Deep learning for robotic strawberry harvesting

An Application Research

Xiaodong Li, Charles Fox & Shaun Coutts

School of Computer Science, University of Lincoln

doli,chfox,scoutts@lincoln.ac.uk

Abstract

A novel R&D of machine learning based robotic strawberry vision system of fruit counting, sizing/weighting, etc. is presented for harvesting management.

Introduction

Strawberries are a high-value crop all around world, it is a challenge for farmers to efficiently manage labor and transport,etc in harvest season due to the weather fluctuations, with heavenly time consuming and labor. Machine vision has become a potential an alternative to the traditional management.

ML based image recognition and classification have high adaptivity to variant image quality.

SVM [5] is impressive under non-linear model, but unable to meet real time requirements and hard to train.

As first practical real time (CPU-based) face detector, Viola-Jones cascade detector [1] is well supported with both Haar-like feature [1] and Local Binary Patterns(LBP) feature [7] [6].

To expand and enhance existing functionality, e.g, ripeness prediction, based on existing traing data, quick prototype development - YOLO [2] [4] is utilized for both classification and refining detection.

System Design and Methodology

1. System are shown in Figure.1.

Figure 1: System Structure

2. Viola-Jones Cascade Detector.
   - Harr-like detector, using a sliding window, calculate and compare with the difference from a learned threshold.
   - AdaBoost, a classifier cascade is used to obtain stronger classifiers by combining single ones.
     A single classifier consists of a weighted sum of many weak classifiers, where each weak classifier is a threshold on a single Haar-like rectangular feature.
     The weight associated with a given sample is adjusted based on whether or not the weak classifier correctly classifies the sample.
   - LBP Feature [7] [6] is an alternative to Harr-like features which is still infeasible on real time with large image.
     A LBP vector can be simply calculated in an image cell (e.g 16x16 pixels of sub-window), where the pixel is compared to each of its 8 neighbors along a circle. The pixel's value is the concatenation of a binary "0" and "1", which is assigned by comparing (less or greater) with each neighbor. This 8-digit binary number is usually converted to decimal for convenience.
     The histogram over the cell for the frequency of each "number" occurring is computed, which can be regarded as a 256-dimensional feature vector.
3. Deep Learning–YOLO [4] [2] is one of the most effective accurate real time object detection algorithms.
   It applies a single neural network to the full image, and then divides the image into regions and predicts bounding boxes and probabilities for each region.
   These bounding boxes are weighted by their predicted probabilities, which output the recognized object after non-max suppression applied.
   Small size of strawberry image (24x24 pixels) in Viola-Jones can not provide enough information for precise classification with Haar/LBP feature.
   Self-trained YOLO model is for color-based object classification to predict the ripeness of strawberries simply by its color category.
   No relabeling needed, the effective and innovative integration of YOLO with existing system is to be achieved.
4. Data Collection and Annotation.
   - Typical Images & Sampling
     Original Figure.2 (Left: indoor,Right: outdoor) and cropped image are in Figure.3.

Figure 2: Typical Strawberry Images

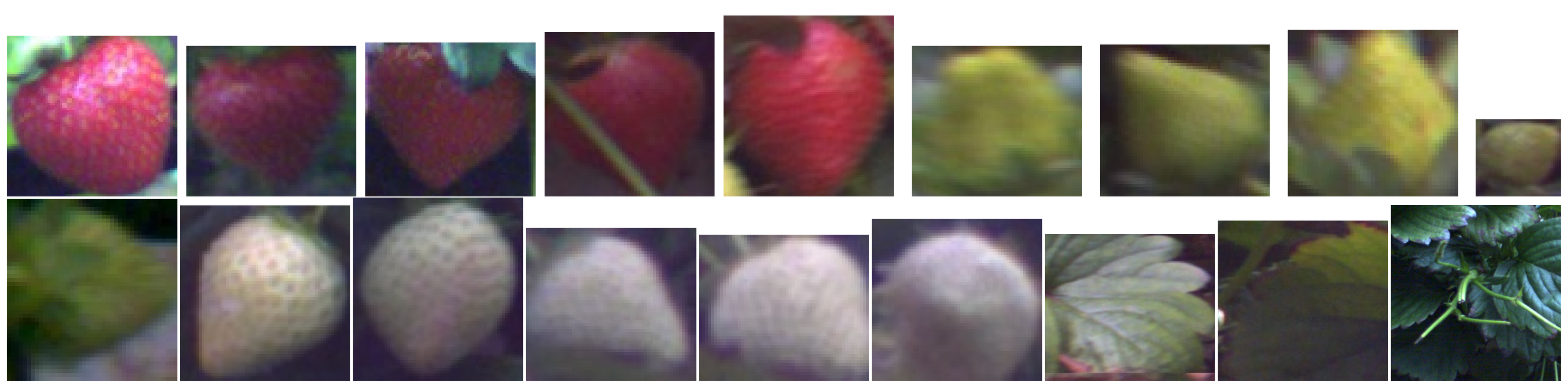

Figure 3: Strawberry Training Images

- Image Annotation
  Inspired by insight into YOLO mechanism, without standard YOLO manual relabeling, an automatic 'one4one' annotation strategy is applied. The labels are formatted as: $center-x = x/w(0.5); center-y = y/h(0.5); width = w/W(1.0); height = h/H(1.0);$
  x: x-coordinate of center of the bounding box;
  y: y-coordinate of center of the bounding box;
  w: width of the bounding box;
  h: height of the bounding box;
  W: width of the whole image;
  H: height of the whole image;

Results

The snapshots shown in Figure.5 include configuration/monitoring, detection/counting, classification, sizing and weighting (displayed in: diameter(mm), weight(gram),etc.

Figure 5: System Features

Summarization

- Pioneer R&D (C/C++) achieved is encouraging, overall hit the goal.
- Tested on a 80 meters strawberry track, compared with manual counting, classification, sizing and weighting , 90% accuracy of counting, and 80% of classification with a model trained on 300 samples for each category.
- Further improvement is to be achieved by training with more sample images.
- Sizing, weighting,etc are under improvement due to inaccurate depth assumption, shape model (only cylinder used) and density setting,etc.

References

[1] Paul Viola , Michael Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features", IJCV 2001.

[2] Joseph Redmon, Ali Farhadi, "YOLOv3: An Incremental Improvement", Computer Vision and Pattern Recognition, April, 2018.

[3] M. Mahmud, M. S. Kaiser, A. Hussain and S. Vassanell, "Applications of Deep Learning and Reinforcement Learning to Biological Data", IEEE Transactions on Neural Networks and Learning Systems, vol. 29, no. 6, pp. 2063-2079, June 2018.

[4] Joseph Redmon, Ali Farhadi, "YOLO9000: Better, Faster, Stronger", Computer Vision and Pattern Recognition, December, 2016.

[5] N. Cristianini , J. Shawe-Taylor, "An introduction to support Vector Machines: and other kernel-based learning methods," , Cambridge University Press, 2000.

[6] L. Wang , DC. He, "Texture Classification Using Texture Spectrum,"IEEE Transactions on Pattern Recognition,vol. 23, pp.905–910, 1990.

[7] DC. He , L. Wang, "Texture Unit, Texture Spectrum, And Texture Analysis," IEEE Transactions on Geoscience and Remote Sensing,vol. 28, pp.509–512, 1990.