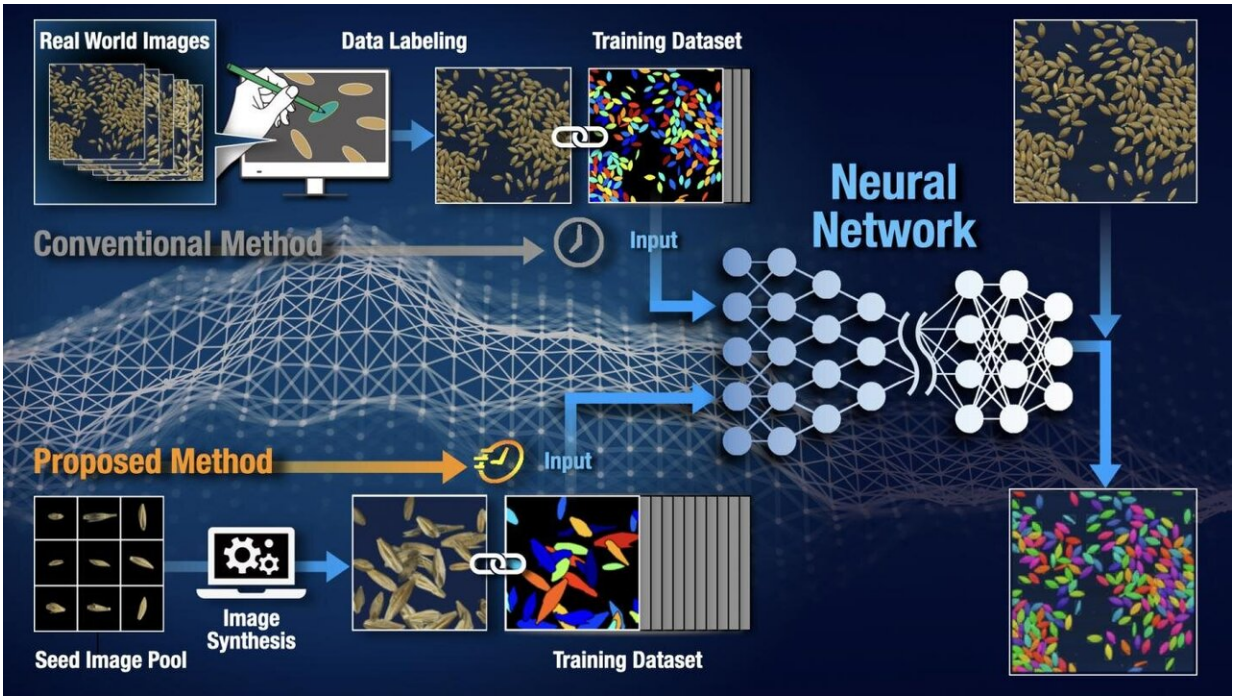
Training instance segmentation neural network with synthetic datasets for seed phenotyping

April 24 2020



Synthesizing training data in virtual space drastically reduces the time and effort required for its creation. The model trained with the synthesized dataset was as accurate as one trained with a handmade dataset. Credit: Issey Takahashi

In the midst of unprecedented climate change and population growth, the establishment of a method to rapidly create elite crop varieties via

selective breeding is a matter of urgency to maintain the food supply. In order to select such cultivars, it is necessary to define and evaluate a metric for what is a 'superior variety' in an efficient manner. For example, the shape of seeds is understood to be a trait closely linked to the quality and yield of crops, and is thus an important factor when conducting selective breeding.

A team of scientists led by Yosuke Toda, Designated Assistant Professor at the Institute of Transformative Bio-Molecules (WPI-ITbM), Nagoya University, and Fumio Okura, Assistant Professor at the Institute of Scientific and Industrial Research, Osaka University, have developed a system which utilizes image analysis and artificial intelligence (AI) to analyze the shape of large numbers of seeds from a single image. The shape of the seed is an important agronomic trait for the yield and quality of crops, and a method for automatically determining and evaluating such from an image is an indispensable tool for plant breeding.

Dr. Toda's research team generated a training dataset to be used for machine learning (deep learning) by synthesizing randomized barley seed images on a virtual canvas. The trained model, using only the synthesized data, was able to detect and segment the individual seeds from images of various barley cultivars as accurately as when done manually, as well as being able to analyze seeds of other crops.

Training data is required to make use of deep learning. Usually, training data is prepared by hand, for example by labeling every object in the images with different colors. However, for objects such as seeds, whose number is vast, creating the training data is very time consuming (for example, having to individually color hundreds of seeds for tens or hundreds of images for each seed variety). Thus, it has been considered difficult to generate a machine learning model that can quickly and simply analyze the seed shapes of different varieties or species.

Dr. Toda's research group succeeded in creating a large volume of training data from only a small number of seeds to effectively train the machine learning (deep learning) model. This approach is called domain randomization, and spares the effort involved in creating the training data, accelerating the development of machine learning models. In the proposed method, sample images of a small number of barley seeds whose shape information was already known were randomly arranged in virtual space, creating a large number and variety of synthesized images. The model trained with this dataset was able to detect the seeds and extract their shape data with the same degree of accuracy as when done by hand. No hand-annotated training dataset was required.

The experiment actually highlighted that the system can clearly identify the characteristic differences in shape of each crop. It is expected that in the future it will be possible to measure fine differences in the growth environment and variety, becoming a powerful tool for plant breeding.

Furthermore, the study showed that the same method can readily be employed to measure the seeds of a variety of different crops, such as rice, wheat, oats, and lettuce. These results strongly suggest that, regardless of crop, it is possible to make the automatic measurement of large numbers of seeds a reality. Beyond just a variety evaluation, this study is expected to contribute to the plant science domain by revealing characteristics of seeds not formerly observed by the human eye.

The majority of research into instance segmentation-based image analysis is conducted using existing datasets including things such as people and cars. On the other hand, plant image analysis has a variety of its own characteristics. Since there is great variation in plants' species, location and individual appearance, different training data is needed for respective applications. While this is also the case for others with multiple applications, the creation of new training data for plants is particularly difficult. The method of generating synthetic training data

employed in this study can be used in a variety of applications. Based on the initiative of this research, it is expected that it will be possible to go beyond the analysis of seeds, and accelerate the development of a machine learning model for the measurement of various plant phenotypes.

More information: Yosuke Toda et al, Training instance segmentation neural network with synthetic datasets for crop seed phenotyping, *Communications Biology* (2020). DOI: 10.1038/s42003-020-0905-5

Provided by Institute of Transformative Bio-Molecules (ITbM), Nagoya University

Citation: Training instance segmentation neural network with synthetic datasets for seed phenotyping (2020, April 24) retrieved 25 April 2024 from https://phys.org/news/2020-04-instance-segmentation-neural-network-synthetic.html

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.