Multiclass Food Classification

Introduction to Computer Vision – Group Project

Syed Zuhair Abbas
Undergraduate
Innopolis University
Innopolis, Russia
s.abbas@innopolis.ru

Kusal KC
Undergraduate
Innopolis University
Innopolis, Russia
k.kc@innopolis.ru

Bekpasha Dursunov
Undergraduate
Innopolis University
Innopolis, Russia
b.dursunov@innopolis.ru

***Abstract*—This document serves as the final submission report for multi-class food classification in video. We used transfer learning, took the pretrained Inceptionv3 model, added some layers finetuning some hyperparameters, and trained upon the 101 Food Dataset and eventually attempted to use this model to try and classify a video containing food items.**

Keywords—classification, food, multiclass, image, video, neural network.

INTRODUCTION

Computer vision and Machine learning have been a hot topic in research in the past decade. The substantial evolvement of technology and rapidly improving computation resources has allowed us to process a large amount of data and derive practical insights from them. Advanced AI models are taking over many industries and performing human tasks with high accuracy and precision. Food classification is one of the applications of image processing. There are quite a number of researches and Machine Learning solutions already available when it comes to classifying food items from images. However, classifying food items directly in a video is a slightly more complicated problem which can be approached in many different ways.

RELATED WORKS

There have been various food classifications already done. Some of those works are briefly reported below. We actually took insights from some of them for our project:

- Food Image Classification with Convolutional Neural Network, https://ieeexplore.ieee.org/document/8550005
- Analysis of food images: Features and Classification, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5448982/
- Multiclass Food Classification, https://www.kaggle.com/theimgclist/multiclass-food-classification-using-tensorflow
- Build Video Classification Model, https://www.analyticsvidhya.com/blog/2019/09/step-by-step-deep-learning-tutorial-video-classification-python/

IMPLEMENTATION

The entire implementation structure consists of two parts: an image classifier and a video stream processor.

The image classifier is trained on top of the Inceptionv3 model with ImageNet weights so as to leverage the benefits of transfer learning. To train this classifier, we used the Food 101 dataset. Here's a quick overview of the dataset:

Total food classes: 101

Total images: 101000

Total training images: 75750

Total test images: 25250

We finetuned the model with the following settings:

Batch size = 32

Two dense layers: one with "Softmax" activation function and L2 regularizer with a value of 0.005, and the other with "relu" activation function and a dropout of 0.2.

The model is compiled with an SGD (Stochastic Gradient Descent) optimizer with hyperparameters: (learning rate) lr=0.0001, and momentum=0.9.

Given the huge size of dataset (approximately 5 GB), the model was trained for only 10 epochs which took almost 6 hours even in Google Collaboratory with GPU runtime enabled.

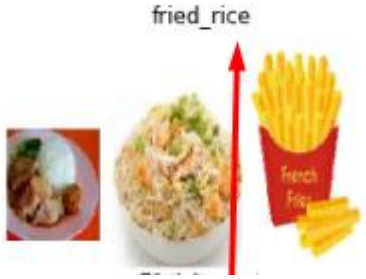
The next step was to create a pipeline to apply the model on a video. We decided to do this on a frame-by-frame basis. This allowed us to dissect the video and get each frame which would then be classified with the model we trained before.

GitHub Link:

https://github.com/thisisbeka/ICV-project

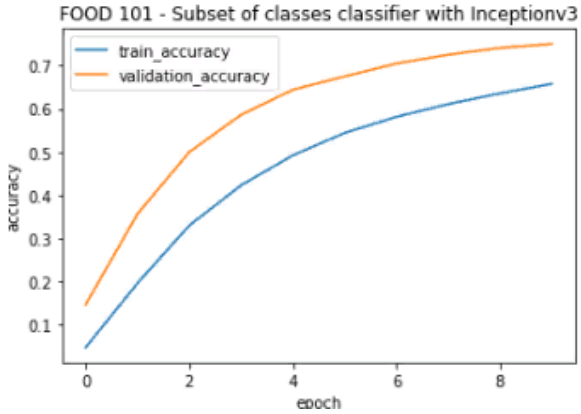
EXPERIMENTS & EVALUATION

In order to see how our algorithm works we have taken the video from youtube and ran it with predefined setup of classifier. At a time 1:58 there is a fried rice image and our algorithm is also detected it as a fried_rice.

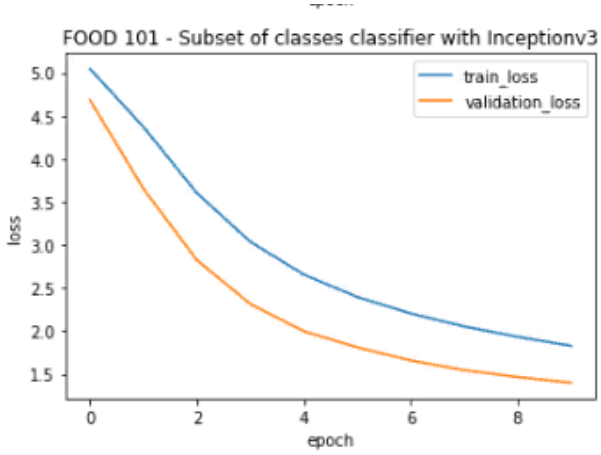


This was rather a difficult frame as it had multiple food items within. But as you will see in demo, we will only consider video with only one food item per frame for now.

We can evaluate the model as we trained for 10 epochs. Now, if we take a look at accuracy vs epoch plot we can see the following: The train accuracy at epoch 10 is around 75 % which is really good result, since the dataset is too big and train process takes a lot of time, we can easily guess that at 15-17 epoch our accuracy would be around 90 %, since the graph is following the exponential growth.



In case of loss we can see that it is getting closer to 0 in each epoch. Which means that there is low difference between the predicted output and the actual output.



ANALYSIS & OBSERVATIONS

To improve the performance of our classifier, we varied the optimizer's learning rate value as well as batch size, and dropout rate. However, since the dataset would take too long to train with such changes, we observed the consequences on only a subset of data. In fact, data belonging to only 3 class labels were chosen and subjected to the analysis.

It was found that having too much learning rate overfit our model. We actually naively tried from 0.01 and exponentially decreased to 0.001 and so on until we found 0.0001 to be effective enough. Also, we did the same with batch size and dropout rate. There was a trade-off when varying all these at once, but we found an effective set with the chosen values which actually gave about 98% accuracy on food images when trained upon only 3 classes. But when we included all 101 classes, our validation rate was only about 75%. We only trained it for 10 epochs.

REFERENCES

Food 101 Dataset used in training:
https://data.vision.ee.ethz.ch/cvl/datasets_extra/food-101/

Multiclass Food Classification,
https://www.kaggle.com/theimgclist/multiclass-food-classification-using-tensorflow

Large Scale Image Classification using pre-trained Inception v3 Convolution Neural Network Model,
https://towardsdatascience.com/classify-any-object-using-pre-trained-cnn-model-77437d61e05f

Image Classification Transfer Learning with Inception v3, https://codelabs.developers.google.com/codelabs/mlimmersion-image-flowerstxf/index.html?index=..%2F..cloud#0