Robotics Inference

Dhawal Wazalwar

Abstract—In this project, robotics inference has been used to work on two tasks: one on supplied dataset sorting conveyor belt objects and secondly on acquired data classifying kitchen objects. Details regarding how training images have been captured and its setup are explained. For this project, NVIDIA DIGITS work-flow was used to train the database and also for inference step to evaluate testing performance. For both tasks, GoogleNet architecture using Stochastic Gradient Descent optimization was employed. In case of Task1, accuracy of 75.4% and average inference time of about 5.4ms was achieved. For Task2 using controlled environment setup, validation accuracy of about 95% and also improved testing performance was seen. Finally, trade off factor between accuracy and inference time for Task2 has been addressed and also how this idea can be used for commercial purposes is also discussed.

Index Terms—Controlled Environment, GoogleNet.

✦

1 INTRODUCTION

DOMESTIC or household service robots have completely transformed the way modern homes look now and that impact is now seen in our kitchens too. Carnegie Mellon University's Home Exploring Robot Butler (HERB) can load and unload dishes from dishwasher [1]. Moley Robotics has created world's first fully automated and intelligent cooking robot, which can recreate recipes from world class chefs [2]. This is just tip of the iceberg, in future many more such intelligent robots that can handle complex tasks in kitchen will be seen.

Any robotics system has three key components: Perception, Decision and Action. In this project, focus is on how robotic inference can be used to perceive our environment in kitchen surroundings. In particular how kitchen objects like bowls, coffee mugs, plates and others can be classified using deep learning. In commercial kitchens, such automated sorting can have tremendous impact in overall productivity improvement and also save considerable time.

2 BACKGROUND / FORMULATION

For the two inference tasks, Task1 on supplied data and Task2 on acquired data, a separate training and validation database (25% of total database size) using gray-scale were created. For Task2, a separate testing database was also created, more details are explained in data acquisition. DIGITS work-flow has 3 predefined networks: LeNet, AlexNet and GoogleNet. GoogleNet because of clever use of inception module has lesser parameters to optimize, 4 Million compared to 60Million in AlexNet. Therefore for both tasks 1 and 2, default GoogleNet architecture was used.

For selecting optimal learning rate, separates runs with varied learning rates ranging from 0.001 to 0.01 were employed for 1 epoch. After careful analysis, 0.005 was selected for both tasks as it had both, steep fall in loss function as well it was more gradual i.e less noisy compared to 0.01.

For Task1, GoogleNet with Stochastic Gradient Descent (SGD) optimization with learning rate of 0.005 and for about 10 epochs was used. For Task2 even though overall same training parameters were used, learning rate with decay factor of 50% and Gamma of 0.2 was also used. Apart from

SGD, use of Adam optimizer was also tried but that did not give expected results and overall SGD seemed to be working better for both tasks.

3 DATA ACQUISITION

The most challenging part for this project was to collect required data for network training. A controlled environment was set up using web cam, where in only one object was present with predefined background. Even though captured images originally were color and of dimension 1280x720, based on task complexity and network requirements, final processed images were gray-scale and of size 256x256. Overall about 25-30 images for each of the 4 categories (CoffeeMugs, Plates, Bowls and Others) were captured. Some images can be seen below in Fig. 1.

Fig. 1. Sample Images for Robotics Inference Task2

For a good classification network though, 25-30 images are not enough and so data augmentation becomes critical here. Fig. 2 shows how for each image, 40 different representations of original image have been created. Also seen in Fig. 3. is sample data augmented images for CoffeeMug class.

A separate testing database was created so that network performance can be analyzed on completely new test objects in all 4 categories. There can be an argument why not use some part of original database itself as testing data, as is done for validation. However it can give a misleading
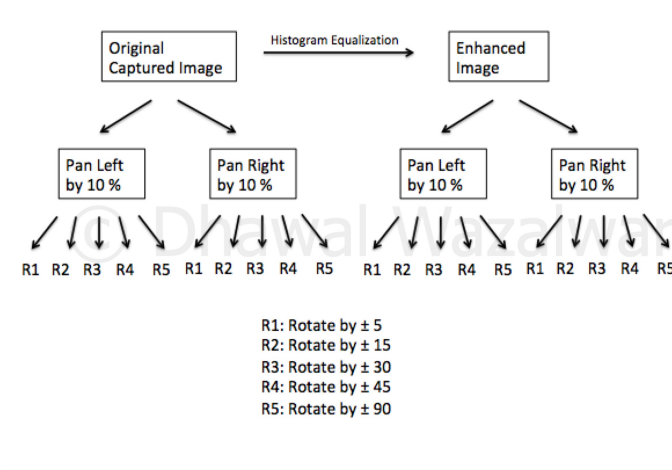
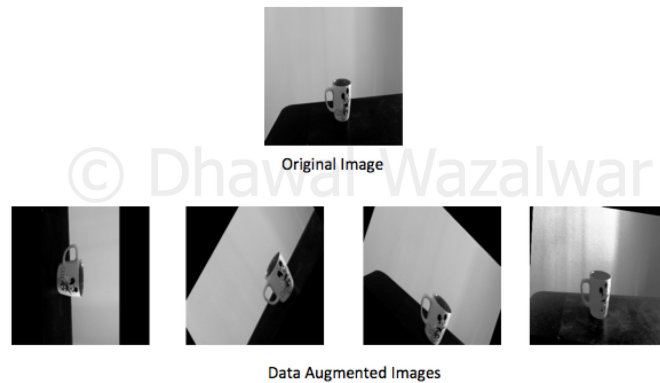Fig. 2. Chart explaining data augmentation done for each captured image

