Self-supervised agricultural insect pest classification

Soumyashree Kar, Koushik Nagasubramanian, Dinakaran Elango, Ajay Nair, Daren S. Mueller, Matthew E. O'Neal, Asheesh K. Singh, Soumik Sarkar, Baskar Ganapathysubramanian, Arti Singh*

Iowa State University, USA

skar@iastate.edu, koushikn@iastate.edu, delango@iastate.edu, nairajay@iastate.edu, dsmuelle@iastate.edu, oneal@iastate.edu, singhak@iastate.edu, soumiks@iastate.edu, baskarg@iastate.edu, arti@iastate.edu*

ABSTRACT

Crop pest detection and mitigation remains an extremely challenging task for the farmers. Majority of the pest classification and detection techniques rely on supervised deep learning frameworks that require significant human intervention in labeling the input data, thereby making the down-stream tasks tedious. Therefore, this study presents a self-supervised learning (SSL) approach to classifying 12 types of agricultural insect pests from 9549 RGB images, by leveraging the Bootstrap your own latent (BYOL) algorithm. SSL uses minimal labeling and is indifferent to data augmentations or distortions. Hence, latent representations from pretrained SSL networks could be generalized well for downstream tasks like classification or object detection. For desirable classification of the insect images, the greatest challenges observed were: i) large intra-class variation (the same insect was found with different colors and patterns), and ii) complex background with inconspicuous foreground. Hence, to overcome these issues and aid generalizability of the representations learned through BYOL, entropy-guided segmentation (segments based on texture not color), is proposed as input to the SSL network in this study. Both raw and segmented images were separately fed to two independent BYOL SSL networks, i.e., with ResNet18 and ResNet50 architectures as the backbone. The efficacy of the latent representations for downstream applications was assessed using linear evaluation, and subsequently compared with traditional transfer learning outcomes from ResNet18 and ResNet50. The results indicated that, while ResNet50 backbone intuitively performed better in all cases, SSL aided with entropy-based segmentation led to ~94% classification accuracy compared to raw images (with ~90% maximum accuracy).