Computers in Industry 99 (2018) 17–28

Contents lists available at ScienceDirect

Computers in Industry

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

Apple flower detection using deep convolutional networks

Philipe A. Dias[a,*], Amy Tabb[b,1], Henry Medeiros[a]

[a] Marquette University, Department of Electrical and Computer Engineering, Milwaukee, WI, USA
[b] U.S. Department of Agriculture (USDA), Kearneysville, WV, USA

ARTICLE INFO

Article history:
Received 25 August 2017
Received in revised form 6 March 2018
Accepted 15 March 2018
Available online xxx

Keywords:
Bloom intensity estimation
Apple flower detection
Deep learning
Convolutional neural networks
Precision agriculture

ABSTRACT

To optimize fruit production, a portion of the flowers and fruitlets of apple trees must be removed early in the growing season. The proportion to be removed is determined by the bloom intensity, i.e., the number of flowers present in the orchard. Several automated computer vision systems have been proposed to estimate bloom intensity, but their overall performance is still far from satisfactory even in relatively controlled environments. With the goal of devising a technique for flower identification which is robust to clutter and to changes in illumination, this paper presents a method in which a pre-trained convolutional neural network is fine-tuned to become specially sensitive to flowers. Experimental results on a challenging dataset demonstrate that our method significantly outperforms three approaches that represent the state of the art in flower detection, with recall and precision rates higher than 90%. Moreover, a performance assessment on three additional datasets previously unseen by the network, which consist of different flower species and were acquired under different conditions, reveals that the proposed method highly surpasses baseline approaches in terms of generalization capability.

1. Introduction

Various studies have established the relationships between bloom intensity, fruit load and fruit quality [1,2]. Together with factors such as climate, bloom intensity is especially important to guide thinning, which consists of removing some flowers and fruitlets in the early growing season. Proper thinning directly impacts fruit market value, since it affects fruit size, coloration, taste and firmness.

Despite its importance, there has been relatively limited progress so far in automating bloom intensity estimation. Currently, this activity is typically carried out manually with the assistance of rudimentary tools. More specifically, it is generally done by inspecting a random sample of trees within the orchard and then extrapolating the estimates obtained from individual trees to the remainder of the orchard [3]. As the example in Fig. 1 illustrates, obstacles that hamper this process are: (1) manual tree inspection is time-consuming and labor-intensive, which contributes to making labor responsible for more than 50% of apple production costs [4]; (2) estimation by visual inspection is characterized by large uncertainties and is prone to errors; (3) extrapolation of the results from the level of the inspected trees to the row or parcel level relies heavily on the grower's experience; and (4) inspection of a small number of trees does not provide information about the spatial variability which exists in the orchard, making it difficult to develop and adopt site-specific crop load management strategies that could lead to optimal fruit quality and yield.

With the goal of introducing more accurate and less labor intensive techniques for the estimation of bloom intensity, machine vision systems using different types of sensors and image processing techniques have been proposed [5]. Most existing methods, which are mainly based on simple color thresholding, have their applicability hindered especially by variable lighting conditions and occlusion by leaves, stems or other flowers [6].

Inspired by successful works using convolutional neural networks (CNNs) in multiple computer vision tasks, we propose a novel method for apple flower detection based on features extracted using a CNN. In our approach, an existing CNN trained for saliency detection is fine-tuned to become particularly sensitive to flowers. This network is then used to extract features from portraits generated by means of superpixel segmentation. After dimensionality reduction, these features are fed into a pre-trained classifier that ultimately determines whether each image region

* Corresponding author at: Marquette University, Department of Electrical and Computer Engineering, 1515 W. Wisconsin Ave., Milwaukee, WI 53233, USA.
E-mail addresses: philipe.ambroziodias@marquette.edu (P.A. Dias), amy.tabb@ars.usda.gov (A. Tabb), henry.medeiros@marquette.edu (H. Medeiros).

[1] Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture. USDA is an equal opportunity provider and employer.

https://doi.org/10.1016/j.compind.2018.03.010
0166-3615/

Fig. 1. Example of image from a flower detection dataset used in this paper.

contains flowers or not. The proposed method significantly outperformed state-of-the-art approaches on four datasets composed of images acquired under different conditions.

Our main contributions are:

(1) a novel CNN-based flower detection algorithm;
(2) an extensive evaluation on a challenging dataset acquired under realistic and uncontrolled conditions;
(3) an analysis of the generalization capability of the proposed approach on additional datasets previously unseen by the evaluated models.

