AAI 800 Project proposal

Research for Object Detection Based on Deep Learning

Yarong Liu

1 Introduction

Computer vision is a field of artificial intelligence that trains computers to interpret and understand the visual world. Therefore detecting object fast and accurately is the basic and first tasks for computers to deal with images. Object detection has received much attention and achieved great success in the last several years, but considering the diversity of conditions, scenes and resolution of pictures and different shapes of objects shown in the pictures, it is still a very challenging task with great potential and space for improvement in the field of computer vision.

This project will implement different object detection algorithms to train different models, which are used to detect kinds of fruits and vegetables in the images. Comparing the accuracy of these models and get thorough comprehension of these algorithms are also the point of this project.

2 Source of Datasets

The whole datasets was downloaded from Kaggle. It contains 82213 images, 61488 images are training set, the size of test set is 20622 images. Other than that, 103 images is multi-fruits, in which there are kinds of fruits and vegetables. Every image is 100*100 pixels.

3 Implementation Plan for Project

After reading some papers about object detection in the preliminary preparations of project, I find three feasible algorithms to detect object.

1) **Single Shot MultiBox Detector**

   SSD is a method for detecting objects in images using a single deep neural network. It discretizes the output space of bounding boxes into a set of default boxes over different aspect ratios and scales per feature map location. At prediction time, the network generates scores for the presence of each object category in each default box and produces adjustments to the box to better match the object shape. Additionally, the network combines predictions from multiple feature maps with different resolutions to naturally handle objects of various sizes.

2) **R-CNN**

   The R-CNN detector first generates region proposals using an algorithm such as Edge Boxes. The proposal regions are cropped out of the image and resized. Then, the CNN classifies the cropped and resized regions. Finally, the region proposal bounding boxes are refined by a support vector machine (SVM) that is trained using CNN features.

3) **CNN combine Adversarial Learning**

   A cascaded CNNs model is first designed as a generator G, which consists of an encoder-decoder network for global saliency estimation and a deep residual network for local saliency refinement. It is hard to explicitly learn such structural information due to the limitation of frequently-used pixel-wise loss functions. Instead, a discriminator D is then designed to distinguish the real salient maps from the

fake ones produced by G, based on which an adversarial loss is introduced to optimize G. G and D are trained in a fully end-to-end fashion by following the strategy of Conditional Generative Adversarial Networks to make G well learn the structural information. At last, G is able to produce high quality salient maps without requiring any post-process to fool D.

4   Expect Result

Using these object algorithms above to achieve model that can detect multi-fruits images fast and accurately. If i have enough time, I will try to train a model which can detect object successfully and can count how many different object on the images.