Conference Abstract

Comparative Study: Evaluating the effects of class balancing on transformer performance in the PlantNet-300k image dataset

José Alejandro Chavarría Madriz‡, Maria Auxiliadora Mora-Cross§, William Ulate|,¶

‡ ITCR, Cartago, Costa Rica
§ ITCR, Alajuela, Costa Rica
| CRBio, San José, Costa Rica
¶ Missouri Botanical Garden, Saint Louis, United States of America

Abstract

Image-based identification of plant specimens plays a crucial role in various fields such as agriculture, ecology, and biodiversity conservation. The growing interest in deep learning has led to remarkable advancements in image classification techniques, particularly with the utilization of convolutional neural networks (CNNs). Since 2015, in the context of the PlantCLEF (Conference and Labs of the Evaluation Forum) challenge (Joly et al. 2015), deep learning models, specifically CNNs, have consistently achieved the most impressive results in this field (Carranza-Rojas 2018). However, recent developments have introduced transformer-based models, such as ViT (Vision Transformer) (Dosovitskiy et al. 2020) and CvT (Convolutional vision Transformer) (Wu et al. 2021), as a promising alternative for image classification tasks. Transformers offer unique advantages such as capturing global context and handling long-range dependencies (Vaswani et al. 2017), which make them suitable for complex recognition tasks like plant identification.

In this study, we focus on the image classification task using the PlantNet-300k dataset (Garcin et al. 2021a). The dataset consists of a large collection of 306,146 plant images

representing 1,081 distinct species. These images were selected from the PI@ntNet citizen observatory database. The dataset has two prominent characteristics that pose challenges for classification. First, there is a significant class imbalance, meaning that a small subset of species dominates the majority of the images. This imbalance creates bias and affects the accuracy of classification models. Second, many species exhibit visual similarities, making it tough, even for experts, to accurately identify them. These characteristics are referred to by the dataset authors as long-tailed distribution and high intrinsic ambiguity, respectively (Garcin et al. 2021b).

In order to address the inherent challenges of the PlantNet-300k dataset, we employed a two-fold approach. Firstly, we leveraged transformer-based models to tackle the dataset's intrinsic ambiguity and effectively capture the complex visual patterns present in plant images. Secondly, we focused on mitigating the class imbalance issue through various data preprocessing techniques, specifically class balancing methods. By implementing these techniques, we aimed to ensure fair representation of all plant species in order to improve the overall performance of image classification models.

Our objective is to assess the effects of data preprocessing techniques, specifically class balancing, on the classification performance of the PlantNet-300k dataset. By exploring different preprocessing methods, we addressed the class imbalance issue and through precise evaluation, conducted a comparison of the performance of transformer-based models with and without class balancing techniques. Through these efforts, our ultimate goal is to assert if these techniques allow us to achieve more accurate and reliable classification results, particularly for underrepresented species in the dataset.

In our experiment, we compared the performance of two transformer-based models, ViT and CvT, using two versions of the PlantNet-300k dataset: one with class balancing and the other without class balancing. This setup results in a total of four sets of metrics for evaluation. To assess the classification performance, we utilized a wide range of commonly used metrics including recall, precision, accuracy, AUC (Area Under the Curve), ROC (Receiver Operating Characteristic), and others. These metrics provide insights into each models' ability to correctly classify plant species, identify false positives and negatives, measure overall accuracy, and assess the models' discriminatory power.

By conducting this comparative study, we seek to contribute to the advancement of plant identification research by providing empirical evidence of the benefits and effectiveness of class balancing techniques in improving the performance of transformer-based models on the PlantNet-300k dataset and any other similar ones.

Keywords

deep learning, image-based identification, plant specimens, transformer-based models

Presenting author

José Alejandro Chavarría Madriz

Presented at

Conflicts of interest

The authors have declared that no competing interests exist.

References

- Carranza-Rojas J (2018) Towards Multi-Level Classification in Deep Plant Identification. Costa Rica Institute of Technology
- Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Houlsby N (2020) An image is worth 16x16 words: Transformers for image recognition at scale. arXiv https://doi.org/10.48550/arXiv.2010.11929
- Garcin C, Joly A, Bonnet P, Servajean M, Salmon J (2021a) Pl@ntNet-300K image dataset (1.0) [Data set]. Zenodo. Release date: 2021-4-29. URL: https://doi.org/10.5281/zenodo.4726653
- Garcin C, Joly A, Bonnet P, Affouard A, Lombardo J, Chouet M, Servajean M, Lorieul T, Salmon J (2021b) Pl@ntNet-300K image dataset. Zenodo https://doi.org/10.5281/zenodo.4726652
- Joly A, Goëau H, Glotin H, Spampinato C, Bonnet P, Vellinga W, Planqué R, Rauber A, Palazzo S, Fisher B, Müller H (2015) LifeCLEF 2015: Multimedia Life Species Identification Challenges. Lecture Notes in Computer Science462-483. https://doi.org/10.1007/978-3-319-24027-5_46
- Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Polosukhin I (2017) Attention is all you need. Advances in Neural Information Processing Systems 30 https://doi.org/10.48550/arXiv.1706.03762
- Wu H, Xiao B, Codella N, Liu M, Dai X, Yuan L, Zhang L (2021) CvT: Introducing Convolutions to Vision Transformers. arXiv https://doi.org/10.48550/arxiv.2103.15808