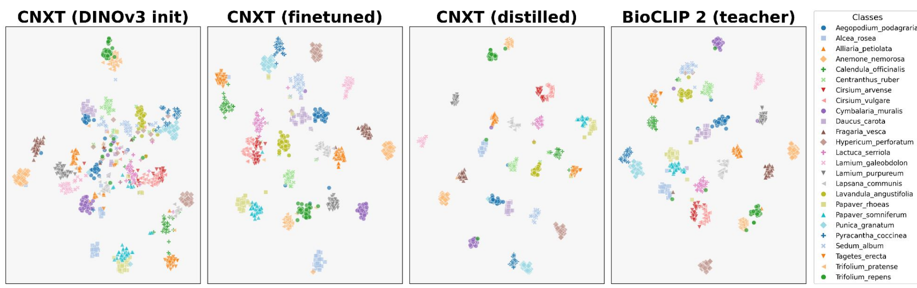

Fig. 2. t-SNE visualization of feature embeddings on the PlantNet MetaAlbum dataset (Pl@ntNet300K subset with 25 most populous classes) for different models.

3.5 Results on Deep-Plant-Disease

Table 4 presents results on Deep-Plant-Disease. Compared to Pl@ntNet300K-v2 (Table 3), overall accuracy is lower for teacher models, reflecting the distribution shift and increased complexity of disease classification relative to the pretraining data.

Effect of distillation. Distillation again leads to consistent improvements across all architectures. For example, ViT-S improves from 80.5% to 82.2% when distilled from BioCLIP-2, while ConvNeXt-T improves from 79.9% to 82.4% when distilled from DINOv3-L. These findings suggest that the benefits of knowledge distillation are not limited to species recognition but also extend to disease classification, despite the substantial distribution shift from the pretraining data.

Teacher comparison. In contrast to Pl@ntNet300K-v2, DINOv3-L emerges as the strongest teacher on this dataset, yielding the highest accuracy among probed models. While teacher probing alone underperforms compared to student finetuning (e.g., DINOv3-L linear probe achieves 76.6% versus 80.5% for a ViT-S student initialized from LVD-1689M), it still provides valuable guidance. Distillation from these teachers consistently improves student performance, indicating that even suboptimal teacher probes can benefit student models when aligned with the target domain.

Model	Init	Strategy	Teacher	Top-1 (%)
		Teachers		
BioCLIP-2 (300M)	TreeOfLife-200M	Linear Probe	–	74.2
Pl@ntCLEF (86M)	Pl@ntCLEF2024	Linear Probe	–	72.9
DINOv3-L (300M)	LVD-1689M	Linear Probe	–	76.6
		Students		
ConvNeXt-T (29M)	LVD-1689M	Finetune	–	79.9
		Distill	BioCLIP-2	82.7
		Distill	PlantCLEF	80.7
		Distill	DINOv3-L	82.4
ViT-S (21M)	LVD-1689M	Finetune	–	80.5
		Distill	BioCLIP-2	82.2
		Distill	PlantCLEF	80.7
		Distill	DINOv3-L	82.1

Table 4. Task-specific distillation of large pretrained visual models on Deep-Plant-Disease, following the same experimental setup as in Table 3.

3.6 Discussion

Across both datasets, our experiments demonstrate that knowledge distillation is a robust and effective strategy for improving compact models on plant-related tasks. It consistently enhances performance across both convolutional and transformer architectures, as well as across different tasks. While pretrained models already provide strong performance, distillation systematically complements pre-training, enabling smaller models to closely match the performance of much larger systems. These results highlight the practical value of knowledge distillation for deploying efficient and accurate models in real-world environmental monitoring applications, particularly when large pretrained models are available as teachers but are costly to fine-tune.

4 Conclusion

In this work, we explored knowledge distillation as a practical approach to reduce the computational footprint of modern plant recognition systems while preserving high predictive performance. Large pretrained models have become central to biodiversity monitoring and plant disease recognition, but their deployment at scale remains constrained by their cost in terms of memory, computation, and energy. Our results show that distillation provides an effective mechanism to transfer their capabilities into significantly smaller models, enabling more efficient inference without sacrificing much accuracy.

