Project: Robotic Inference

Rastri Dey

Abstract—This report illustrates an embedded robotics inference system, leveraging Nvidia's Deep Learning tools, to create models on a DIGITS workflow platform that can be deployed in real time robotics applications. Real time image analysis and recognition of different features of Indian Bank Currency Notes– Original and Fake, using GoogLeNet neural network has been proposed. A cheaper solution for Data collection is well devised and has been executed for the model to train on varying orientation, size, texture and shape of various real and fake denominations. Furthermore, a pre-supplied dataset has also been tested and performance is evaluated using Neural Networks on DIGITS platform.

Index Terms—Classification, Deep Neural Network, Currency recognition, Data extraction.

I. Introduction

Image Classification, object recognition, Neural networks are a few of the most crucial fields of interest for our rapidly changing world today. From a vehicle's license plate, a sign on a sign board, face reognition, document authentication to a person's iris in his eyes and blood cells in his veins, there's possibly nothing that can't be trained, solved and recognized by an Inference System. Rather the field is so wide and evolving that possibly in future, there will be aritificial brains inside human body, trained in Deep Neural Networks. Amongst the various challenges that could be solved using Neural Networks, authentication in Commerce is one significant application, the financial world is dealing with today. Each day there are millions to trillions of monetary transaction in world banks, business dealings, import and export, share market etc. In the banks where a large transaction of money is an everyday affair, sorting the notes, classifying them, verify for authentication is a highly critical job. An object classification system trained to tackle this issue, will not only save a lot of human-effort hours, but can also potentially save the financial market in dealing with misclassification of notes or even authentication of forged notes at a highly stable, system accuracy. This report presents a model trained for bank notes authentication and classification on Indian Rupees for various denominations under different conditions of orientation, luminance, texture etc. Data is collected using a very simple yet efficient way on which a standard neural network is applied, which is then allowed for performance evaluation. The trained model results in a high accuracy classifier without misclassifying a single class.

Number of researchers, in recent years have contributed to the subject of paper currency recognition system, identification, sorting and recognizing counterfeit notes[1-5],[8]. Traditional methods include linear aray of photo-emitters and photo-detectors in one side of the banknotes that determines the authenticity of the notes from reflected and transmitted light[6]. This is highly infeasible where the software based training is more robust because of its high flexibility to adopt for upgraded applictaions such as, classifying for newer denominations– where changing the hardware could be a rather expensive proposition.

II. Background & Formulation

On analysis of Deep Neural Network models, for choosing a network best suited for the application in hand, multiple metrics as presented in [7], are reoptimized. The hard constarints for the optimization are stressed on Accuracy, Inference time, Operation count, Parameters and Information Density. The power utilization for Data acquisition and model training is executed through a wall connected power socket.

As stated in [7], accuracy and inference time are in hyberbolic relationship, a lot of additional computational time may not lead to a high increase in accuracy. Following curve shows, the performance comparison of a few state of the art DNN architectures in terms of percent accuracy with respect to Inference images per second, for the classification of 1000 labels on ImageNet.

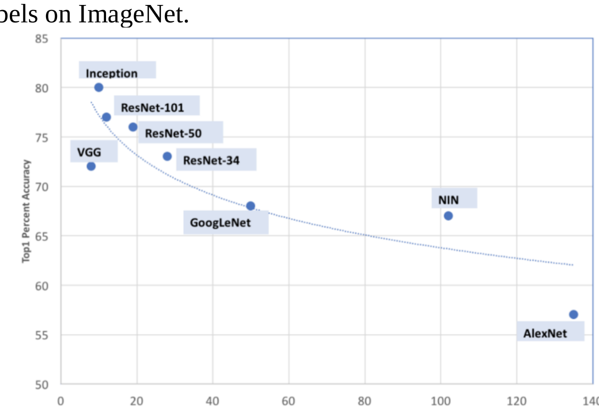


Fig. 1. Performance comparison: DNNs architecture[7]

It can be inferred from the figure that, GoogLeNet is an optimum balance of accuracy and inference fps among other existing DNNs.

Next, the comparison is performed on the optimization parameter: Operation count. Operation count is a rough estimate of inference time and hardware deployed. As the number of operations increases, there is an increase in Inference time. This relationship trend is linear in the ImageNet classification shown in the below curve. Here the size of the blobs is used to depict the number of network parameters used. It can be deduced that the performance of DNN architectures such as, batchnormalisedNetwork In Network (NIN) (Lin et al., 2013) (Red), ENet(Paszke et al., 2016) (Black),GoogLeNet (Szegedy et al., 2014) (Blue) are a good balance of operations count, inference time and the number of network parameters, whereas AlexNet(Zagoruyko,2016) (Orange) even though incurs less inference time, posses a large number of network parameters.