Fig. 3. Sample images for CoffeeMug Class

picture of testing performance and also affects inference model robustness. For an accurate testing analysis, complete separation between testing and training database is a must.

After manual inspection, some images where object were not clearly visible, are not included for network training. Overall count for all categories used in network training is shown in Table 1 below.

TABLE 1
Total Number of Images in image category

Bowls	986
CoffeeMugs	1286
Plates	1160
Others	877

4 RESULTS

For Task1 on supplied dataset, GoogleNet with SGD and fixed learning rate of 0.005 gave expected accuracy of 75.4% and average inference time of 5.4ms. Validation accuracy very close to 99% was achieved. Fig. 4 and Fig. 5 below
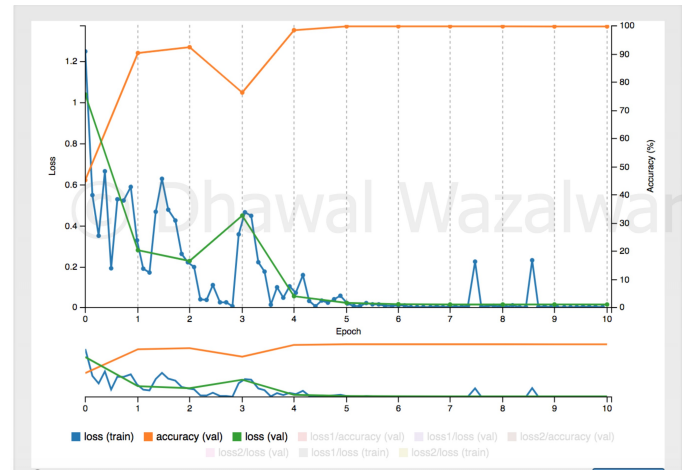

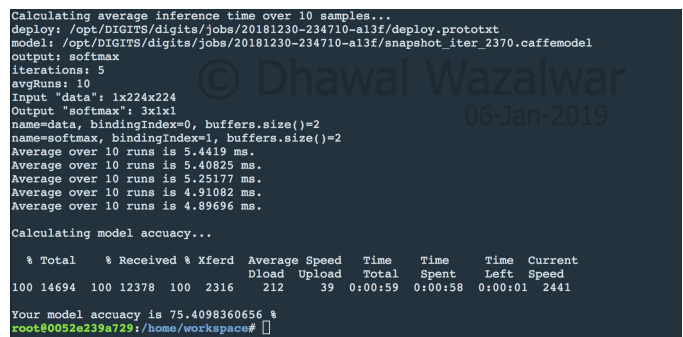Fig. 4. Training Graph for Robotics Inference Task1

Fig. 5. Task1 Evaluate Command Output

show training-validation graph and also output of evaluate command respectively.

For Task2 on acquired dataset, as mentioned in Background/Formulation two approaches were tried: firstly with fixed learning rate and another with decay rate factor. Fig. 6. shows training graph with using fixed learning rate of 0.005.
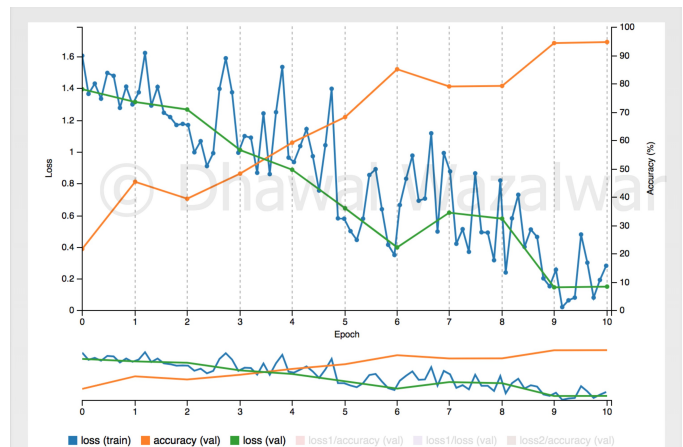

Fig. 6. Training Graph for Inference Task2: fixed learning rate 0.005

As seen in graph above, validation accuracy close to 95% was achieved. It can also be seen, learning loss is decreasing but it is too noisy and that impacts robustness and testing

performance of model. Fig. 7. below also shows model inference on a separate testing image list created for all 4 categories. Overall test results seem good but look at the 4th test case it is less only 57%.

