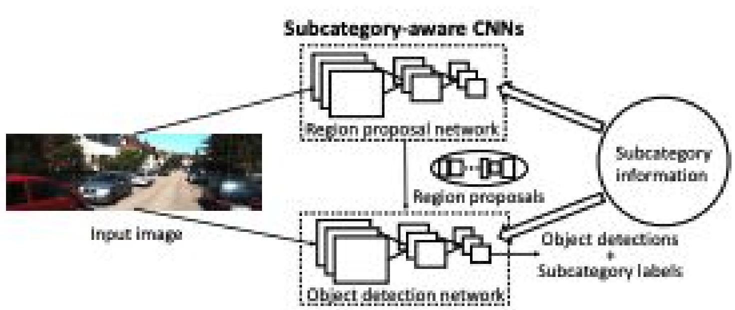
Types of Fruits identification with Convolutional Neural Networks

Xin Liu Qing Hu
Northeastern University
{liu.xin10, hu.qing}@northeastern.edu

Abstract
Deep Convolution Neural Networks (CNNs) have shown impressive performance in various vision tasks such as image classification, object detection and semantic segmentation. For object detection, particularly in still images, the performance has been significantly increased last year thanks to powerful deep networks (e.g. GoogleNet) and detection frameworks (e.g. Regions with CNN features (RCNN))[1]. In this work, we modify the input dataset like autoencoder technology and change different parameters and combined different layers of CNN model to find the relationship between those changes and the model performance (accuracy and loss rate).

1.Introduction
Deep learning has been widely applied to various computer vision tasks such as image classification, etc. As the amount of image and video information available increases, robust, configurable object detection systems for managing this data will become indispensable. There has been an explosion in the amount of information presented on the Internet as it quickly transitions from a text-based medium to one of image and video content; object detection systems will be used to search through the growing number of image and video databases. This technology will also be used in surveillance applications, driver assistance systems, and as front ends to recognition systems.
Convolutional Neural Networks (CNNs) have become dominating in solving different recognition problems recently. CNNs are powerful due to their capability in both representation and learning. With millions of weights in the contemporary CNNs, they are able to learn much richer representations from data. In object detection, we have witnessed the performance boost when CNNs are applied to commonly used benchmarks such as PASCAL VOC and ImageNet.[2]



This paper addresses the problem of object and pattern detection in static images.

2.Related Works
CNN-based Object Detection.
We can categorize the state-of-the-art CNN-based object detection methods into two classes: one-stage detection and two-stage detection. In one-stage detection, such as the Overfeat framework, a CNN directly processes an input image, and outputs object detections. In two-stage detection, such as R-CNNs, region proposals are first generated from an input image, where different region proposal methods can be employed. Then these region proposals are fed into a CNN for classification and location

refinement. It is debatable which detection paradigm is better. We adopt the two-stage detection framework in this work, and consider the region proposal process to be the coarse detection step in coarse-to-fine detection. We propose a novel RPN motivated by and demonstrate its advantages.[3]

3. Method

In this section, we will introduce the dataset, preprocessing, data split and the model.

3.1 Dataset

a. 79 kinds of fruits
b. 65000+ images
c. 100 * 100 pixels

3.2 Preprocess data & dataset split

Step 1. Split all data into 3 different sets. Train : Validation : Test = 60% : 15% : 25%