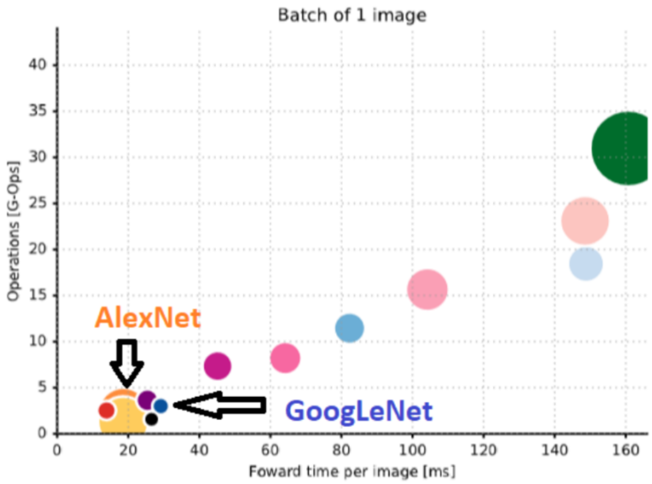


Fig. 2. Relationship curve of Operations vs Forward time per image [ms], among state of the art DNNs [7]

As much like the number of operations, parameter size also impacts the accuracy.

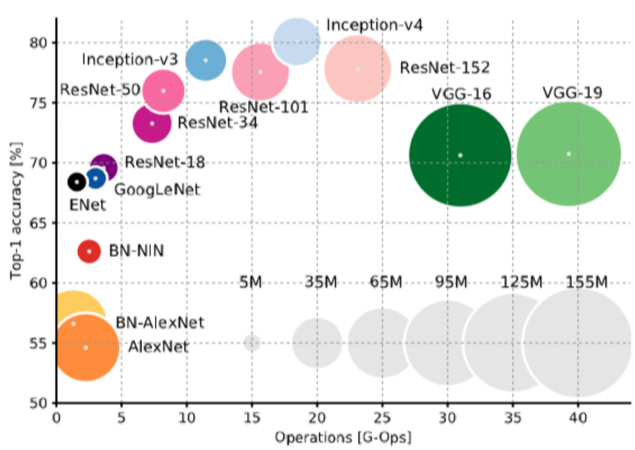


Fig. 3. Trend for Network Parameter size, accuracy and operations for DNN architectures in ImageNet classification challenge[7]

As shown in the figure above, it can be observed that VGG-16 and -19 (Simonyan & Zisserman, 2014) (Green), are having accuracy more than AlexNet (Orange), though at a cost of increase in operations and the network parameter size. The real figures for the number of parameters in the ImageNet classification challenge are 4 million for GoogLeNet, 60 million for AlexNet and a whooping 138 million for VGGNet.

A large increase in parameter size also impacts Information Density. The efficiency of a network to utilize its parametric space is quantified in terms of its information density. An improvement in network architecture would require change in its parameters, a higher improvement in the network performance owing to lesser parametric change will reflect in a higher information density. A bar graph relating to this is depicted below, comparison makes GoogLeNet and ENet stand out among others.

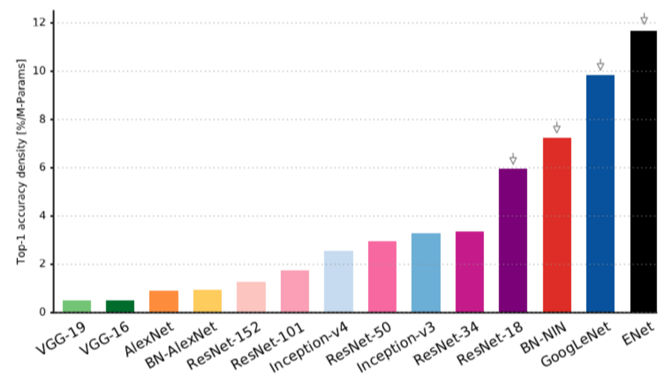


Fig. 4. Accuracy density for Networks[7]

The number of classes for the supplied dataset and the Banknotes dataset, is 3 and 5 respectively. This is much lower in comparison to the ImageNet classification of 1000 classes. Hence, the accuracy levels and inference time would be enhanced from the ones depicted in Figure 1 for the current application. Moreover, on the basis of the number of network parameters and information density, as shown in the comparison figures 2, 3 and 4, GoogLeNet is observed as one of the notified performers.

After scrutinizing various network performances and weighing each one of them under different performance metrices, GoogLeNet is identified as the best network solution on the supplied dataset and the dataset of bank currency notes. Its an optimum balance of lower computational burden, higher accuracy and lower inference time, leading to lower power consumption and memory requirements.

The implementation of a dataset in GoogLeNet architecture inside NVIDIA's Digits platform, requires the images in dataset to be— RGB colored pixels with 256 X 256 dimension. The supplied dataset for Candy boxes, Bottle and Nothing are therefore assigned as 'squash' in Resize Transformation of Digits workflow.

Similarly, the model for the Banknotes dataset is also trained on standard GoogLeNet architecture on Digits, with similar resize transformation of 'squash' for the required pixel dimensions. The banknotes are typically having dimensions (in mm) – (150 x 60) for 500 Rupees to (123 X 63) for 10 Rupees. Each denomination is having intricate markings on them, which requires the pixel resolution to be effective enough to not miss any valuable information. The images are captured in colored pixels to emphasize on varying colors on different denominations. The Red, Green and Blue channels inside the input images aids to the image classification even further.

