Plant Phenotyping with Limited Annotation: Doing More with Less

Koushik Nagasubramanian[1,1,1], Sarah Jones[1], Asheesh Singh[1,1,1], Arti Singh[1,1,1], Soumik Sarkar[1,1,1], and Baskar Ganapathysubramanian[1,1,1]

[1]Iowa State University

November 30, 2022

Abstract

Deep learning (DL) methods have transformed the way we extract plant traits – both under laboratory as well as field conditions. Evidence suggests that "well-trained" DL models can significantly simplify and accelerate trait extraction as well as expand the suite of extractable traits. Training a DL model typically requires the availability of copious amounts of annotated data; however, creating large annotated dataset requires non-trivial efforts, time, and resources. This has become a major bottleneck in deploying DL tools in practice. Self-supervised learning (SSL) methods give exciting solution to this problem, as these methods use unlabeled data to produce pretrained models for subsequent fine-tuning on labeled data, and have demonstrated superior transfer learning performance on down-stream classification tasks. We investigated the application of SSL methods for plant stress classification using few labels. Plant stress classification is a fundamentally challenging problem in that (1) disease classification may depend on abnormalities in a small number of pixels, (2) high data imbalance across different classes, and (3) there are fewer annotated and available plant stress images than in other domains. We compared four different types of SSL methods on two different plant stress datasets. We report that pre-training on unlabeled plant stress images significantly outperforms transfer learning methods using random initialization for plant stress classification. In summary, SSL based model initialization and data curation improves annotation efficiency for plant stress classification tasks.

Plant Phenotyping with Limited Annotation: Doing More with Less

Koushik Nagasubramanian[1], Asheesh K Singh[2], Arti Singh[2], Soumik Sarkar[3], Baskar Ganapathysubramanian[3]

1. *Department of Electrical Engineering, Iowa State University, Ames, USA*
2. . *Department of Agronomy, Iowa State University, Ames, USA*
3. *Department of Mechanical Engineering, Iowa State University, Ames, USA*

Abstract

Deep learning (DL) methods have transformed the way we extract plant traits – both under laboratory as well as field conditions. Evidence suggests that "well-trained" DL models can significantly simplify and accelerate trait extraction as well as expand the suite of extractable traits. Training a DL model typically requires the availability of copious amounts of annotated data; however, creating large-scale annotated dataset requires non-trivial efforts, time, and resources. This has become a major bottleneck in deploying DL tools in practice. Self-supervised learning (SSL) methods give exciting solution to this problem, as these methods use unlabeled data to produce pretrained models for subsequent fine-tuning on labeled data, and have demonstrated superior transfer learning performance on down-stream classification tasks. We investigated the application of SSL methods for plant stress classification using few labels. Plant stress classification is a fundamentally challenging problem in that (1) disease classification may depend on abnormalities in a small number of pixels, (2) high data imbalance across different classes, and (3) there are fewer annotated and available plant stress images than in other domains. We compared four different types of SSL methods on two different plant stress datasets. We report that pre-training on unlabeled plant stress images significantly outperforms transfer learning methods using random initialization for plant stress classification. In summary, SSL based model initialization and data curation improves annotation efficiency for plant stress classification tasks.

Keywords: Deep Learning, Self-Supervised Learning, Plant Phenotyping, Plant Stress Classification