Robotic Inference Using Nvidia Digits

Karim Hassanieh

Abstract—The following technical report details the steps & methods used to complete Robotics Inference Project. The main purpose of the project is to implement a neural network which can be used in real life vision applications. Two classification applications are implemented and detailed in below report.The first application classifies images into 3 classes : bottles, candy boxes and nothing.The dataset provided by Udacity RoboND team is used to train the Google LeNet network model. The trained model obtained after training is evaluated and provides an accuracy of 75.4 % with an average runtime of 5ms. The second application classifies images which contain open doors,closed door, and no door. The trained model is successfully capable of classifying the status of the door with high accuracy.

Index Terms—Robotics, IEEE, Computer Vision, Classification, Deep Learning.

1 Introduction

Large databases and advances in computing power during the 2000's had given the machine learning research community tools to further explore ideas which were not feasible in the past [1]. Special interest was given to solve computer vision applications such as image classification, semantic segmentation, and depth estimation using monocular cameras. As a result the need to give to neural networks further depth and data was essential to solve such stated problems. This led to the rise of state of the art neural networks such as Alex Net [2] , Seg Net [3] and Google LeNet [4].

Building on such advanced and deep neural networks we preform image classification in this report using deep nueral networks on two different real life applications. The first application classifies images between Bottles, Boxes, and No object. The second application classifies images which contains closed doors, open doors, and no doors.

1.1 Bottle - Candy Box Classification

In the first part of the project. We intend to solve the problem of classifying objects between bottles and boxes through images obtained. This can be of benefit in factories which are in need of a method to sort the different material passing through. A simple example could be material passing through on a conveyor belt for processing.

1.2 Open - Close Door Classification

In the second part of the project we try to use deep learning with deep neural networks to solve the problem of classifying weather a door is available in there first place and if it is, if it is open or not. This application is important for autonomous indoor mobile such autonomous vacuum cleaning robots, who need to maneuver in and out a room. If the door is open the robot can take the decision to maneuver outside the room otherwise stay inside.

2 Background / Formulation

Our objective is to examine deep neural networks models that could be accurate enough to solve computer vision classification problems such as the ones mentioned above. Several state of the art network architectures are available for testing. The basic components of these network models consists of convolution layers followed by pooling layers. The addition of deeper layers helps improve the accuracy of these networks models at the computational cost. Moreover the more layers there are the more parameters are needed to be tuned. Various network models are available at our disposal to complete the project : Alex Net [2] , Seg Net [3], Google LeNet [4] and LeNet [5]. When choosing between the different models it was important to choose the model which is applicable for an image classification problem. Therefore it was obvious that SegNet should not be considered as it is focused primarily on image and pixel wise segmentation which does not fit our purpose here. In addition LeNet only takes grayscale images with sizes of 28 x 28 [5] which does not suit our purpose.

A comparison was done between Google LeNet and Alex Net network models. The main advantage which Google LeNet has according to the literature is the reduced number of parameters [4]. AlexNet's parameters are around 60 million whereas Google LeNet's is found to be at 4 million [6]. Moreover after experimentation with the dataset AlexNet was unable to achieve the required accuracy needed to pass the evaluation. Google LeNet was developed and inspired by LeNet . The neural model was focused on being deeper, wider and sparser then existing deep network models [4]. The model basically consists of 22 layers, and implemented an inception module. The inception module consists of very small convolutions to drastically reduce the number of parameters.

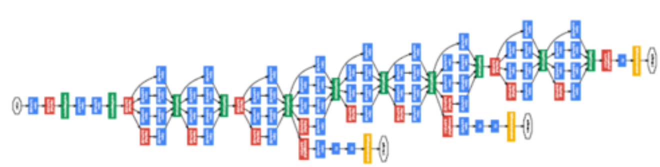

Fig. 1. Google LeNet Neural Network Architecture

The model would be trained on the NVIDIA DIGITS workspace using ADAM optimization method [7]. The ADAM optimization algorithm would be used which is an extension to the stochastic gradient decent algorithm. Number of epochs used for the Bottle-Candy Box Classification was around 15, while 30 epochs were used for the Open-Close door Classification. The learning rate for both classification problems was chosen to be at 0.001. The choice of the learning came after much tuning and experimentation to achieve the needed and required accuracy.

