Robotics Nanodegree
Making currency accessible to the visually impaired with robotics inference

Benjamin Ong

Abstract

This project is part of the Robotics Nanodegree Programme and the objective is to prototype a vision inference system used on robots to help the visually impaired recognize different cash denominations. In this project, a dataset of different cash denominations ($2, $5 and $10) was used to train a deep convolutional neural network that would be used on a robot for cash denomination recognition. The model achieved an accuracy of more than 98% after 30 epochs and an inference time of around 5ms.

Introduction

Paying for items using cash is a task that is simple and familiar to many. Yet, this simple act has been a constant struggle for the visually impaired. Without a good sense of sight, the visually impaired risk being cheated and often have to rely on the honesty of the cashier. Many visually impaired often have to resort to applying different strategies to recognize cash notes, such as feeling for a unique texture and to estimate the size of the note by folding. However, with countries often redesigning their currency, many of these strategies tend to fall short.

With the mission to make currency accessible to the visually impaired, this project aims to create a deep learning model that can be used to recognize different cash denominations. A robotic system can be built for the visually impaired to allow them to verify the cash denominations that they have received. Using the GoogLenet convolutional neural network, the deep learning model was trained to recognize the different cash denominations in Singapore – namely the $2, $5 and $10 cash denominations.

Background

The NVIDIA Digits environment was used in this project for training, validation and testing of the convolutional neural network.

GoogLeNet model was chosen for its effectiveness for image classification. In the project, GoogLeNet network was used with Stochastic Gradient Descent with 30 epochs and a starting learning rate of 0.01. A separate test dataset of 20 images for each class was used for testing the models accuracy. The speed of each individual classification run was also evaluated.

Data Acquisition

400 images were collected with approximately equal number of images for the following classes:

1. $2 dollars note
2. $5 dollars note
3. $10 dollars note
4. Background of images used

Using different backgrounds, 100 photos of each cash denominations were being captured and compressed. There were a total of 400 photos being captured - 100 photos of SGD $2, 100 photos of SGD $5, 100 photos of SGD $10 and 100 photos of the varied backgrounds that were used in the photos of the different cash denominations.

To ensure that the trained model would be able to recognize both the front and back of the cash notes, images of the front and back of each cash denomination were also captured in approximately equal proportions.

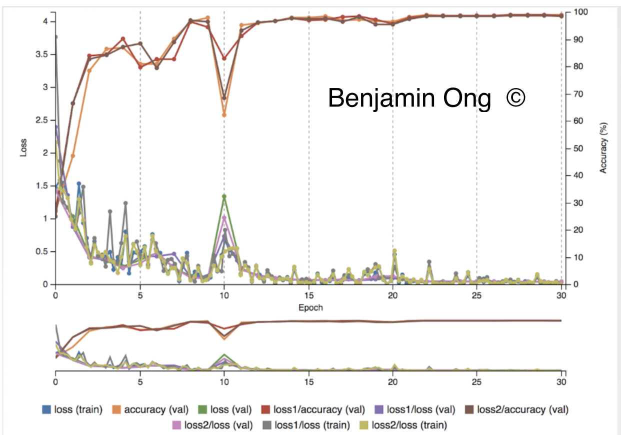
Data Preparation and Augmentation

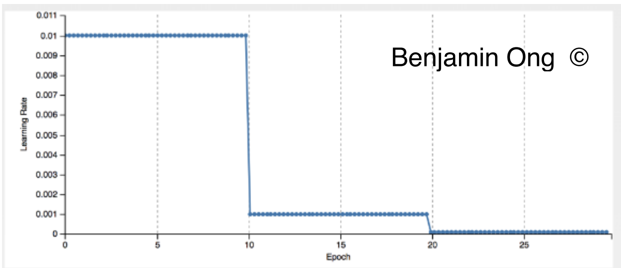
The images were first cropped to RGB colored images of size 256 by 256. To increase the training size and to improve the accuracy of the model, the amount of data was increased by rotating the images by 90, 180 and 270 degrees. This resulted in 400 images for each class.

Three different datasets for training, testing and validation were then created. For each class, there were 280 images used for training (70%), 100 images used for validation (25%) and 20 images used for testing (5%).

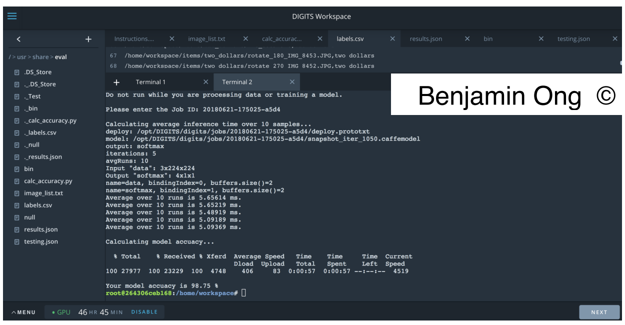
Results

The model was first trained using Google LeNet with Adam Optimizer. However, the accuracy was poor after the first 10 epochs. Google LeNet with SGD Optimizer was then used and the model performed well and achieved a good accuracy.



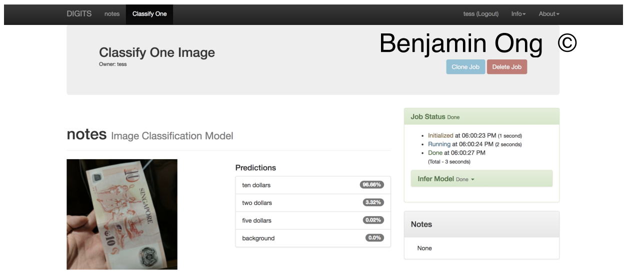


It can be observed that the training loss decreased significantly after 10 epochs. The learning rate was also decreased from the 10 epochs onwards.



Confusion matrix

	background	five dollars	ten dollars	two dollars	Per-class accuracy
background	19	0	0	1	95.0%
five dollars	0	20	0	0	100.0%
ten dollars	0	0	20	0	100.0%
two dollars	0	0	0	20	100.0%



Text

As seen from the figure above, the model was able to predict the image to be a ten dollar note with high confidence of more than 96%. Overall, the trained model was able to achieve an accuracy of 98.75% correct classification of all the different cash denominations after 30 epochs. The inference time was around 5ms.

Discussion

The classification accuracy is more than 98% and exceeds the given requirements. The average evaluation time is approximately 5ms, which also meets the requirement for evaluation time. Using data augmentation methods, it helped the model to be more robust to the differences in images and contributed to its high classification accuracy.

The high classification accuracy of above 98% could be due to the fact that the 3 different cash denominations are each of different colors (red, purple and green). In future experiments, less common denominations could also be used to train the model (e.g $1000 dollar note which is also red).

Future Work

The project achieved better than expected results with accuracy level above 98% after 30 epochs, which falls above the rubric requirements for the project. Using this model, it can be used to build a robot that can help to guide the visually impaired, with one of its features being that of helping the visually impaired recognize and count money.

Future work would include counting other cash denominations such as coins as well as the less common money denominations such as $20, $50, $100 and $1000. Photos of cash denominations with more varied backgrounds can also be used to further improve the robustness of the model.

References

Das, S. (2017, November 16). CNN Architectures: LeNet, AlexNet, VGG, GoogLeNet, ResNet and more. Retrieved June 24, 2018, from https://medium.com/siddharthdas_32104/cnns-architectures-lenet-alexnet-vgg-googlenet-resnet-and-more-666091488df5