

## The TerraByte Project

<https://www.acs.uwinnipeg.ca/terrabite/>

Christopher J. Henry, Christopher P. Bidinosti, Michael A. Beck, Manisha Ajmani

University of Winnipeg, Manitoba, Canada

**BACKGROUND/OBJECTIVES:** A lack of sufficient training data, both in terms of variety and quantity, is often the bottleneck in the development of machine learning (ML) applications in any domain. For agricultural applications, ML-based models designed to perform tasks such as autonomous plant classification will typically be coupled to just one or perhaps a few plant species. Consequently, each crop-specific task is very likely to require its own specialized training data, and the question of how to serve this need for data now often overshadows the more routine exercise of actually training such models.

**METHOD:** To tackle this problem, we have developed an embedded robotic system to automatically generate and label large datasets of plant images for ML applications in agriculture. The system can image plants from virtually any angle, thereby ensuring a wide variety of data; and with an imaging rate of up to one image per second, it can produce labeled datasets on the scale of thousands to tens of thousands of images per day. As such, this system offers an important alternative to time- and cost-intensive methods of manual generation and labeling.

**RESULTS:** The result is a labelled data of over 1.5 million labelled images and over 500 thousand multi-plant images with bounding boxes (Figs. 1 & 2). The images will be made publicly available via a web portal being developed in collaboration with the Enterprise and Machine Intelligence Learning Initiative (EMILI) and Compute Canada.



Figure 1. Labelled single-plant images.



Figure 2. Multi-plant images with bounding boxes.

**CONCLUSION/IMPLICATION:** Whether one speaks of precision agriculture, digital agriculture or Agriculture 4.0, this movement is expected to bring critical innovations based on automated methods and algorithms that are dependent on large amounts of data. Our hope is that our public dataset will spur innovation by researchers and industry in the agricultural sector in much the same way that ImageNet did for computer vision.