Classification of spices using Neural Network

Shweta Ruparel

Abstract—This paper presents how neural networks can be used effectively to classify spices and use the power of deep learning to create robotic kitchen in nearly every restaurant and household.

Index Terms—Robot, IEEEtran, Udacity, LaTeX, deep learning,spices classification.

1 INTRODUCTION

IMagine, after a tiring day when you get back home and you're left with no energy to cook something for yourself,you would rather think to order online.

How about thinking of owning a robotic arm(your robotic chef) just like a human arm that could cook for you. Unbelievable isn't it ? With the image classification and object recognition using Deep Neural Networks it is possible that soon we will encounter robotic kitchen in nearly every restaurant and household, where robot chefs will cook delicious foods for us on demand.

It's already a reality in restaurants. Here are a few stated below-

- Spyce, a restaurant based in Boston, replaced human cooks with robot chefs to elevate their customers' taste buds.
- A restaurant named Nagoya based in Japan aims to use robots to attract customers while driving labor efficiency.
- A UK-based robotics company, Moley, has built a robotic kitchen that learns recipes, cooks, and cleans up the mess.
- Developing countries like India are wholeheartedly welcoming this tech innovation now.

In this paper, you will find classification of few spices viz, cumin seeds, mustard seeds,clove and cinnamon sticks using two DNN Architectures Alexnet and GoogLeNet.The data is collected using integrated camera of the laptop and a python script.The collected dataset is trained with GoogLeNet and Alexnet for accurate classification. Once the network is trained and classify the images accurately, then we can apply it to real time kitchen environment where a robotic arm with a camera and other mechanism can use the trained network to classify the real time data from the spice rack and pick the right spice and ingredients to cook a meal.The results from two different networks will be compared and analyzed.

There are innumerable ideas and information available online. Find below few references.

- Will Chefs Lose Relevance In A Robotic Kitchen?
- MIT engineers replace chefs with machines at world first robotic kitchen
- Image-to-Recipe Translation with Deep Convolutional Neural Networks

1.1 Evaluation using supplied Data

The Supplied data was trained and tested to find out the speed and accuracy of different pre-supplied training models/external network or even customized. The trained model should have at least an inference time of 10 ms or less on the workspace and an accuracy greater than 75 percent. Evaluate command is used on DIGITS Server to check the inference speed of the trained model for a single input averaged over ten attempts for five runs.It uses Tensor RT 3.0 to achieve this in a very fast time. Then the trained model is tested on a test set that is not used for testing or validation.

Alexnet and GoogLeNet was used for training the supplied dataset. Both models achieved far less time that 10 ms and GoogLeNet achieved accuracy of 99.7, just few percentile higher than Alexnet.

1.1.1 *GoogLeNet trained Model*

GoogLeNet Trained Model performed better both in classification and accuracy.Find below the figures that shows the sample dataset provided,loss and accuracy graph for GoogLeNet Trained Model, Learning Rate and Epoch Graph for GoogLeNet Trained Model and Evaluate response.

- Figure 1, shows the supplied dataset.
- Figure 2 , shows the graph for loss and accuracy while training.
- Figure 3,shows the graph for learning rate at different iterations of going through the whole dataset.Epoch set to 8.
- Figure 4, shows the evaluate command on DIGITS server.It uses Tensor RT 3.0 to achieve this in a very fast time.