Step 2. Merge similar files into one file

e.g. Apple Golden 1, Apple Golden 2, Apple Golden 3 => Apple Golden

Step 3. Mapping from String to Number

e.g. Apple Golden => 2

Step 4. Normalization picture data

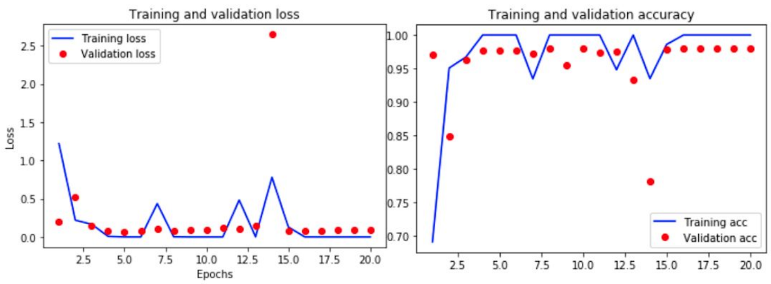
3.3 Model

a. We use the conv2D function which windows are 3*3. After we use 2 times, the final feature map has 96*96*32 = 294912 total coefficient, which is huge.
b. keras.layer.MaxPooling2D(pool_size = (2,2))
c. we choose Max-pooling to reduce the size of the feature maps. After we make the pool_size parameter equals to 2*2, the feature map is halved. Whenever we add a layer of conv2D, we will add another Max-pooling layer below.
d. The fruit classification is a multi-class, single-label classification, so we find the best last activation is softmax(if the problem is binary or multi-class, multil-label classification, sigmod is better). And the loss function is categorical-crossentropy.

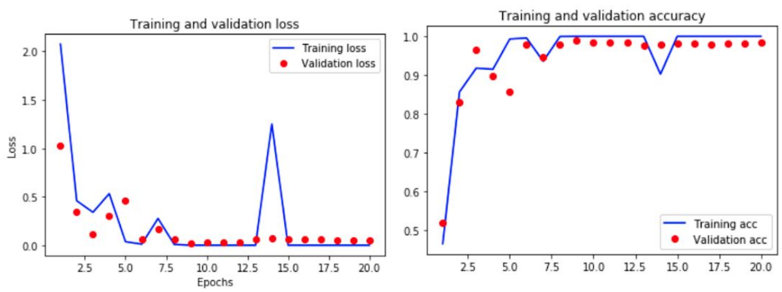
4.Result

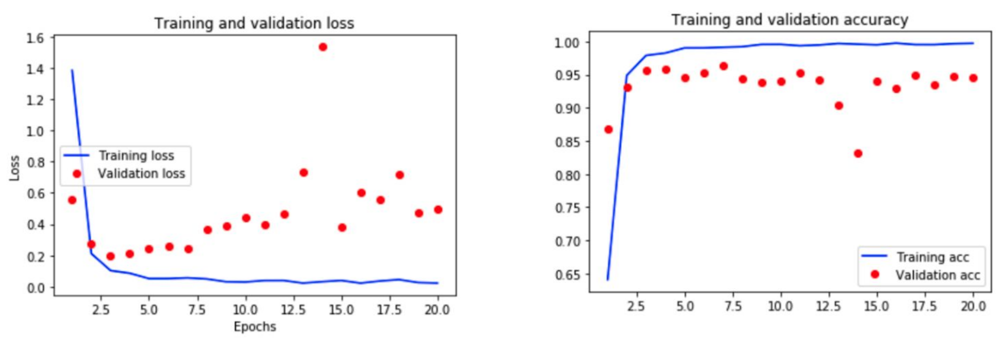
We use matplot to draw the loss value and accuracy. From the output of chart below, the model overfit after 11 epochs, so we tentatively decide 11 is a good epoch times in our 5000 training and valid set. The final epochs will be decided by all of the train and valid sets, around 40000 images.

Padding = valid
Means no
padding



Padding = same
Means that the
output has the
same length as
the original input





Performance Evaluation

Accuracy:

```
scores = model.evaluate(test_set, test_labels)

print ("Accuracy: %.2f%%" %(scores[1]*100))
```

```
16421/16421 [==============================] - 105s 6ms/sample - loss: 0.
2192 - acc: 0.9719
Accuracy: 97.19%
```

5.Conclusion

In this work, we propose a complete CNN model for fruit classification. The framework efficiently and accurately classifies 79 types of fruit.

Acknowledgment

This work is completed by Xin Liu and Qing Hu.

References

[1] K. Kang, W. Ouyang, H. Li, and X. Wang. Object detection from video tubelets with convolutional neural networks. In CVPR, 2016
[2] CONSTANTINE PAPAGEORGIOU AND TOMASO POGGIO. A Trainable System for Object Detection. 2000
[3] Yu Xiang, Wongun Choi, Yuanqing Lin, and Silvio Savarese. Subcategory-aware Convolutional Neural Networks for Object Proposals and Detection.