3 Data Acquisition

3.1 Bottle - Candy Box Classification

The data chosen for this application was based on Udacity's P1 dataset. The dataset contains 10,094 images. The images were separated in three different folders to differentiate between the three image classes (bottles,candy box, and nothing). 7,570 images were used for training the neural network model while the remaining 2,524 images were separated for validation. The dataset images are colored RGB images with 256 x 256 size. The image size (256 x 256) is a suitable input for Google LeNet. The dataset contains images of the object which we desire to classify which makes the dataset suitable for use to such application.

3.2 Open - Close Door Classification

The dataset was for the open - close door classification was taken from a Logitech camera connected to a Dell Computer laptop. OpenCV python script was used to capture RGB images of size 256 x 256 . It was important to obtain 256x256 sized images since this was the normal input expected by the Google LeNet Nueral network. The images were taken on several different doors,different lighting condition, and different distances from the door itself. The collected data was separated into 3 folders/categories (open door, closed door, wall). The closed doors contained 410 images,open door 203 images , the wall or no door category contained 203.Which in total resulted in a database of 816 images.

4 Results

The results obtained for each classification problem were as follows.

4.1 Bottle - Candy Box Classification

For classification between Bottle and Candy Box an accuracy of 75.04% with an average time of 5 ms. The time of training reached approximately 28 minutes. Figure 2 and Figure 3 provide further insight and details to the results obtained.

4.2 Open - Close Door Classification

For the Open - Close Doors the Training Chart Figure 4 describes the progress of the trained neural network. The overall accuracy reached is around 100%.

Images in Figure 5, Figure 6, and Figure 7 have shown moreover the capability of the trained model to successfully determine if the image contains an open,closed or no door.

