Indian Food Image Classification with Transfer Learning

Rajayogi J R
Department of CSE,
RVCE, Bengaluru
rajayogijr.scs18@rvce.edu.in

Manjunath G
Department of CSE,
RVCE, Bengaluru
manjunathg.scs18@rvce.edu.in

Shobha G
Department of CSE,
RVCE, Bengaluru
shobhag@rvce.edu.in

Abstract— **Image classification has become easier with deep learning and availability of larger datasets and computational resources. The Convolutional neural network is the most popular and widely used image classification technique in the recent days. In this paper image classification is performed on Indian food dataset using different transfer learning techniques. The food plays important role in human's life as it provides us different nutrients and hence it is necessary for every individual to keep a watch on their eating habits. Therefore, food classification is a quintessential thing for a healthier life style. Unlike the traditional methods of building a model from the scratch, pre trained models are used in this project which saves the computation time and cost and also has given better results. The Indian food dataset of 20 classes with 500 images in each class is used for training and validating. The models used are InccceptionV3, VGG16, VGG19 and ResNet. After experimentation it was found that Google InceptionV3 outperformed other models with an accuracy of 87.9% and loss rate of 0.5893.**

Keywords— Convolutional neural network, Google inception v3 model, VGG16, VGG19, ResNet, Transfer learning, Food classification Introduction (Heading 1)

I. INTRODUCTION

Research in Computer vision and machine learning has made image classification easier because of availability of huge data and computational resources. There are various techniques that can be used for image classification like KNN classifier on local and global features used in [9], Artificial Neural networks, SVM and Random forest technique used to classify 11 different classes using different bag-of-features in [10] and many. But these methods fail when the dataset is large. Since the Convolution neural network can easily handle large amount of data and provide high classification accuracy, it has gained attention in the area of image classification recently.

Training of CNN for image classification can be done mainly in 2 ways: training the CNN from the scratch or using the concept of transfer learning. Transfer learning is a deep learning technique where a model is trained to learn and store the knowledge from one problem and use the same model to other similar problems. i.e., fine tuning already trained CNN models from the huge dataset to food image classification task. The Pre-trained models used are Inception V3 [1], VGG16 [2], VGG19 and ResNet. These are the top performing models in the annual Imagenet large scale visual Recognition challenge (ISLVRC) [5] giving considerable accuracy and less validation loss.

Food Classification is gaining popularity slowly because of consciousness of health and food among people. According to the World Health Organization (WHO) [7], it is estimated that more than 1.9 billion adults who are aged above 18 years were overweight. It is very shocking to know that 13% of the world's population including both men and women (11% men and 15% women) are obese. In fact, the number of people across the world that are suffering from obesity has doubled since 1980. The numbers above show that food plays a vital role in physical health of a person.

Statistics show that 95% of the people do not follow any nutritional plan as these are very strict and restricts people from consuming their day-to-day food. Old aged who want to monitor their food intake, patients who want to monitor their health through food due to different dietary restrictions and mainly youth who want to track the calories and nutrition intake to maintain fitness, the importance of food classification has increased. Over the past couple of years, image based dietary and calories extraction has been a challenging task and a lot of research is going on the same.

Through this research an effort has been put to classify Indian food images into their respective classes using transfer learning. Image Classification with deep learning techniques such as Convolution neural network are getting incredible consideration because of their efficiency in learning and classifying complex features. A comparison has been made between the models with respect to accuracy and validation loss.

The organization of the paper is as follows: the related works are presented in section II. Section III deals with the proposed methodology followed by the experimental results, conclusion and future work in section IV and section V respectively.

II. RELATED WORK

The emergence of deep learning has proved to be very beneficial in dealing with huge datasets of images. Regarding the food image recognition there are many advancements in the literature in the past few years. The efforts from [3] proposed a new deep learning convolution neural network configuration that detects and recognizes the local food images. The work used the local Malaysian food dataset for their study which was collected from the publicly available internet sources including the commonly used image search engines. It was observed that the CNN method achieved far better accuracy score when compared to the traditional methods. However, it was quite evident from [3] that network depth is significant for better performance of the model. Indian food dataset used in this research is collected from the internet from various websites as in [3].

