Deep Convolutional Neural Network's Applicability and Interpretability for Agricultural Machine Vision Systems

Harshana Habaragamuwa

The future capability of Agricultural Machine Vision Systems (AMVS) and the interpretability of their decisions holds the key to agricultural automation. The capability of AMVS is currently restricted by detection algorithms that use human selected characteristics for object category detection. These characteristics, such as color and shape, need specific imaging equipment and specific environmental conditions, which restricts their use to one object category. Thus, AMVS algorithms that can learn characteristics of multiple categories, are needed. In turn, to improve interpretability of such algorithms users will need to know what characteristics the algorithms have learned and their relationship to the detection. Such interpretations will help users to gain knowledge, predict the behavior, trust and ultimately improve AMVS that have a learning algorithm. Recently, Deep Convolutional Neural Networks (DCNNs) have performed well in such detection tasks.

We first developed the architecture and tested the applicability of a DCNN in AMVS for detection of mature and immature greenhouse grown strawberries as a model. Results were evaluated using the following parameters: average precision (AP), a parameter which measures maturity detection accuracy; and bounding box overlap (BBOL) which measures fruit boundary box occupancy. The DCNN attained an AP of 88.03% and 77.21%, and a BBOL of 0.7394 and 0.7045 respectively for mature and immature categories.

To improve the interpretability, a new pixel level decision interpretation method, which doesn't need to modify layers of the DCNN for interpretation, was developed. Performances of this method were evaluated by R^2 (sum of pixel effects vs adjusted class score), localization and visualization, using ImageNet data set (includes agricultural images), and two quite different DCNN architectures (GoogLeNet, VGG16). Results were compared with other interpretation methods (Deep Taylor Decomposition and Sensitivity Analysis). The sum of pixel level explanations described adjusted DCNN class score for database categories with an R^2 of 0.99. In addition, localization results were also competitive for both

DCNN architectures.

To improve the understandability of the DCNN decision interpretations a foundational feature level (patch level) decision interpretation method, which first selects foundational features and then calculates their effects to obtain a more understandable explanation of the DCNN decision to the user was developed. This method was evaluated in a similar manner to the previously developed pixel level interpretation method for both DCNN architectures, achieving an R^2 (sum of foundational feature effects vs adjusted class score) of 0.99. For this method, the localization performances were improved, and visualization was competitive. In this research, we established the applicability of using DCNNs in AMVS and improved the interpretability of the DCNN decisions.