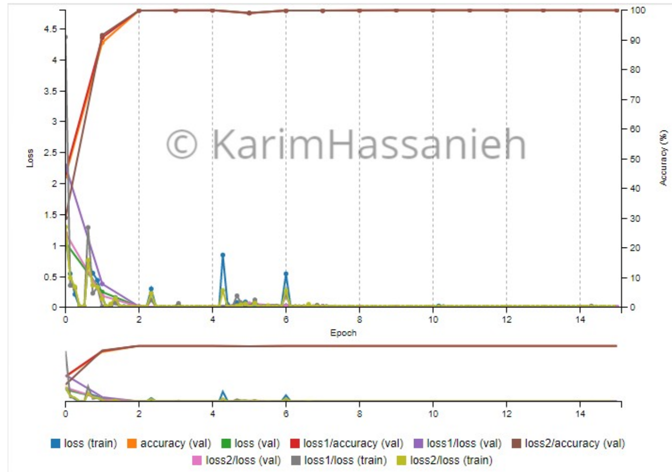

Fig. 2. Training Progress Chart - Bottle Box Classification

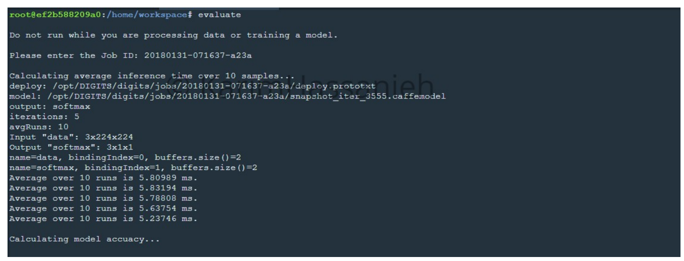

Fig. 3. Evaluation Results - Bottle Box Classification

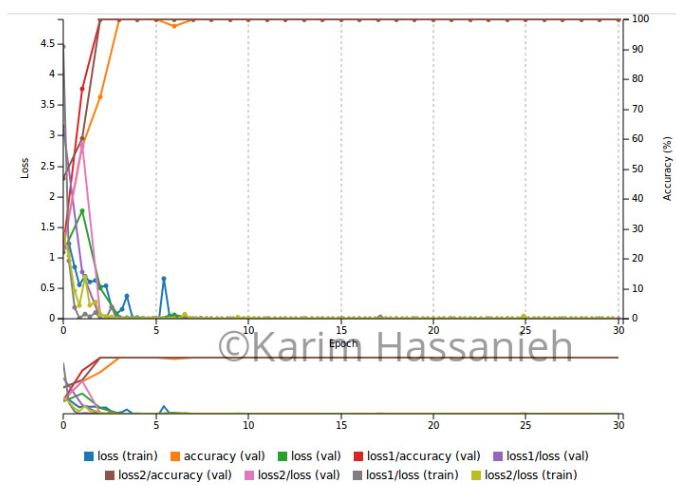

Fig. 4. Training Progress Chart - Open/Closed Door Classification

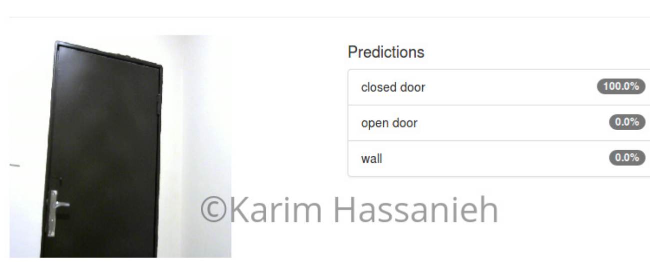

Fig. 5. Closed Door Classification

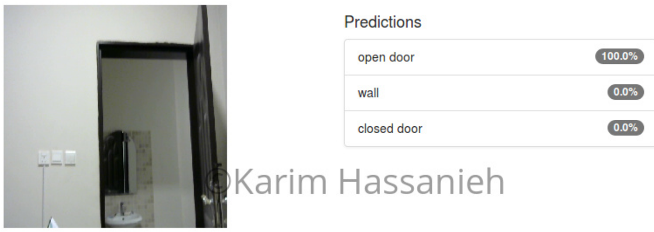

Fig. 6. Open Door Classification

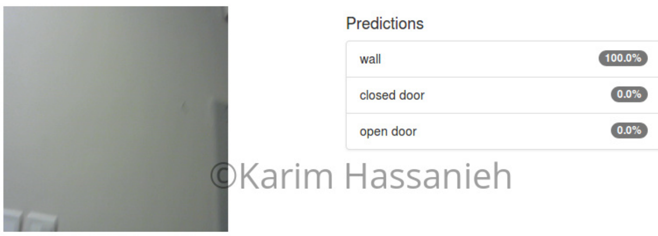

Fig. 7. Wall-No Door Classification

5 Discussion

The above application was achieved with acceptable accuracy and timing. An accuracy above 75% was reached which meets the project requirement with a runtime of 5ms. It is critical to have an acceptable balance between accuracy and runtime. Most machine learning process require quick decision making to carry on the process or provide feedback to an end-client or user. This however will depend on the computational capacity carried by the machine itself.

6 Future work

There is still room for future work on the provided applications. For the application to be deemed robust and commercially viable further testing needs to be done.The Open-Closed door classification can be integrated to indoor mobile robots which carry camera sensors. However further testing needs to be done on more types and models of doors with varying color and nature.

References

[1] R. Miikkulainen, J. Z. Liang, E. Meyerson, A. Rawal, D. Fink, O. Francon, B. Raju, H. Shahrzad, A. Navruzyan, N. Duffy, and B. Hodjat, "Evolving deep neural networks," *CoRR*, vol. abs/1703.00548, 2017.

[2] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*, NIPS'12, (USA), pp. 1097–1105, Curran Associates Inc., 2012.

[3] V. Badrinarayanan, A. Handa, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for robust semantic pixel-wise labelling," *arXiv preprint arXiv:1505.07293*, 2015.

[4] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. E. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," *CoRR*, vol. abs/1409.4842, 2014.

[5] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, pp. 2278–2324, November 1998.

[6] S. Das, "Cnns architectures: Lenet, alexnet, vgg, googlenet, resnet and more ," 2017.

[7] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *CoRR*, vol. abs/1412.6980, 2014.