III. Data Acquisition

Data acquisition is of paramount importance in any Neural network architecture. For the pre-supplied dataset, there were 10,094 images of Bottle, Candybox and nothing. 25% of this dataset is allowed for Validation and 75% for Training. The resulting count for images under different classes for training and validation is represented in a tabular form as below:

Class	Image Dimension	Training Images	Validation Images
	(Width X Height)		
Bottle	256 X 256	3426	1142
Candybox	256 X 256	1871	624
Nothing	256 X 256	2273	758
		7570	2524
			Total = 10,094

Table. 1. Representation of classes for Pre-supplied dataset

The supplied dataset information in the Digits workflow looks like:

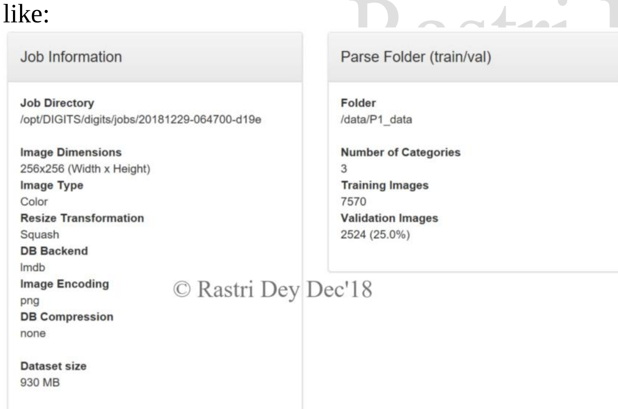


Fig. 5. Pre-supplied dataset information on Digits

A typical representation of images in the pre-supplied dataset:

Fig. 6. Sample Images from pre-supplied dataset

The dataset for Candybox, bottle and Nothing are provided by Udacity sample dataset. The Dataset for banknotes, on the other hand are captured using the laptop webcam. At first, a collection of bank notes in random orientations are made to stick on a blackboard, which is then vertically placed with the support of a wall. A wheeled base is constructed for the laptop to move and collect images covering the entire length of the blackboard. The height of the laptop onto the wheeled base is about 20 cm, this is just enough to capture the complete note in one frame. Every bank note is made to stick on the blackboard using a double sided tape.

On a single run, denominations of one class are captured, by moving the wheeled base infront of the blackboard for about 1-2 minutes. The entire setup for Data Acquisition is shown in the figure below.

Fig. 7. Lab setup for data collection

Finally, for some variation in the luminance texture of picture quality, images are captured in two different lightning conditions. Few of the images for every label are captured in natural daylight conditions while the rest few are captured through the camera flash from a mobile handset. This is adjusted alongside the laptop onto the wheeled base which makes both the laptop and the mobile handset move together while capturing images.

All images are captured in RGB with 640 X 480 pixel dimensions at a time span of every 100ms. The brightness of every image is adjusted using a function called increase_brightness(img, value) as shown in the code snippet below:

```
import cv2,time

video=cv2.VideoCapture(0)

def increase_brightness(img, value):
    hsv = cv2.cvtColor(img, cv2.COLOR_BGR2HSV)
    h, s, v = cv2.split(hsv)

    lim = 255 - value
    v[v > lim] = 255
    v[v <= lim] += value

    final_hsv = cv2.merge((h, s, v))
    img = cv2.cvtColor(final_hsv, cv2.COLOR_HSV2BGR)
    return img
count=0
while count<600:

    count = count+1
    check, frame = video.read()
    print(check)
    print(frame)

    cv2.imshow("Capturing", frame)
    frame = cv2.resize(frame,(265,265), interpolation = cv2.INTER_NEAREST)
    frame = increase_brightness(frame, value=5)
    cv2.imwrite("Original_Note50_" + str(count) + ".png", frame)
    print(count)
    key=cv2.waitKey(100)
    cv2.waitKey(100);
    if key == ord('q'):
        break

video.release()
cv2.destroyAllWindows
```


Fig. 8. Code for Data acquisition

The number of images required for classification are adjusted as per the model requirement for the corresponding class, through a while loop count.

For each denomination there is approximately, 700 images collected across 5 class. The denomination classes are – INR 10, INR 20, INR 50, INR 500 and Fake notes. The Fake notes are article cuts from daily newspaper and shopping bills having dimensions similar to an original bank note.
The INR currency dataset, consists of 2992 training images and 996 validation images of 256 X 256 image size. Every image of RGB color type is assigned as 'squash' in the resize transformation of Digits workflow. The total dataset size is of 329 MB. Similar to supplied dataset, the number of training images are 75% and number of validation images are 25% for the banknotes dataset. The images are taken using OpenCV tools for video capturing and are saved in the disk at an assigned fps.
On generating this dataset on 'INR Currency bank notes', the resulting summary from Digits platform looks like:

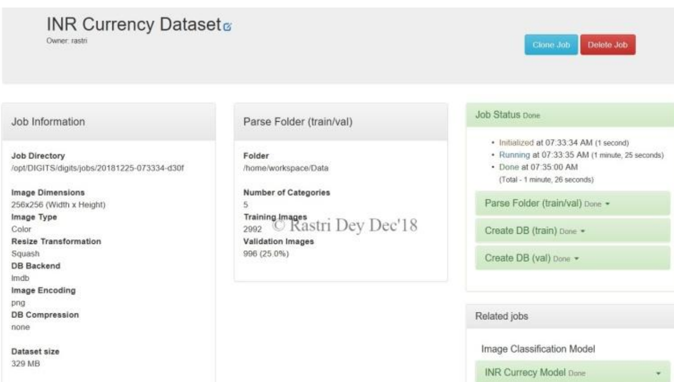


Fig. 9. INR currency dataset summary on Digits

A typical sample of images from this dataset can be shown as below. It can be noticed that every image is captured with a common black background.

