Neurocomputing 000 (2018) 1–15

Contents lists available at ScienceDirect

Neurocomputing

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

Multi-class fruit detection based on image region selection and improved object proposals

Hulin Kuang[a,*], Cairong Liu[b], Leanne Lai Hang Chan[c], Hong Yan[c]

[a] Department of Clinical Neurosciences, University of Calgary, Alberta, Canada
[b] Department of Mathematics, Chinese University of Hong Kong, Tai Po Road, Shatin, NT, Hong Kong, China
[c] Department of Electronic Engineering, City University of Hong Kong, 83 Tat Chee Avenue, Kowloon, Hong Kong, China

ARTICLE INFO

Article history:
Received 30 October 2015
Revised 4 September 2017
Accepted 31 December 2017
Available online xxx

Communicated by Prof. Zidong Wang

Keywords:
Object detection
Multi-class fruit detection
Image region selection
Score-level multi-feature fusion
Object proposal

ABSTRACT

This paper proposes a novel approach for multi-class fruit detection using effective image region selection and improved object proposals. Five complementary features, namely local binary patterns (LBPs), histograms of oriented gradient (HOGs), LBP based on magnitude of Gabor feature (GaborLBP), global color histograms, and global shape features, are utilized to improve the detection accuracy. An optimal combination of regions (*i.e.*, features) is selected using an image region selection method based on feature similarity and cross-validation accuracy. To combine the strength of the five complementary features, a weighted score-level feature fusion approach based on the average confidence coefficient is used. Moreover, during detection, an object proposal method, "EdgeBoxes," is improved by calibrating scores considering the image region similarity to generate windows that are likely to contain fruits and to speed up detection. The experimental results show that the image region selection method can select an effective and optimal combination of regions, which exhibits better recognition accuracy than the method without image region selection. This proposed method demonstrates a low miss rate (0.0377) at 0.0682 false positives per image (FPPI) and outperforms some baselines: multi-class fruit detection using the traditional sliding window mechanism, the well-known deformable parts model (DPM) method, convolutional neural networks features (CNN) with support vector machine (SVM) for classification (CNN + SVM), cascade detection framework and faster RCNN in terms of the miss rate *vs.* FPPI and precision *vs.* recall curves. The proposed multi-class fruit detection can detect multi-class and multiple fruits in a variety of sizes, backgrounds, angles, locations, and image conditions.

1. Introduction

Object detection, *i.e.*, detecting the location and identifying the type of an object in an image, is a challenging task and has become one of the most important areas in computer vision and pattern recognition. Object detection has made impressive progress in the recent decades. Several global grand challenges are held to evaluate algorithms for object detection and image classification on a large scale, *e.g.*, the Pascal Visual Object Classes challenge [1] and ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [2]. Recently, most state-of-the-art methods in object detection have focused on feature descriptors and machine learning with a sliding window framework [3–7]. Further, various technologies have been utilized to increase the detection speed and improve the detection performance, such as fast rejection based on a simple linear classifier [7,8] before detection using sophisticated classifiers, and object proposals [9–11].

Fruit recognition/detection is a subfield of object detection. Fruit recognition refers to the recognition of a class of fruit within an image in which there is only one fruit. Fruit detection not only recognizes the class of each fruit but also localizes each fruit within an image, and is thus similar to face detection and pedestrian detection. In fruit detection, many classes of fruit, *e.g.*, red apple, pear, and orange, are considered. Therefore, multi-class fruit detection should not only distinguish all classes of fruit from backgrounds but also differentiate each class of fruit from the other classes, which is more difficult than other object detections, which merely distinguish all classes of objects from the background.

Fruits play an indispensable role in our everyday life and are categorized as healthy food. However, the sugar in fruits has an adverse impact on diabetic individuals. An automatic fruit recognition system with the information dataset of each fruit can help us

* Corresponding author.
E-mail addresses: hulinkuang@gmail.com (H. Kuang), crliu@math.cuhk.edu.hk (C. Liu), leanne.chan@cityu.edu.hk (L.L.H. Chan), h.yan@cityu.edu.hk (H. Yan).

https://doi.org/10.1016/j.neucom.2017.12.057
0925-2312/

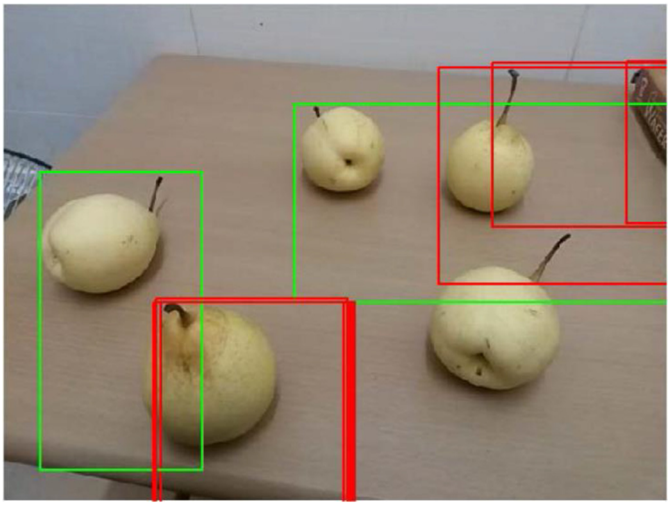

Fig. 1. Examples of the top-10 object proposals obtained by EdgeBoxes.