The remainder of paper is organized as follows. Section 2 discusses the most relevant existing approaches for automated flower and fruit detection. Our proposed approach is described in Section 3, which also includes a description of three baseline comparison methods as well as implementation details. Experiments performed to evaluate the impact of specific design choices are described in Section 4, followed by an extensive comparison of our optimal model against the baseline methods on four different datasets. Our concluding remarks are presented in Section 5.

2. Related work

While existing techniques employed for flower detection are based only on color information, methods available for fruit quantification exploit more modern computer vision techniques. For this reason, in this section we first review the most relevant works on automated flower detection, followed by a discussion of the relevant literature on fruit quantification. Moreover, to make this article self-contained and therefore accessible to a wider audience, we also provide a brief introduction to the fundamentals of CNNs.

2.1. Flower and fruits quantification

Aggelopoulou and colleagues presented in [7] one of the first works using computer vision techniques to detect flowers. That method is based on color thresholding and requires image acquisition at specific daylight times, with the presence of a black cloth screen behind the trees. Thus, although its reported error in predicted yield is relatively low (18%), such approach is applicable only for that controlled scenario.

Similar to the work of Thorp and Dierig [8] for identification of *Lesquerella* flowers, the technique described by Hočevar et al. [9] does not require a background screen, but it is still not robust to changes in the environment. The image analysis procedure is based on hard thresholding according to color (in the HSL color space) and size features, such that parameters have to be adjusted whenever changes in illumination (daylight/night), in flowering density (high/low concentration) or in camera position (far/near trees) occur.

Horton and his team described in [10] a system for peach bloom intensity estimation that uses a different imaging approach. Based on the premise that the photosynthetic activity of this species increases during bloom period, the system relies on multispectral aerial images of the orchard, yielding an average detection rate of 84.3% for 20 test images. Similarly to the aforementioned methods, the applicability of this method also has the intrinsic limitation of considering only color/spectral information (thresholding near-infrared and blue bands), such that its performance is sensitive to changes in illumination conditions.

More advanced computer vision techniques have been employed for fruit quantification [5]. A multi-class image segmentation for agrovision is proposed by Hung et al. [11], classifying image pixels into leaves, almonds, trunk, ground and sky. Their method combines sparse autoencoders for feature extraction, logistic regression for label associations and conditional random fields to model correlations between pixels. Some other methods are based on support vector machine (SVM) classifiers that use information obtained from different shape descriptors and color spaces as input [12,13]. Compared to existing methods for flower detection, these methods are more robust since morphological characteristics are taken into account. As many other shape-based and spectral-based approaches [14–17], these techniques are, however, still limited by background clutter and variable lighting conditions in orchards [3].

Recent works on fruit quantification include the use of metadata information. Bargoti and colleagues [18] built on [11] to propose an approach that considers pixel positions, orchard row numbers and the position of the sun relative to the camera. Similarly, Cheng et al. [19] proposed the use of information such as fruit number, fruit area, area of apple clusters and foliage area to improve accuracy of early yield prediction, especially in scenarios with significant occlusion. However, the inclusion of metadata is highly prone to overfitting, particularly when limited training data is available and the variability of the training set is hence low [18].

2.2. Deep learning

Following the success of Krizhevsky's model [20] in the ImageNet 2012 Challenge, deep learning methods based on CNNs became the dominant approach in many computer vision tasks. The architecture of traditional CNNs consists of a fixed-size input, multiple convolutional layers, pooling (downsampling) layers and fully connected layers [21]. Winner of the ImageNet 2013 Classification task, the Clarifai model is one such network [22]. Illustrated on the right side of Fig. 2, it takes input image portraits of size 227×227 pixels, which traverse a composition of 5 convolutional layers (C1–C5) and 3 fully connected layers (FC6–FC7 and the softmax FC8). Each type of layer plays a different role within the CNN architecture: while convolutional layers allow feature extraction, the latter fully connected layers act on this information to perform classification.

In computer vision and image processing, a *feature* corresponds to information that is meaningful for describing an image and its regions of interest for further processing. Feature extraction is therefore crucial in image analysis, since it represents the transition from pictorial (qualitative) to nonpictorial (quantitative) data representation [23]. Rather than relying on hand-engineered features (e.g., HOG [24]), deep CNNs combine multiple convolutional layers and downsampling techniques to learn hierarchical features, which are a key factor for the success of these models [25]. As described in [22], the convolutional layers C1–C2 learn to identify low-level features such as corners and other edge/color