All classifications

	Path	Top predictions			
1	/home/workspace/Robotics-Inference/test_images/image-bowls-55.jpg	bowls 99.94%	coffeemugs 0.04%	others 0.02%	plates 0.0%
2	/home/workspace/Robotics-Inference/test_images/image-bowls-66.jpg	bowls 99.58%	coffeemugs 0.3%	others 0.12%	plates 0.01%
3	/home/workspace/Robotics-Inference/test_images/image-coffeemugs-4.jpg	coffeemugs 98.81%	others 0.81%	bowls 0.38%	plates 0.0%
4	/home/workspace/Robotics-Inference/test_images/image-coffeemugs-8.jpg	bowls 57.15%	coffeemugs 39.8%	others 3.04%	plates 0.01%
5	/home/workspace/Robotics-Inference/test_images/image-plates-10.jpg	plates 100.0%	others 0.0%	coffeemugs 0.0%	bowls 0.0%
6	/home/workspace/Robotics-Inference/test_images/image-plates-17.jpg	plates 100.0%	others 0.0%	coffeemugs 0.0%	bowls 0.0%
7	/home/workspace/Robotics-Inference/test_images/image-others-1.jpg	plates 85.46%	others 14.54%	coffeemugs 0.0%	bowls 0.0%
8	/home/workspace/Robotics-Inference/test_images/image-others-22.jpg	others 77.64%	plates 22.3%	coffeemugs 0.06%	bowls 0.0%

Fig. 7. Inference Task2: Result on Testing image list, fixed case

On the other hand if learning rate with decay factor is used, loss function decrease is less noisy see Fig. 8.
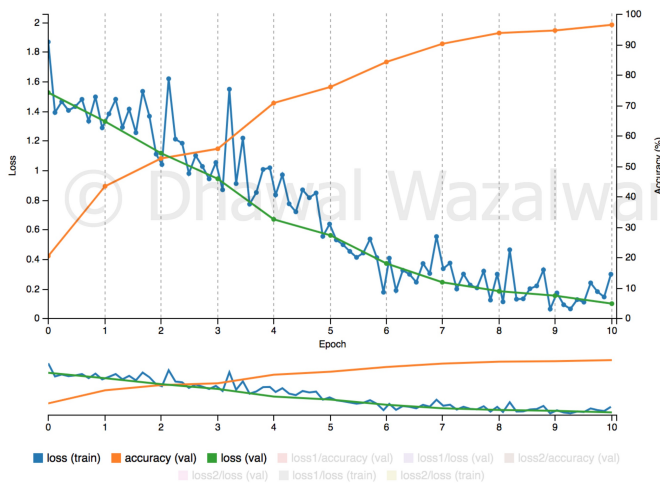

Fig. 8. Training Graph for Inference Task2: learning rate 0.005 with decay factor

Also in terms of testing performance on same image list, much improved performance was seen on 4th test case with percentage accuracy of 98% (see Fig. 9)

All classifications

	Path	Top predictions			
1	/home/workspace/Robotics-Inference/test_images/image-bowls-55.jpg	bowls 99.99%	coffeemugs 0.01%	others 0.0%	plates 0.0%
2	/home/workspace/Robotics-Inference/test_images/image-bowls-66.jpg	bowls 99.98%	coffeemugs 0.02%	others 0.0%	plates 0.0%
3	/home/workspace/Robotics-Inference/test_images/image-coffeemugs-4.jpg	coffeemugs 88.32%	bowls 11.61%	others 0.07%	plates 0.0%
4	/home/workspace/Robotics-Inference/test_images/image-coffeemugs-8.jpg	coffeemugs 98.08%	bowls 1.73%	others 0.19%	plates 0.0%
5	/home/workspace/Robotics-Inference/test_images/image-plates-10.jpg	plates 99.97%	others 0.03%	coffeemugs 0.0%	bowls 0.0%
6	/home/workspace/Robotics-Inference/test_images/image-plates-17.jpg	plates 99.87%	others 0.13%	coffeemugs 0.0%	bowls 0.0%
7	/home/workspace/Robotics-Inference/test_images/image-others-1.jpg	plates 75.6%	others 24.4%	coffeemugs 0.0%	bowls 0.0%
8	/home/workspace/Robotics-Inference/test_images/image-others-22.jpg	others 98.83%	coffeemugs 1.08%	plates 0.09%	bowls 0.0%

Fig. 9. Inference Task2: Result on Testing image list with decay factor

Fig. 10 and 11 below also shows testing performance on individual test images.