Figure 1. Supplied Data

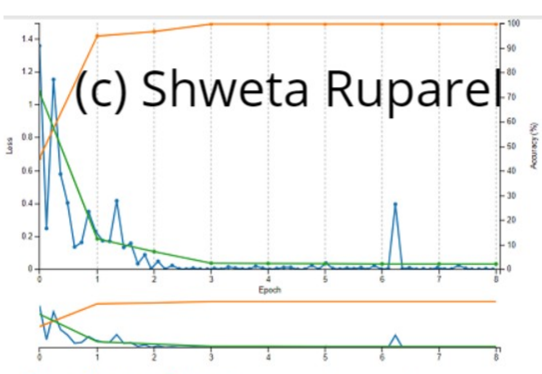

Figure 2. Loss Accuracy Graph

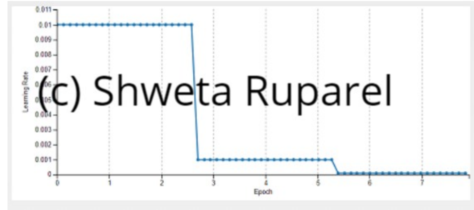

Figure 3. Learning Rate and Epoch

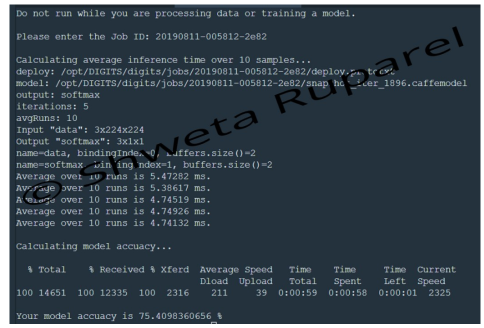

Figure 4. Evaluate Command

2 BACKGROUND / FORMULATION

The NVIDIA GPU was used to train the supplied data as well as collected data.The supplied data by Udacity was taken from a Jetson mounted over a conveyor belt.Pictures of candy boxes, bottles, and nothing (empty conveyor belt) is trained for the purpose of real time sorting. Huge supplied dataset of 7570 images was used to find out how different models achieve.Alexnet and GoogLeNet was used to train using epoch of 8 , which was enough for this large dataset.After training the evaluation was done using evaluate command and the result for GoogLeNet is shown in figure 4.

The collected data is a set of images of common spices that are used in cooking. A small set of 169 images was taken for each of the spice for mustard seeds, cumin seeds, cinnamon sticks , clove and nothing. Alexnet and GoogLeNet was used, starting with epoch of 10. The results were not good.Then epoch was changed to 30 and the results improved drastically.Both the models achieved much higher accuracy than before.GoogLeNet worked pretty well with higher accuracy and less loss. It was successful in classifying the test images with very high accuracy. This is a big milestone towards creating a larger dataset of hundreds of spices and food ingredients and training the model.Refer the following figures to find out training results.

- Figure 5, shows the sample of the collected dataset.
- Figure 6, shows the model trained with Alexnet with epoch set to 30.
- Figure 7, shows the model trained with GoogLeNet with epoch set to 30.

Figure 5. Collected Data

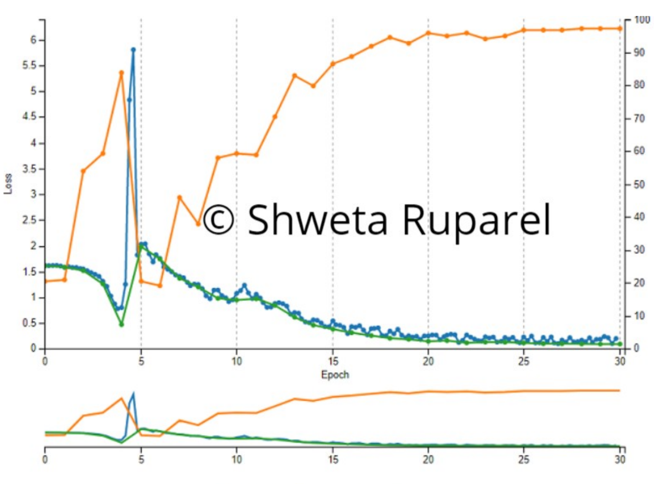

Figure 6. Alexnet Trained Model

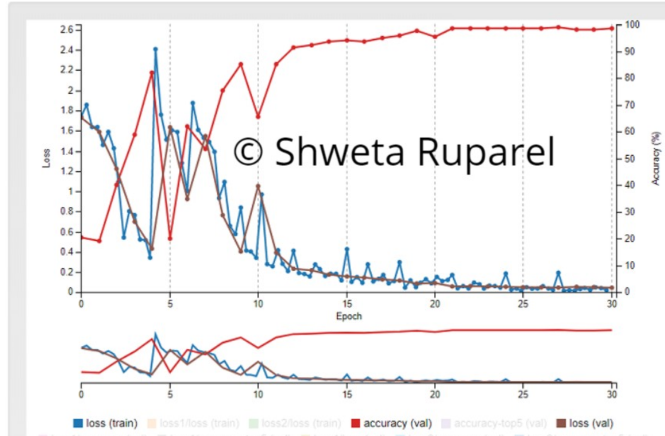

Figure 7. GoogLeNet Trained Model

3 Data Acquisition

The collected dataset has 845 images in total for 5 classes, 169 images in each class, viz. clove, cinnamon, mustard, cumin and nothing.It was a small dataset for the first trial iteration and it turned out to work very effective.It is always an important parameter to decide how big the dataset should be.Each image was taken using laptop camera and a python script.Each image is of dimension 640X480, RGB for clarity of picture details.The python script was run on the laptop and it started the integrated camera of the laptop.The images at different angles were taken by pressing escape button on the laptop and once the picture was taken, it was saved by the script in the right directory set in the beginning and changed for each class in the python script.Once the preferred amount of pictures were taken the script stopped, in this case it was 169.