Through a comprehensive empirical study across datasets, architectures, we demonstrated that distilled models consistently outperform their non-distilled counterparts and can closely match the performance of much larger teacher models. Importantly, these gains hold across both species recognition and disease classification tasks, and complement strong pretrained initializations. This suggests that distillation can be a key tool for improving the efficiency–accuracy trade-off in environmental monitoring applications. By reducing model size and inference cost, distillation facilitates deployment on resource-constrained platforms such as mobile devices, embedded sensors, or edge computing infrastructures, which are essential for large-scale and continuous biodiversity observation.

Looking forward, combining distillation with other model compression techniques, such as pruning and quantization, represents a promising direction to further decrease energy consumption and hardware requirements. Overall, this work contributes to the development of more sustainable machine learning approaches for biodiversity monitoring, and we hope it will encourage further research at the intersection of model efficiency, computer vision, and environmental applications.

Acknowledgements This project was funded by the French National Research Agency (ANR) through the grant Pl@ntAgroEco 22-PEAE-0009.

References

1. Chai, A.Y.H., Jee, K.L.Z., Lee, S.H., Tay, F.S., Vandeputte, J., Goeau, H., Bonnet, P., Joly, A.: Deep-plant-disease dataset is all you need for plant disease identification. In: Proceedings of the 33rd ACM International Conference on Multimedia. p. 12578–12584. MM '25, Association for Computing Machinery, New York, NY, USA (2025)
2. Garcin, C., Joly, A., Bonnet, P., Affouard, A., Lombardo, J.C., Chouet, M., Servajean, M., Lorieul, T., Salmon, J.: Pl@ntNet-300K: a plant image dataset with high label ambiguity and a long-tailed distribution. In: Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks (2021)
3. Goëau, H., Bonnet, P., Joly, A., Bakić, V., Barbe, J., Yahiaoui, I., Selmi, S., Carré, J., Barthélémy, D., Boujemaa, N., et al.: Pl@ ntnet mobile app. In: Proceedings of the 21st ACM international conference on Multimedia. pp. 423–424 (2013)
4. Goeau, H., Espitalier, V., Bonnet, P., Joly, A.: Overview of PlantCLEF 2024: Multi-species plant identification in vegetation plot images. In: Working Notes of CLEF 2024 – Conference and Labs of the Evaluation Forum. CEUR Workshop Proceedings (2024)
5. Gong, Y., Liu, L., Yang, M., Bourdev, L.: Compressing Deep Convolutional Networks using Vector Quantization. arXiv preprint arXiv:1412.6115 (2014)
6. Gu, J., Stevens, S., Campolongo, E.G., Thompson, M.J., Zhang, N., Wu, J., Kopanev, A., Mai, Z., White, A.E., Balhoff, J., Dahdul, W., Rubenstein, D., Lapp, H., Berger-Wolf, T., Chao, W.L., Su, Y.: BioCLIP 2: Emergent properties from scaling hierarchical contrastive learning. In: The Thirty-ninth Annual Conference on Neural Information Processing Systems (2025)
7. Hinton, G., Vinyals, O., Dean, J.: Distilling the Knowledge in a Neural Network. arXiv preprint arXiv:1503.02531 (2015)
8. iNaturalist community: inaturalist. https://www.inaturalist.org (2025)