Fig. 10. Sample of images from collected dataset

The dataset image count from Training and Validation set for the 5 labels can be represented as:

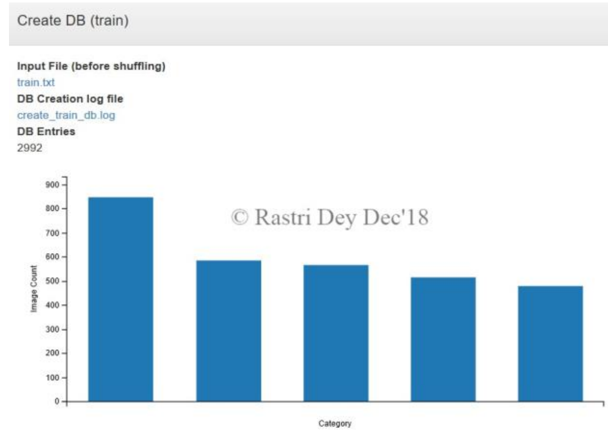


Fig. 11. Image count bar graph on Training dataset of Banknotes

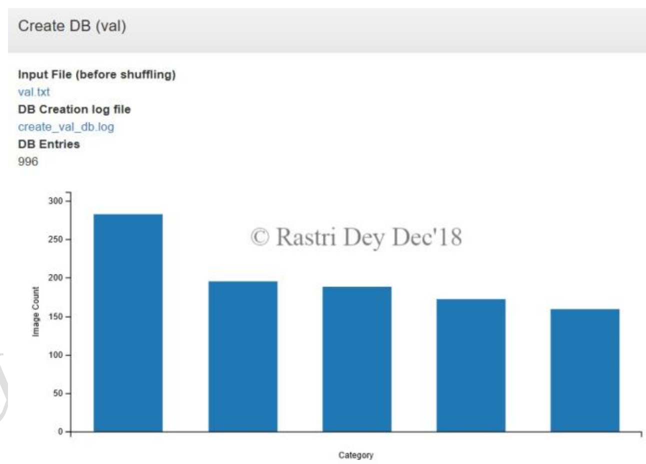


Fig. 13. Image count bar graph on Validation dataset of Banknotes

It can be noticed that the count for Fake Notes are maximum for both training and validation set. This has been done for the model to learn from the wide difference gap in every other fake note image. A good dataset helps the model to learn from random newspaper clips and shopping bills taken into account for the Fake Note dataset.
A summary of the INR Currency banknotes dataset is represented in a tabular form as shown below:

Class	Image Folder Path	Training Images	Validation Images
OriginalNote 10INR	home/workspace/data/OriginalNote_10INR	566	188
OriginalNote 20INR	home/workspace/data/OriginalNote_20INR	479	159
OriginalNote 50INR	home/workspace/data/OriginalNote_50INR	515	172
OriginalNote 500INR	home/workspace/data/OriginalNote_500INR	585	195
FakeNote	home/workspace/data/FakeNote	847	282
		2992	996
			Total = 3,988

Table. 2. Representation of classes for collected dataset

The methodology for Data acquisition been implemented here, is extremely cost efficient. The whole setup was arranged from easily available household things. The entire processing time for data capturing, setup arrangement and processing the data, took only about an hour. Hence, such a procedure saved time, cost and efforts while rendering fruitful datsets, vital for any Neural network architecture.

IV. Results

Two case studies have been demonstrated in this report to illustrate the classification performance on two separate datasets. The datasets are individually simulated over GoogLeNet neural network models on digits workflow.

a. Results with Pre-supplied data:

The pre-supplied data from [/data/P1_data] folder path is extracted inside Digits platform and the classification model with an initial learning rate of 0.001 and a solver type Stochastic Gradient Descent (SGD) is made to run over the generated data 'P1_dataset'. The inputs from digits workflow for the classification model is shown below:

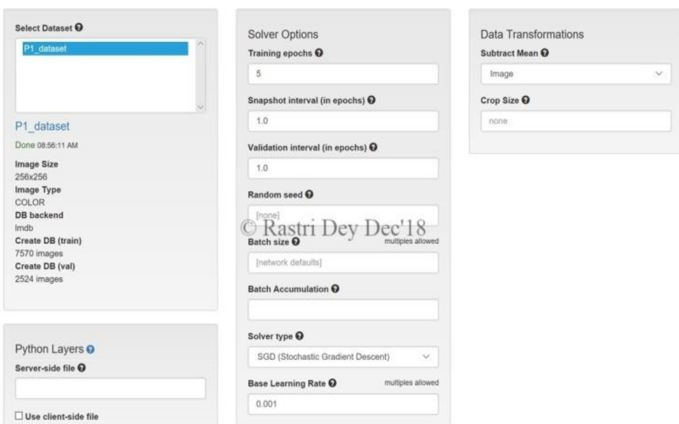


Fig. 11. Input Workflow for classification model on supplied data

The model is allowed to run for 5 training epochs at an initial learning rate of 0.001 with GoogLeNet, which then generates the following model summary:

