Enhancing Precision Agriculture: Deep Learning for Winter Crop Nutrient Deficiency Classification

Duong Anh Kiet
Faculty of Science and Technology, University of Limoges
23,Avenue Albert Thomas, 87060 Limoges, France
anh-kiet.duong@etu.unilim.fr

Abstract

This report outlines our approach and findings in the "CVPPA@ICCV'23: Image Classification of Nutrient Deficiencies in Winter Wheat and Winter Rye" competition. We applied deep learning techniques to identify nutrient deficiencies in winter crops using the DND-Diko-WWWR dataset. Our model leveraged similarity learning, concatenation of two models, and pseudo-labeling for improved accuracy, showcasing the potential of computer vision in precision agriculture.

1. Introduction

This report summarizes our approach and outcomes from the CVPPA@ICCV'23 competition, part of the 8th Workshop on Computer Vision in Plant Phenotyping and Agriculture (CVPPA) at ICCV 2023. The competition's objective was to tackle a crucial agricultural challenge: identifying nutrient deficiencies in winter wheat and winter rye crops, which profoundly affect crop yield and food security.

Our work centered on the DND-Diko-WWWR dataset, comprising 3,600 high-resolution UAV-based RGB images of winter wheat (WW2020) and winter rye (WR2021) [8, 9], captured at various stages of growth under different soil/nutrient conditions. Our report details our deep learning-based classification approach, following competition guidelines and using only the provided dataset for training. We also harnessed publicly available pre-trained models for improved performance.

We delve into our model architecture, hyperparameters, and training techniques. The report discusses data preprocessing, model selection, and fine-tuning strategies that optimized our performance on both WW2020 and WR2021 datasets. Our aim is to contribute to sustainable agriculture and advance computer vision in plant phenotyping. We look forward to sharing our insights with the CVPPA@ICCV'23 community and fellow participants. [code]

2. Method

In this section, we present our methodology for addressing the "CVPPA@ICCV'23: Image Classification of Nutrient Deficiencies in Winter Wheat and Winter Rye" competition. Our methodology comprises several crucial components designed to enhance classification accuracy and robustness.

2.1. Similarity Learning

We employ similarity learning techniques to enable the model to understand the relationships between nutrient deficiency patterns in winter wheat and winter rye. This allows for more effective feature representation and discrimination.

2.1.1 Loss function

To train the model, we employed the ArcFace [2] loss function with the use of cotangent instead of cosine, as introduced in [4]. The margin parameter was randomly sampled from a standard normal distribution, following the approach described in [1]. We set the parameters as follows: margin = 0.3, scale = 64, and sigma = 0.1 * margin.

2.1.2 Feature Extraction

To perform classification, we utilize the model to extract features from images within the training dataset, resulting in a feature vector of dimensions [1, 512] for each image. Subsequently, we compare the feature vector of the image to be classified with all feature vectors from the training set. We employ the k-Nearest Neighbors (kNN) algorithm for image classification based on this feature comparison. We implement (KNN) as a post-processing step to refine the classification results. KNN helps improve the overall accuracy by considering the similarities between samples.

2.1.3 Class Incremental

While Class Incremental learning wasn't a specified requirement of the competition, we recognized its potential significance for accommodating future label expansions, such as introducing new fertilizer techniques or plant treatments. This underscores an intriguing aspect of similarity learning, enabling the system to adapt and integrate new concepts without discarding previously acquired knowledge.

2.2. Model Architecture

In the following section, we will provide an in-depth exploration of the model's architecture, elucidating the intricate details of its structure and parameters, facilitating comprehensive reproducibility.

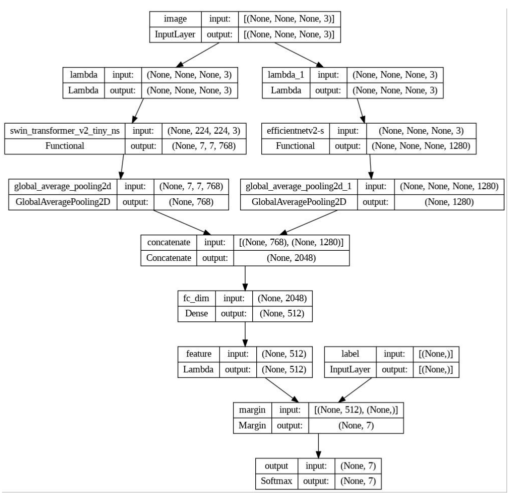

Figure 1. Model architecture

2.2.1 Architecture

We amalgamate two backbones: EfficientNetV2S [6, 7] and SwinTransformerTiny [5], one being a Convolutional Neural Network (CNN), and the other a Transformer [3]. This amalgamation serves the purpose of enhancing the diversity of features extracted. The approach involves utilizing both backbones to extract features and concatenating them together.