9. Jetz, W., Cavender-Bares, J., Pavlick, R., Schimel, D., Davis, F.W., Asner, G.P., Guralnick, R., Kattge, J., Latimer, A.M., Moorcroft, P., et al.: Monitoring plant functional diversity from space. Nature plants **2**(3), 16024 (2016)
10. LeCun, Y., Denker, J., Solla, S.: Optimal Brain Damage. Advances in neural information processing systems **2** (1989)
11. Mäder, P., Boho, D., Rzanny, M., Seeland, M., Wittich, H.C., Deggelmann, A., Wäldchen, J.: The Flora Incognita app – Interactive plant species identification. Methods in Ecology and Evolution **12**(7), 1335–1342 (2021)
12. Marrie, J., Arbel, M., Mairal, J., Larlus, D.: On Good Practices for Task-Specific Distillation of Large Pretrained Visual Models. Transactions on Machine Learning Research (2024)
13. Menghani, G.: Efficient Deep Learning: A Survey on Making Deep Learning Models Smaller, Faster, and Better. ACM Computing Surveys **55**(12), 1–37 (2023)
14. Observation International: Obsidentify: Wildlife and plant identification app. https://observation.org/apps/obsidentify/ (2026)
15. Oquab, M., Darcet, T., Moutakanni, T., Vo, H., Szafraniec, M., Khalidov, V., Fernandez, P., Haziza, D., Massa, F., El-Nouby, A., Assran, M., Ballas, N., Galuba, W., Howes, R., Huang, P.Y., Li, S.W., Misra, I., Rabbat, M., Sharma, V., Synnaeve, G., Xu, H., Jégou, H., Bojanowski, P., LeCun, Y., Caron, M.: DINOv2: Learning Robust Visual Features without Supervision. Transactions on Machine Learning Research (TMLR) (2023)
16. Sariyildiz, M.B., Weinzaepfel, P., Lucas, T., Larlus, D., Kalantidis, Y.: UNIC: Universal Classification Models via Multi-teacher Distillation. In: European Conference on Computer Vision (ECCV) (2024)
17. Savary, S., Willocquet, L., Pethybridge, S.J., Esker, P., McRoberts, N., Nelson, A.: The global burden of pathogens and pests on major food crops. Nature ecology & evolution **3**(3), 430–439 (2019)
18. Schwartz, R., Dodge, J., Smith, N.A., Etzioni, O.: Green AI. Communications of the ACM **63**(12), 54–63 (2020)
19. Siméoni, O., Vo, H.V., Seitzer, M., Baldassarre, F., Oquab, M., Jose, C., Khalidov, V., Szafraniec, M., Yi, S., Ramamonjisoa, M., Massa, F., Haziza, D., Wehrstedt, L., Wang, J., Darcet, T., Moutakanni, T., Sentana, L., Roberts, C., Vedaldi, A., Tolan, J., Brandt, J., Couprie, C., Mairal, J., Jégou, H., Labatut, P., Bojanowski, P.: DINOv3 (2025), https://arxiv.org/abs/2508.10104
20. Singh, D., Jain, N., Jain, P., Kayal, P., Kumawat, S., Batra, N.: PlantDoc: A Dataset for Visual Plant Disease Detection. In: Proceedings of the 7th ACM IKDD CoDS and 25th COMAD, pp. 249–253 (2020)
21. Tuia, D., Kellenberger, B., Beery, S., Costelloe, B.R., Zuffi, S., Risse, B., Mathis, A., Mathis, M.W., Van Langevelde, F., Burghardt, T., et al.: Perspectives in machine learning for wildlife conservation. Nature communications **13**(1), 792 (2022)
22. Ullah, I., Carrión-Ojeda, D., Escalera, S., Guyon, I., Huisman, M., Mohr, F., Van Rijn, J.N., Sun, H., Vanschoren, J., Vu, P.A.: Meta-Album: Multi-domain Meta-Dataset forFew-Shot Image Classification. Advances in Neural Information Processing Systems **35**, 3232–3247 (2022)
23. Vendrow, E., Pantazis, O., Shepard, A., Brostow, G., Jones, K.E., Mac Aodha, O., Beery, S., Van Horn, G.: INQUIRE: A natural world text-to-image retrieval benchmark. Advances in Neural Information Processing Systems **37**, 126500–126514 (2024)
24. Wei, T., Chen, Z., Huang, Z., Yu, X.: Benchmarking In-the-Wild Multimodal Plant Disease Recognition and A Versatile Baseline. In: Proceedings of the 32nd ACM International Conference on Multimedia. pp. 1593–1601 (2024)