5 DISCUSSION

From results, it can be seen testing performance even for such small size dataset was quite good and validation accuracy also was close to 95%. One of the key reasons for this performance is use of controlled environment as mentioned
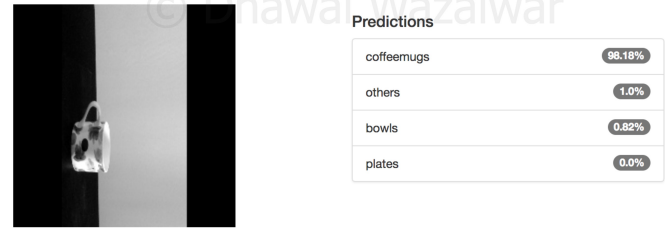
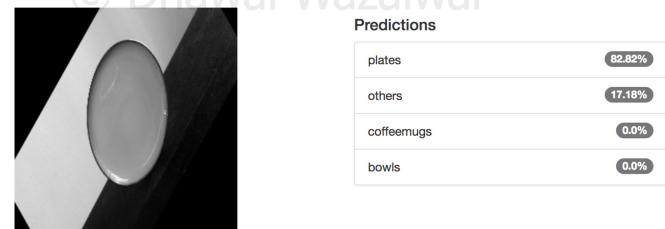Fig. 10. Test Case 1 showing probability for each class

Fig. 11. Test Case 2 showing probability for each class

in data acquisition section. Using such setup results in less noisy images, there by less complex data and thus even with such small size dataset, these results can be achieved.

Data augmentation is also a key aspect when working with small datasets and even on bigger datasets it can play crucial role in deciding overall testing results. It is important to know and be aware of expected test environment. For example here in Task2 testing, it can be expected that coffee mugs, bowls or even plates will have different orientation and model needs to handle them. So in data augmentation step, original image was rotated at 5 different angles and added to training database. This helped to achieve improved performance on tilted images too.

Accuracy vs Inference time is a key trade-off point and a very tough decision to make in robotics systems some times. In case of Task2 for kitchen objects sorting, project scale becomes important factor. For example assume aim is to develop a robotic system that can sort all objects in commercial kitchen and then later placed on a separate big shelf. Now there are 3 such systems developed: System 1 has accuracy of 75% only but inference time is low 1ms, System 2 has accuracy of 85% but inference time is more about 5ms and finally System 3 has accuracy of 95% but inference time is very high 20ms. For commercial kitchens or say restaurants for that matter, time is a very important factor since primary requirement for them is to cater to more customers. 20ms Inference is bit too high for them and so they would not prefer 3rd system even though accuracy is close to 95%. They would also not want System 1 with very low accuracy of 75%, as that will increase their manual work and thereby reduce overall productivity. So as a trade-off assuming both accuracy and inference time expectations are met, 2nd system may be the recommended one.

6 CONCLUSION / FUTURE WORK

For Task1 and Task2, in both cases using GoogleNet from DIGITS workflow expected performance metrics were achieved. As discussed above in case of Task2, it was only possible as a controlled setup was used. However for most commercial systems, thats not always the case and robotic system has to handle lot of uncontrolled factors like different lighting conditions, presence of more than one object in an image and sometimes-noisy conditions in terms of presence of dirt etc. For developing a robotic inference solution to handle such tasks, one has to make sure all such cases are included in our training database.

Commercial kitchen robotics market is going to increase and with increased research, smart kitchens are no longer only a dream anymore. Specific to Task2, such sorting can be a first step in some robotic systems. Sorted kitchen objects means place of each object is known to robot and then based on this information it can act to either to load/unload say a dishwasher or pick up required object to make online recipe.

REFERENCES

[1] S. S. Srinivasa, D. Ferguson, C. J. Helfrich, D. Berenson, A. Collet, R. Diankov, G. Gallagher, G. Hollinger, J. Kuffner, M. V. Weghe, and et al., "Herb: a home exploring robotic butler," *Autonomous Robots*, vol. 28, no. 1, p. 520, 2009.

[2] M. Gibson, "Meet The Robot Chef That Can Prepare You Dinner." http://time.com/3819525/robot-chef-moley-robotics/, 2015. [Online; accessed 06-Jan-2008].