2.2.2 Parameters

The input images, with dimensions [224, 224, 3], are utilized within the model. Subsequently, global average pooling is applied to convert each backbone's output into a feature vector. These two feature vectors, extracted by each model, are then concatenated into a single vector of length 2048. Following this, we employ a fully connected layer to reduce the dimensionality of the feature vector to 512 and introduce labels for training, as per the description of the loss function.

We utilize a learning rate of $1e-5$ and a total of 10 epochs. This choice is driven by limited resources, and the utilization of similarity learning and loss functions as described. Increasing the number of epochs may result in only marginal improvements in accuracy, which may not be proportional to the resource investment.

2.3. Data Augmentation

Data augmentation plays a crucial role in preventing overfitting and improving the model's generalization capabilities. We employ various data augmentation techniques to increase the diversity of the training dataset and enhance model robustness.

We completely separate the two datasets, meaning that the model is trained exclusively on either WW2020 or WR2021, but not both simultaneously.

2.4. Image transformation

We utilize image transformation techniques to introduce variability into the dataset, making our model more adaptable to different conditions and viewpoints. We employed several basic data augmentation techniques in our approach, including: resize, scale, rotate, blur, random brightness, flip.

2.5. Pseudo Labeling

Pseudo labeling is a semi-supervised technique that utilizes unlabeled data to augment the training process. We incorporate pseudo labeling to make the most of available data resources and enhance model performance.

Our strategy for employing pseudo-labeling is straightforward. During each epoch of training, if the validation accuracy surpasses the previous best, we proceed to make predictions on the test dataset. We then consider the predicted label if its probability is greater than $threshold = 0.9$. We add these labels to the training set and continue training the model.

The utilization of pseudo-labeling significantly contributed to a substantial increase in the accuracy of our approach.

3. Results

	WW2020	WR2021	Mean
EfficientNetV2S +SwinTiny	66.9	77.1	72.0
EfficientNetV2S +SwinTiny +Pseudo-labeling	69.2	82.1	75.6

Table 1. Classification Results on WW2020 and WR2021 Datasets

4. Conclustion

In the "CVPPA@ICCV'23" competition, our deep learning approach, powered by similarity learning and pseudo-labeling, successfully identified nutrient deficiencies in winter crops. Although we did not achieve a top-three ranking, our solution showcases its adaptability for future label expansions. The scalability of our approach positions it well to accommodate an increase in the number of labels, meeting the evolving demands of agricultural research.

References

[1] Fadi Boutros, Naser Damer, Florian Kirchbuchner, and Arjan Kuijper. Elasticface: Elastic margin loss for deep face recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1578–1587, 2022.

[2] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4690–4699, 2019.

[3] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

[4] Anh-Kiet Duong, Hoang-Lan Nguyen, and Toan-Thinh Truong. Large margin cotangent loss for deep similarity learning. In *2022 International Conference on Advanced Computing and Analytics (ACOMPA)*, pages 40–47. IEEE, 2022.

[5] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 10012–10022, 2021.

[6] Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*, pages 6105–6114. PMLR, 2019.

[7] Mingxing Tan and Quoc Le. Efficientnetv2: Smaller models and faster training. In *International conference on machine learning*, pages 10096–10106. PMLR, 2021.

[8] Jinhui Yi, Lukas Krusenbaum, Paula Unger, Hubert Hüging, Sabine J Seidel, Gabriel Schaaf, and Juergen Gall. Deep learning for non-invasive diagnosis of nutrient deficiencies in sugar beet using rgb images. *Sensors*, 20:5893, 2020.

[9] Jinhui Yi, Gina Lopez, Sofia Hadir, Jan Weyler, Lasse Klingbeil, Marion Deichmann, Juergen Gall, and Sabine J Seidel. Non-invasive diagnosis of nutrient deficiencies in winter wheat and winter rye using uav-based rgb images. *Available at SSRN 4549653*.