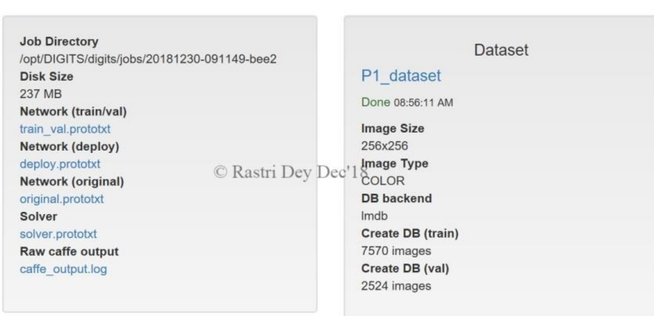


Fig. 12. Classification Model summary for supplied dataset

The model while learning, begins with an initial learning rate of 0.001 that reduces to its one-tenth in every 1.5 epochs approximately. Around 5th learning epoch, the learning rate comes down to 0.00001 as shown below:

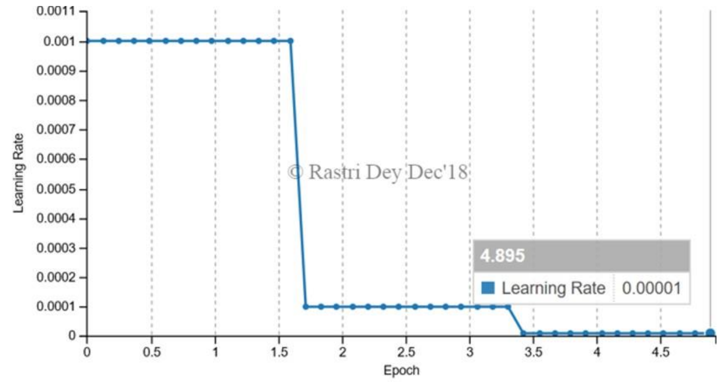


Fig. 13. Model learning rate for supplied data

The transition in learning rate is owed to the adaptive learning as the model progresses through every epoch. The accuracy vs epoch curve for the learned model looks like:

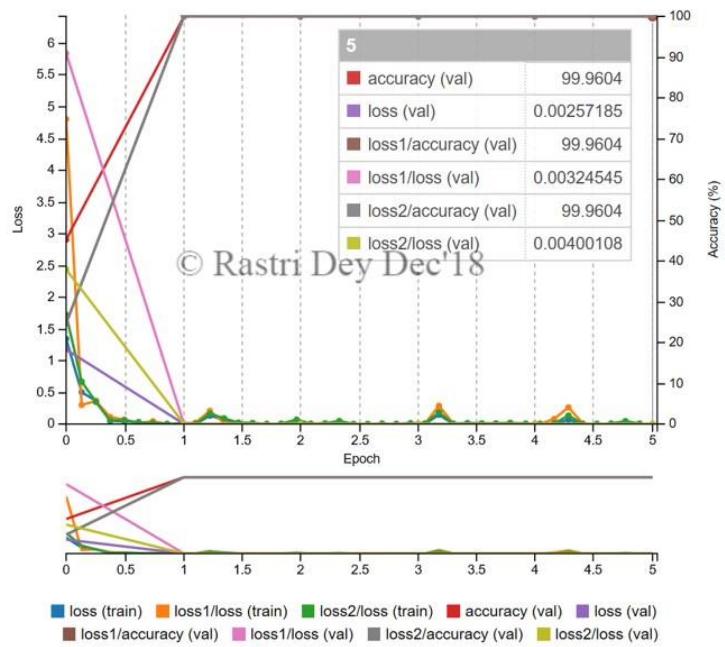


Fig. 14. Accuracy Vs epoch curve for classification model of supplied dataset

The validation accuracy generated after 5 epochs resulted in a value of 99.9%.The evaluate command in the terminal for this Job Id, gives an accuracy value of 75.41% at an inference time of around 5ms. The exact values for inference time and accuracy with the model Job Id can be verified from the terminal figure below:

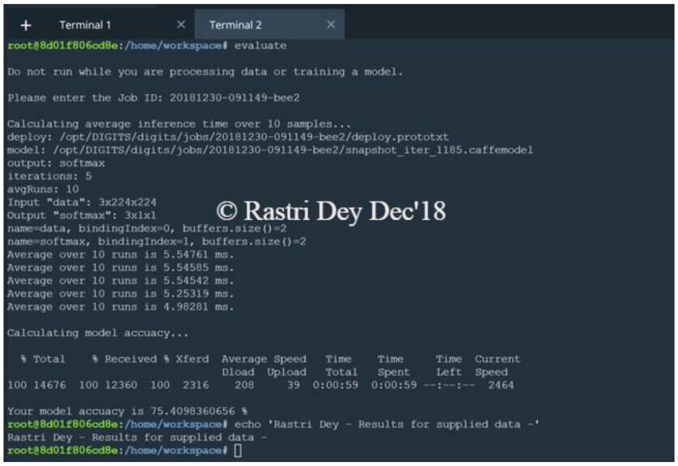


Fig. 15. Results on Accuracy and Inference Time for supplied dataset

b. Results with captured data:

The captured data of INR Currency Banknotes is simulated in Digits workflow with 30 training epochs at an initial learning rate of 0.001 with Stochastic Gradient Descent (SGD) solver type:

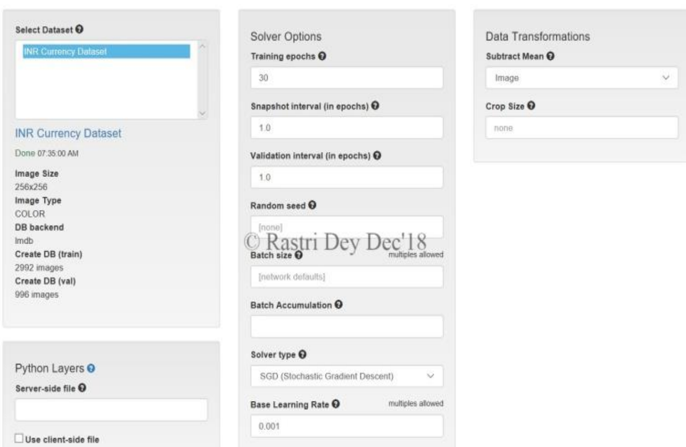


Fig. 16. Input Workflow for classification model on INR Currency Dataset

The model generated from the above inputs, results in the following model summary:

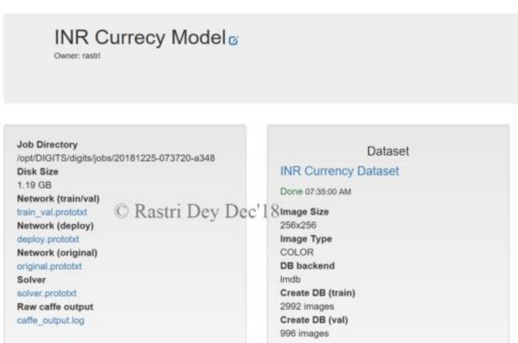


Fig. 17. Classification Model summary of INR Currency Model

The model transits thrice from its initial learning rate of 0.001. Every 10th epoch succeeds with a reduced learning rate. The learning rate curve from 30 training epochs looks like:

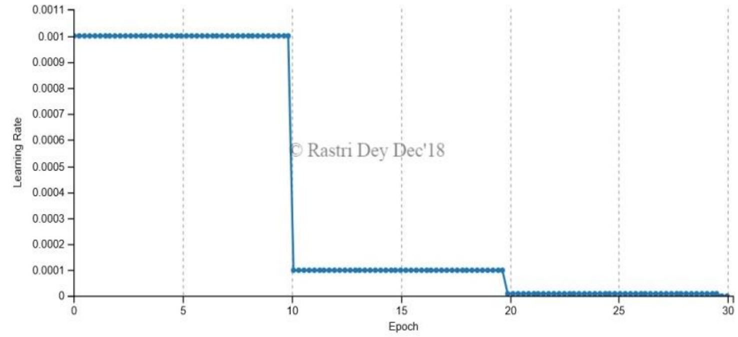


Fig. 18. Model Learning rate for Captured Data

The model keeps improving with every epoch and a new learning rate. The training for 30 epochs, converges very quickly to a minimal loss value after 15th epoch. This finally ends at an accuracy level of 99.9% approximately. Following curve shows the training loss, validation loss and accuracy levels for the INR Currency banknotes datasets, with a GoogLeNet architecture:

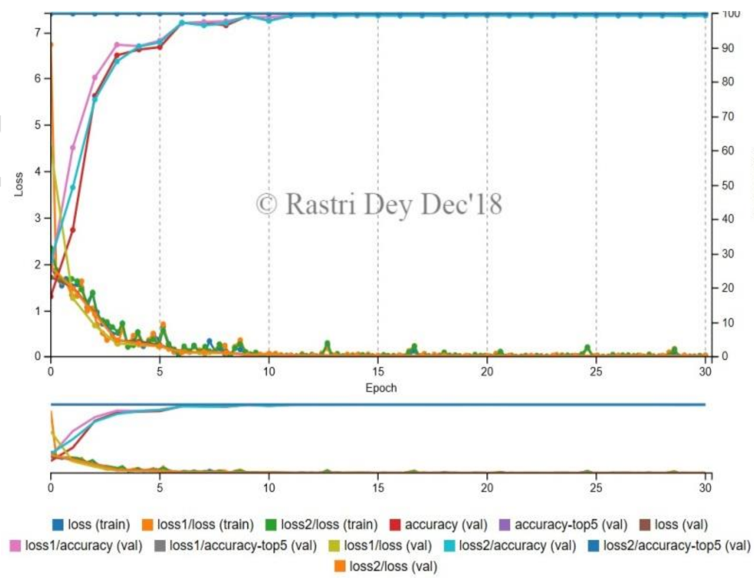
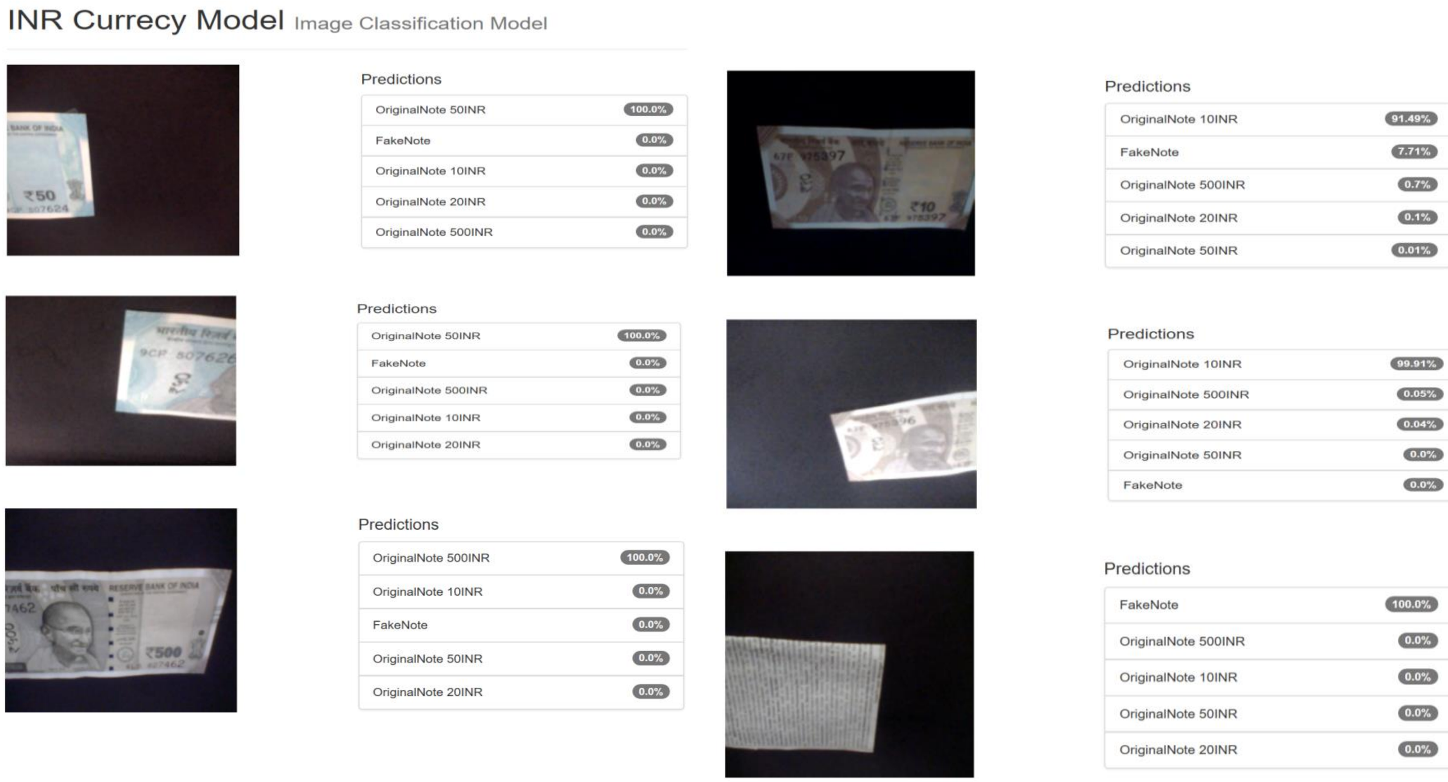
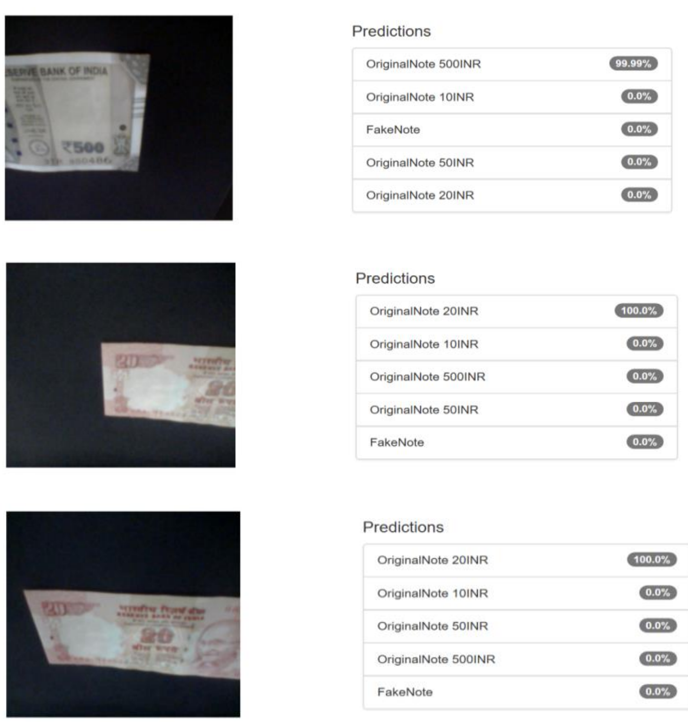


Fig. 19. Accuracy Vs epoch curve for classification model of INR Currency Banknotes dataset

The generated model is then allowed to evaluate the classification performance over a few test images. These were separated from training and validation data beforehand. The test dataset compises of 2 test images from each class, a total of 10 test images for 5 class. The prediction on test dataset with the trained classification model accurately determines all the test images with over 90% accuracy levels. The resulting predictions on test dataset is shown below:





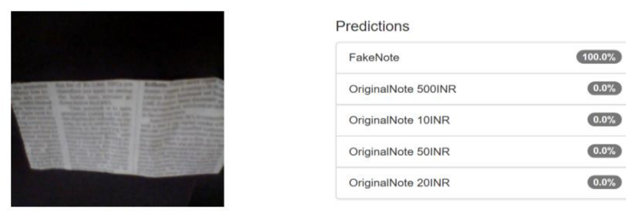


Fig. 20. Predictions of INR Currency classification model

The accuracy levels for the INR currency model for test image 50 INR, 20 INR and Fake Note produces an accuracy level of 100 % for both the test cases. While the prediction for 10 INR in one of the case is 91.49%, assumingly due to low luminance conditions. Similarly for 500 INR the prediction accuracy levels are over 99% for both the test cases.
Hence, its concluded that the INR Currency Model with a test accuracy level of over 99% performs remarkably well for all the classes under varying test environmental changes.

V. Discussion

The DNN architectures of LeNet, AlexNet and GoogLeNet are very compelling in creation of classification models. Its observed that GoogLeNet attains a very high accuracy level within a few epochs of 4 or 5 in comparison to AlexNet where similar accuracy levels are obtained after a several many epochs or sometimes not even then. This is presumably because of the set upper bound on accuracy for a given architecture at a particular frame rate. However, the time required by GoogLeNet to train a model is more in comparison to AlexNet. This although can be justified with the obtained accuracy levels of GoogLeNet. Clearly, if the prime focus is to build a classifier, the basic requirement is to achieve, atleast a minimum level of accuracy that can classify all the labels flawlessly. Hence accuracy is more vital than the inference time. Yet a real world classification model cannot take indefinite time as that will mean more power consumption and thus increase in costs which is not viable in a real world screnario. Hence an optimum balance is what’s preferred.

The results obtained for the supplied dataset, meets the minimum criteria of obtaining 75% accuracy and inference time less than 10ms with the help of GoogLeNet architecture. Though, a preference on even better architectures such as Inception or ResNet or ENet are not tested on these datasets. The INR Currency Model also performs exceptionally well resulting in 100% accuracy levels for most of the test cases. Yet an even better test case would have involved, a wide spectrum of difference in the notes structure such as torned notes or old notes etc. The data acquisition methodology for INR Currency Dataset is also very efficient besides being cheap. The data acquisition takes just a few minutes to collect data for a particular class. Apart from, inference time and accuracy, for the application of Banknotes recognition, power consumption is also an important factor, considering the Automated Teller Machines (ATM) or other applications, where small power is a limiting factor.

VI. Conclusion & Future Work

Paper currency recognition is one of the vital areas of research to prevent banknotes forgery, theft and currency frauds. A product to prevent these will not only help the Government but the entire country and the world economy in achieving prominent heights. This would be a great commercially viable product that can be used in banks, offices, markets, schools or anywhere, where there is a lot of financial dealings occurring, on everyday basis. The performances of the INR Currency classifier using GoogLeNet as presented in this report performed exceptionally well, giving a 100% classifying performance for all the classes. But there is still room for improvement. There are different kinds of Notes available for a single denomination which are not taken into account during this classification. For example, there are two kinds of 50 Rupee Notes valid in Indian market, a better classifier should be able to classify both as 50 Rupee note under different categories of New and Old. Moreover in foreign exchange, there are monetary transactions within different countries, a classifier to classify the notes for a particular country and then to a particular denomination of that country would be highly appreciable. This would although mean increase in number of classes, more time, more power consumption and more memory requirement. Yet this would be a good encouragement in exploring better Neural Networks which would be helpful in building even further, sophisticated classifiers.

References

[1] Zhang EH, Jiang B, Duan JH, Bian ZZ. Research on paper currency recognition by neural networks. In Proceeding of the second international conference on machine learning and cybernetics, p. 2193–2197; 2003.

[2] Liu Q, Tang L. Study of Printing Identification Based on Multi-spectrum Imaging Analysis, Proceedings of the International Conference on Computer Science and Software Engineering, p. 229 – 232; 2008.

[3] Ahmadi A, Omatu S, Yoshioka M. Implementing a Reliable Neuro-Classifier for Paper Currency Using PCA Algorithm, Proceedings of the41st SICE Annual Conference, p. 2466-2488; 2002.

[4] Takeda F, Omatu S. High Speed Paper Currency Recognition by Neural Networks, IEEE Transactions on Neural Networks 1995; 6(1):73-77.

[5] Takeda F, Omatu S, Onami S, Kadono T, Terada K. A Paper Currency Recognition Method by a Small Size Neural Network with OptimizedMasks by GA, Proceedings of the IEEE World Congress on Computational Intelligence, Orlando, USA, 7, p.42434246; 1994.

[6] Currency paper verification and denomination device. United States Patent 5416307.

[7] Canziani, Alfredo, Adam Paszke, and Eugenio Culurciello. "An analysis of deep neural network models for practical applications." arXiv preprint arXiv:1605.07678 (2016)

[8] Jyothi, Ch.Ratna, Krishna, Y.K. Sundara, Rao, V Srinivasa, "Bank Notes Authentication System based on Wavelet Features and Artificial Neural Network," International Journal of Computer Applications, 2016, Vol. 142, No.2, pp. 24-28.