Refer figure 5 for sample of collected dataset and figure 8 for python script.

```python
import cv2
# Run this script from the same directory as your Data folder
# Grab your webcam on local machine
cap = cv2.VideoCapture(0)
# Give image a name type
name_type = 'nothing'
# Initialize photo count
number = 0
# Specify the name of the directory that has been premade and be sure that it's the name of your class
# Remember this directory name serves as your datas label for that particular class
set_dir = 'nothing'
print ("Photo capture enabled! Press esc to take photos!")
while True:
    # Read in single frame from webcam
    ret, frame = cap.read()

    # Use this line locally to display the current frame
    cv2.imshow('Color Picture', frame)
    # Use esc to take photos when you're ready
    if cv2.waitKey(1) & 0xFF == 27:

        # If you want them gray
        #gray = cv2.cvtColor(frame,cv2.COLOR_BGR2GRAY)

        # If you want to resize the image
        # gray_resize = cv2.resize(gray,(360,360), interpolation = cv2.INTER_NEAREST)

        # Save the image
        cv2.imwrite('Data/' + set_dir + '/' + name_type + "_" + str(number) + ".png", frame)

        print ("Saving image number: " + str(number))

        number+=1
    # Press q to quit the program
    if cv2.waitKey(1) & 0xFF == ord('q') or number == 170:
        break
cap.release()
cv2.destroyAllWindows()
```

Figure 8. Python Script for taking images using Laptop

4 RESULTS

The test results are compared below in a table for both the models Alexnet and GoogLeNet trained for collected dataset for epoch set to 30.The classification was done using both the model , GoogLeNet classified each image 100 percent.Refer the figure shown below.

	accuracy	loss
Alexnet	0.973214	0.093379
GoogLeNet	0.986607	0.045578

Figure 9. This figure shows the classification done on an image by Alexnet Trained Model.The classification was done poorly.

Figure 10. This figure shows the classification done on an image by GoogLeNet Trained Model.The classification was done 100 percent.

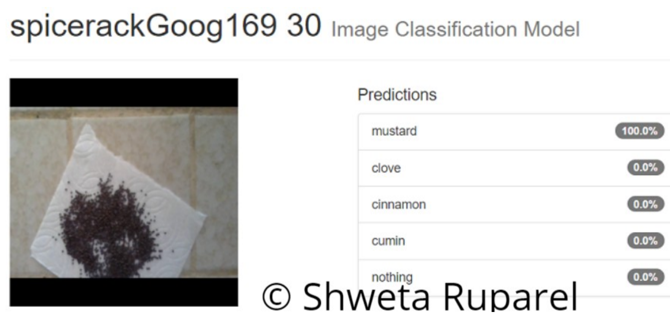

Figure 11. This figure shows the classification done on an image by GoogLeNet Trained Model.The classification was done 100 percent.

5 DISCUSSION

The accuracy on the collected dataset was great and the classification was 100 percent achieved using GoogLeNet with en epoch of 30.This is a big milestone to move further in this direction and imagine a robotic arm with other moving and grabbing mechanisms with a camera attached to take the images of the available ingredients and classify which ingredient is needed. Once classified the arm can successfully pick and place using kinematics.

Once the dataset is more complex which means more classes , it is recommended to take more images for each class and epoch can be experimented to reduce so that the time taken for training is not too high.

Both speed and accuracy is important for cooking. Accuracy is more important because one wrong ingredient can spoil the dish.

6 CONCLUSION / FUTURE WORK

GoogLeNet achieves greater accuracy and classification even for the smaller dataset and that makes it a great model.More classes for spice and other food ingredients can be created and trained with larger dataset, since less number of classes (number of items to classify, 5 in this project), means less complexity in terms of shape, size, color. This could become very complex as the number of items increases, so a larger dataset is recommended to train the model. This could be a great milestone in making the Robotic Kitchen in every household a reality.

There are many research papers that focuses on deep learning the food item from the cooked food image to classify the ingredients.We need to focus more on reversing the situation and deep learn the food ingredients and program to cook meals by feeding thousands of recipes available online.Find below papers referred.

DeepFood: Automatic Multi-Class Classification of Food Ingredients Using Deep Learning [2], Deep learning -based Ingredient Recognition for Cooking Recipe Retrieval [1]

REFERENCES

[1] Jingjing Chen and Chong-Wah Ngo. "Deep-based Ingredient Recognition for Cooking Recipe Retrieval". In: *ACM Multimedia*. 2016.

[2] Lili Pan et al. "DeepFood: Automatic Multi-Class Classification of Food Ingredients Using Deep Learning". In: Oct. 2017. DOI: 10.1109/CIC.2017.00033.