The tasks involving identification, extraction and classification exclusively from image data such as the pixels is in itself a very challenging thing to do. However, efforts were made in [4] to use the convolution neural network for classification of the food images which got 86.97%

978-1-7281-2619-7/19/$31.00 ©2019 IEEE

accuracy. The authors have used the popular food-101 dataset in their study and also compared their results with other models in a well-defined table. The only flaw as mentioned by the authors is that it requires a lot of computational time to train the CNN network. They have concluded that although it takes more computational time and resources, CNN stands out in performance when it comes to image classification with more number of classes. Hence, CNN is used for Image classification in our research and is carried out using Google colab platform which provides free GPU for building deep learning models.

In [6], a prediction model to classify the images of Thailand fast food has been proposed. In [6] authors have used Thai Fast Food dataset (TFF) and Inception V3 to classify the images and have achieved an accuracy of 88.33%. The TFF dataset contained about only 3960 images with 11 classes, which is very less for cnn model to learn as per the best practices.

Four different pre-trained image classifying models namely the Inception v3 [1] model from the Google, VGG16 [2] and VGG19 by Oxford's renowned Visual Geometry Group (VGG), which achieved very good performance on the ImageNet dataset and ResNet based on residual network [8] have been used here. These models are widely used in image recognition, image classification and image processing as it works efficiently with large number of images and provides high accuracy results.

In VGG16 [2] and VGG19 only 3x3 convolution layers and 2x2 pooling layers are used throughout the network. VGG model depicts that depth of the network plays a key role as deeper networks give better results by learning more features. There are 2 fully-connected layers, each with 4,096 nodes followed by a softmax activation layer as a classifier. Both VGG16 and VGG19 networks has an image input size of 224-by-224. Both the models are known for their simplicity, using only *3×3* convolutional layers stacked on top of each other increasing the depth of the network. The "16" and "19" stand for the number of weight layers in the network

In the VGG16 and VGG19 architectures, there are more consecutively stacked 3x3 convolutional layers in the higher layers than the lower layers due to speed and memory constraints. More convolutional layers at the higher layers will make everything slower and consumes more GPU memory and lower layers have a high spatial resolution and hence dominate the storage and computational constraints. To overcome this the most recent models like google inception and ResNet reduce the spatial resolution at first before adding any convolution layers.

ResNet uses micro architecture models which are also known as network-in-network architectures. The Network-in-network architectures in ResNet are referred as building blocks which are used to build the network. This is not the case in traditional sequential networks like AlexNet and VGG. These micro-architectures are grouped to form macro architecture. This was first introduced by He et al. in their 2015 paper [8]. The ResNet architecture has turned into an important work, illustrating those extremely deep networks can be trained using Stochastic Gradient Descent through the use of residual modules. Regardless of the way that ResNet is deeper than VGG16 and VGG19, the model size is smaller on account of the utilization of global average pooling rather than fully-connected layers which decreases the model size down to 102MB for ResNet50.

The GoogleNet [1], refers to the Inception architecture developed by Szegedy et al. is a deep convolution neural architecture that was codenamed, Inception. The objective of the inception module is to behave like a multi-level feature extractor by using 1×1, 3×3, and 5×5 convolutions inside a single module of the network, then the result of this module are fed as input to the next layer in the network. It scored really well with detection and classification in the ImageNet Large Scale Visual Recognition Challenge 2014 (ILSVRC14) and also bagged the first place. It was implied by [1] for vision networks and covering the hypothesized outcome by dense, readily available components. With a bit of tuning, modest gains were observed compared to the other reference networks. Inception V3, the latest version has been used to build a classifier in this paper .