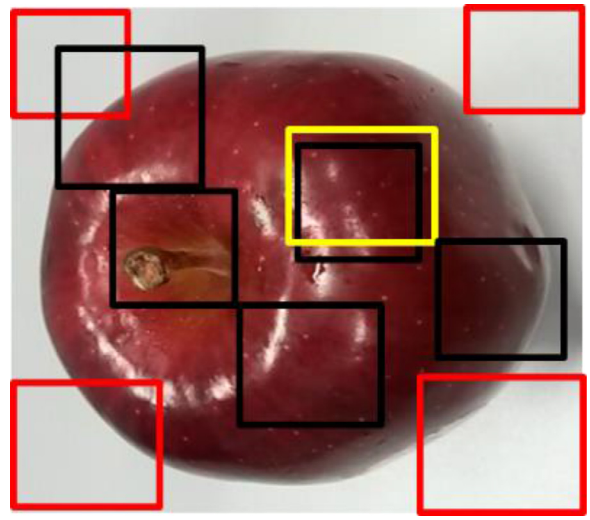

Fig. 2. Examples of meaningless regions and useful regions for object recognition.

to select fruit that is suitable for us. Meanwhile, such a system can help educate children or visually impaired individuals and familiarize them with fruits. In supermarkets and grocery stores, such a system can be utilized to provide customers with the options of self-service and self-checkout. A fruit detection system that can localize a certain class of fruit has been utilized for automating fruit harvester machines in large gardens and greenhouses [12–14].

Recently, researchers have focused on fruit recognition/detection. State-of-the-art methods for fruit recognition are based on feature descriptors and machine learning [15–26] and consider a variety of efficient features such as simple statistical color features, global color histograms, shape or morphological features, and texture features. Although good performance has been reported, optimal features are not selected in these methods, and sophisticated features such as local binary patterns (LBPs) [27] and histograms of oriented gradient (HOGs) [3] are seldom utilized in the current fruit recognition methods. In addition, multiple features used in present fruit recognition are combined by concatenating one feature after another [17,21] or using PCA to construct a new feature vector [22]. These types of feature fusion might lead to misclassification [24]. Most of the fruit detection methods are focused on a single class [12–14,28] and have been applied to fruit harvesters in gardens or greenhouse scenes [12–14,28,29]. However, these systems lack multi-class fruit detection. In these fruit detection methods, the locations of fruits are obtained by image segmentation, edge detection, and background reduction, which are sensitive to the background. Thus, more powerful fruit detection and recognition techniques need to be developed.

Object proposal methods have been widely and successfully used for reducing the computational cost of object detection [9–11,30–33]. In general, "objectness" (score) of each proposal (window), *i.e.*, the likeliness of each window to be an object, is computed, and then, a set of top-score windows after sorting the scores of windows from high to low is input into the classifiers. By observation, we find that the scores of proposals might be inaccurate, as shown in Fig. 1. In this figure, we show the top-10 object proposals obtained by EdgeBoxes [30]. Despite of the simple background, some pears are not included in these top-10 proposals with a high overlap and two pears are at the corner of the proposals denoted in green. Therefore, scores obtained by EdgeBoxes is calibrated by adding the image region similarity between the center region and the four corner regions within the windows. In addition, HOG, LBP, and GaborLBP are extracted from variable-size blocks (*i.e.*, regions) in this study. Hereafter, we use "region" to represent "block". However, there exist some meaningless regions, *e.g.*, regions in corners containing only the background (red rectangles in Fig. 2), that could lead to feature redundancy and increase feature dimension. In Fig. 2, the black regions contain parts of fruits or edges of fruits and have relatively similar features that are useful for object recognition. The two regions with a high overlap (yellow region and the black region close to it) have very similar features, and this might lead to feature redundancy. A novel image region selection method based on feature similarities and considering the similarity of feature vectors and the overlap of two regions along with cross-validation accuracy is proposed in this paper in order to select an optimal combination of regions for the extraction of optimal HOG, LBP, and GaborLBP features. Besides, in this study, we extract five complementary features to represent most attributes of fruit. These complementary features are combined at score-level by summing score vectors of each feature (*i.e.* classifier) with weights to improve recognition performance. Finally, after detection a fruit might be surrounded by two or more bounding boxes in different classes. To deal with these cases, a post-processing step based on the overlap and the score is utilized, which is complementary to non-maximum suppression (NMS) [34]. The framework of the proposed method is shown in Fig. 3.

To summarize, in this work, we make the following contributions: (1) A novel image region selection method is proposed to select an effective and optimal combination of regions, which can reduce the feature dimension and improve the accuracy for fruit recognition. (2) The original EdgeBoxes method is improved by constructing a new calibrated score function by combining the image region similarity with the scores obtained by EdgeBoxes with learnt weights to extract more accurate proposals. (3) Because no universal dataset is used in the fruit detection, a new fruit dataset containing five classes of fruit is built for the multi-class fruit detection task.

The remainder of this paper is organized as follows: Related work is described in Section 2. Section 3 provides an overview of the proposed multi-class fruit detection. The proposed image region selection method is introduced in detail in Section 4. Section 5 demonstrates the score-level feature fusion based on the average confidence coefficient. The details of improved object proposals are presented in Section 6. In Section 7, the experimental settings and the evaluation results are given. Finally, the conclusions are discussed in Section 8.

Please cite this article as: H. Kuang et al., Multi-class fruit detection based on image region selection and improved object proposals, Neurocomputing (2018), https://doi.org/10.1016/j.neucom.2017.12.057