ARTICLE IN PRESS

Engineering in Agriculture, Environment and Food xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Engineering in Agriculture, Environment and Food

journal homepage: www.elsevier.com/locate/eaef

Detecting greenhouse strawberries (mature and immature), using deep convolutional neural network

Harshana Habaragamuwa[a,*], Yuichi Ogawa[a], Tetsuhito Suzuki[a], Tomoo Shiigi[b], Masanori Ono[a], Naoshi Kondo[a]

[a] Graduate School of Agriculture, Kyoto University, Kitashirakawa-Oiwakecho, Sakyo-ku, Kyoto, 606-8502, Japan
[b] Department of Ocean Mechanical Engineering, National Fisheries University, 2-7-1, Nagata Honmachi, Shimonoseki, Yamaguchi, 759-6595, Japan

ARTICLE INFO

Keywords:
Deep learning
Deep convolutional neural network
Computer vision
Strawberry maturity

ABSTRACT

Existing agricultural detection algorithms mainly detect a single object category (class) under specific conditions which restricts the farmer's ability to utilize them in different conditions and for different classes. What is needed are generic algorithms that can learn to detect objects from examples, thereby reducing the technical burden required to adapt to local circumstances. Among generic algorithms, deep learning methods recently are beginning to outperform other generic algorithms. In this study, we investigate the possibility of using a deep learning algorithm for recognition of two classes (mature and immature strawberry) of agricultural product using a deep convolutional neural network (DCNN) and greenhouse images taken under natural lighting conditions. To the best of our knowledge, this is the first application of deep learning to the detection of mature and immature strawberries as two classes. We evaluated the results using the following parameters: average precision (*AP*), a parameter that combines detection success and confidence of detection; and bounding box overlap (*BBOL*) which measures localization accuracy. The developed deep learning model achieved an *AP* of 88.03% and 77.21%, and a *BBOL* of 0.7394 and 0.7045 respectively for mature and immature classes.

1. Introduction

Autonomous robots, of which more and more are appearing in agricultural settings, are smart machines that can carry out tasks without intensive human intervention (Bekey, 2005), such as navigation, recognition, etc. A crucial step for automation in the agricultural environment will be successful detection (identification and localization) of objects; a function carried out by object detection algorithms. To date most object detection algorithms though are task specific and developed under highly constrained artificial environments. This means that they are limited to identifying only one object class at a time under very specific conditions. Algorithms which can be readily expanded to detect other object categories in a complex, real world situation would provide much greater flexibility (Csurka et al., 2004). Thus, increasing the farmer's capacity to flexibly use these autonomous robots in different situations, for example: different field conditions or crop types.

For example, recognition algorithms in harvesting robots to detect strawberries (Hayashi et al., 2010), apples (Ji et al., 2012), corn tussles (Kurtulmuş, and Kavdir, 2014), and immature peach (Kurtulmus et al., 2013) have been recently developed. These algorithms are, however, specifically designed for a particular object using specific features of that object, such as color, roundness, or size. Although color information can provide faster running speed, it doesn't increase the accuracy of detection in mature strawberry (Xu et al., 2013).

Examples using this color information include: color segmentation and textural features for mango fruit identification (Payne et al., 2013); local texture features such as contrast entropy and a scale invariant key point (SIFT) to detect the texture of green citrus (Sengupta and Lee, 2014); color features and radial symmetry transform (Kurtulmus et al., 2013) to identify Region Of Interest (ROI); and color based features to detect mature strawberries for a strawberry harvesting robot (Rajendra et al., 2011). Such algorithms though are difficult to quickly adapt to new object classes, such as flowers, calyxes, or humans, and are also very sensitive to illumination conditions, so specific illumination conditions become a necessity.

This kind of specificity constraints the robot's performance by restricting its ability to understand the environment, thus preventing agricultural robots from achieving a higher degree of autonomy. Another disadvantage of having a task specific algorithm in an agricultural robot is that it will not be allow the robots capacities to be fully utilized. An example would be the hardware of a harvesting robot, which can harvest fruit, could be used to also thin out immature fruit,

* Corresponding author.
E-mail address: habaragamuwa.harshana.87z@st.kyoto-u.ac.jp (H. Habaragamuwa).

https://doi.org/10.1016/j.eaef.2018.03.001
Received 21 September 2016; Received in revised form 16 October 2017; Accepted 17 March 2018

Please cite this article as: Habaragamuwa, H., Engineering in Agriculture, Environment and Food (2018), https://doi.org/10.1016/j.eaef.2018.03.001

but if the identification algorithm can only detect mature fruit such capacity cannot be accessed. Such disadvantages suggest agricultural detection algorithms should be more biased towards generic detection algorithms which can be readily adapted to identify multiple classes using only some examples of that class taken under natural light conditions.

Generic algorithms can be divided into two classes: deep and shallow learning. Shallow algorithms are those with an architecture two to three layers deep, while deep algorithms are those that contain more than three layers of nonlinear operations (Bengio, 2009). Examples of shallow learning models include: a Support Vector Machine (SVM) used for fruit identification and classification (Song et al., 2014); a SVM classifier inside their algorithm to identify and count green citrus fruits in ambient light conditions (Sengupta and Lee, 2014); and a neural network with two hidden layers, SVM and other statistical classifiers to identify immature peach fruit (Kurtulmus et al., 2013).

A major disadvantage, however, with such shallow algorithms is that they need a large amount of data (Less Generalization) to obtain an accurate classification (Bengio, 2009). On the other hand, deep learning methods do not potentially suffer from this disadvantage. In the field of image recognition, deep learning methods have been shown to display significant performance improvements over shallow learning classifiers (Chatfield et al., 2014). This is in part because deep learning can represent the world as a hierarchy of features, for example: Pixels → Edges or blobs → Parts of Objects → Objects → Collections of Objects (Lee et al., 2009). Also, it can represent complex nonlinear factions due to their deep architecture. And it has the ability to use unlabeled or semi labeled data to achieve higher generalization.

One such deep learning method in particular, deep convolutional neural networks (DCNN), has recently been shown to be the most successful algorithm for image recognition in the field compared to deep Boltzmann machines, convolutional deep belief networks (Lee et al., 2009), and deep auto encoders (Kuang et al., 2014). Due to the success of DCNN in identification, many researchers have started to use DCNN in detection. The R-CNN (Regions with CNN features) (Girshick et al., 2014) algorithm combined DCNN and semantic segmentation to achieve state of the art results in detection.

Most of this research on deep learning, though, has been carried out using internet image databases, such as the PASCAL Visual Object Classes (VOC) Challenge (Everingham et al., 2010), IMAGENET Large Scale Visual Recognition Challenge 2012 (ILSVRC2012) (Russakovsky et al., 2014), Label me (Russell et al., 2007). These non-agricultural image databases contain images taken under a wide variety of situations and sources. More recently, Sa et al. (2016) studied a fruit detection system using deep neural networks. They used RGB and NIR images which they obtained themselves. They used sweet pepper and rock melon at different ripeness levels. Even though they used strawberry images those images were downloaded from the internet. Detection of strawberry classes (mature and immature) in known conditions, such as illumination, distance from camera, using deep learning remains unexplored at this point.

Many studies related to strawberry harvesting robots are being carried out in Japan, such as a machine vision algorithm for robots to harvest strawberries in tabletop culture greenhouses (Rajendra, 2008), and field operation of a movable strawberry-harvesting robot using a travelling platform (Hayashi et al., 2014). Our lab is also currently developing a strawberry harvesting drone. So, as our model fruit for this generic algorithm model we chose strawberry.

In strawberry field operations, many kinds of object classes (Mature, Immature, Flowers etc.) need to be detected in order to optimize field efficiency, an area of research yet to be explored for strawberry. Moreover, localization accuracy is important for robotics applications such as harvesting. So, in this initial study the behavior of a deep learning algorithm to detect strawberries using three classes (Mature, Immature and Background) was undertaken under known conditions. We selected mature and immature strawberry as our objects because for harvesting robots it is important to distinguish between these two classes clearly and, also for operations, such as pruning and yield estimation detection of immature strawberries from the background is important. We adopted the structure of appropriate deep learning method (R-CNN with only DCNN) for immature and mature strawberry detection and analyzed the results of detection.

Fig. 1. (a) Greenhouse view (b) Capturing images.

Our objective was to train and evaluate R-CNN: by its training behavior, classification accuracy, localization accuracy and the errors in detection of mature and immature fruit.

2. Materials and methodology

2.1. Materials and conditions

Pictures of strawberry were taken at a greenhouse near Matsuyama, Japan (Fig. 1(a)), from Feb. 04th to Feb. 06th 2015. The Akai Shizuku strawberry variety was used in the experiments. Average weight for this variety was 16.1 g/fruit, which is almost the same as the Amaotome or Benihoppe varieties, but this variety has smaller leaves, a short plant height compared to the above two varieties, its shape is uniform, and it has a good appearance (Matsuzawa et al., 2015). Because of these favorable characteristics we used the Akai Shizuku variety for our experiment. A total of 421 images were captured. Out of these 421 images, 48 images were removed uniformly (every 9th image until 48) for use in other experiments. Thus, we were left with 373 images. These images were used in the training and testing. From the 373 images, 551 mature and 923 immature fruits were identified and used in this experiment. The images were captured by a Nikon COOLPX S6500 camera. The camera was placed approximately 20 cm from the table of the strawberry (Fig. 1(b)). It was important to get as much variation in the images as possible. To reduce redundancy in the training data for this task, a lower ISO was selected. The camera had an original ISO range of 125–3200. The white balance of the camera was set using a standard whiteboard. The focal length of the camera varied from 4.5 to 54.0 mm and the f-number from 3.1 to 6.5; as the camera automatically selected the appropriate focal length and f-number within these ranges.

The camera was set to ISO 125, to get the lowest gain as possible so that the natural illumination conditions were not altered and noise was also reduced. The images were captured at a resolution of 4608 × 3456 pixel. Images were saved using the JPG format and resized to 500 × 375 to reduce computational burden. Each fruit was marked with a rectangle. The criteria of marking will be described in section 2.2.2. After resizing images, Table 1 shows the mean pixel sizes of the marked rectangles.

Light conditions were recorded using a Minolta CL-200A chroma meter, color temperature and illumination data was taken close to the

Table 1
Pixel sizes of marked rectangles.

Class	Mean height	Mean width	Standard deviation of height	Standard deviation of width
Mature	76.91	60.38	14.10	12.03
Immature	57.14	46.56	14.54	11.89