Similar kinds of work were done in [7] and [9]. The study in [7] has come up with a design model using CNN to distinguish the nutrition group of food. For this purpose, two pre-trained models namely Alexnet and Caffenet were fine tuned. The dataset was generated from Food-11, Food-100 and even from web archives. The result shows that fine-tuned models from transfer learning performed better compared to building the network from scratch.

III. PROPOSED METHODOLOGY

A. The Indian Food Dataset.

The dataset considered for our study is the Indian Food dataset. It contains 20 different classes of food and each class has 500 sample images. The dataset inherently comes with a lot of noise since there are images in which there is more than one food item. The image samples also contain a lot of color and few of them are wrongly labeled too. The figure below shows the sample food images from the Indian Food dataset.

Fig. 1. Samples from the indian food dataset

B. Image preprocessing

The dataset contains 20 different classes of food images. Each class of image is divided into training and testing images wherein 400 images from each class are considered as training samples and the remaining 100 samples as test samples. Overall, there are 8000 training samples and 2000 test samples. The training set images are fed to the CNN model and validation is made using the test dataset. Figure 2 below depicts the proposed methodology for our research.

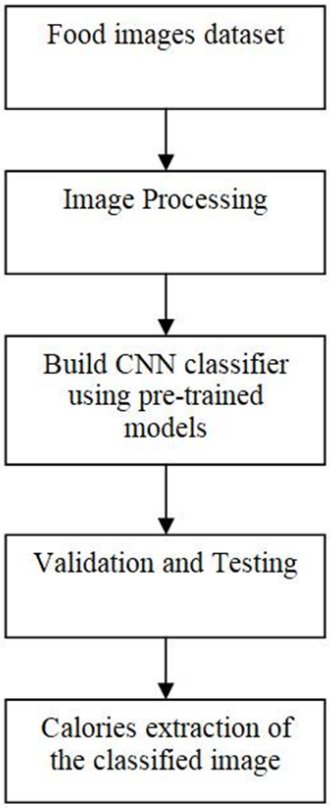

Fig. 2. The proposed methodology

C. *Training the CNN classifier using pretrained models*

The greater part of the Computer Vision Problems doesn't have exceptionally huge datasets (10,000 images—50,000 images). Indeed, even with extraordinary data augmentation procedures it is hard to accomplish respectable accuracy. Building these networks with many parameters over fit the model. So Transfer learning avoids this over fitting. In transfer learning early layers will detects edges, middle layers detect the shapes and the last layers will detect some high level data features. These transfer learning models are useful in many computer vision and image classification problems.

80% images were used for training and 20% was used for validation. Each model was run for 30 epochs with a batch size of 32. The input image size for InceptionV3 was 299x299 and 224x224 for the rest of the models. Stochastic Gradient descent optimizer was used which update and tune the model's parameters in backward direction so that we can minimize the Loss function. Loss function used is Categorical cross entropy which is a measure of how good a prediction model does in terms of being able to predict the expected outcome. Evaluation metrics used is categorical accuracy which is a function that is used to judge the performance of the model. Dropout of 0.2 is used, which helps prevent over fitting. The pooling layer used here is global average pooling and softmax function is used at the end for classification which classifies images based on the probabilities.

D. *Validation and Testing*

Once the model is trained using the train dataset (the sample of data used to fit the model) then validated using validation dataset (The sample of data used to provide an unbiased evaluation of a model fit on the training dataset while tuning model hyper parameters.) and finally tested using the test dataset (The sample of data used to provide an unbiased evaluation of a final model fit on the training dataset.)

E. *Calories extraction of the classified image.*

Finally, our classifier can be used to estimate the calorific content of the classified food from the internet. Suitable python or any scripts can be used to perform web scraping to fetch the nutrition facts for the classified image from the web and provide it to the user.

IV. EXPERIMENTS AND RESULTS

The experiment was carried out using Google Colaboratory, a research tool for machine learning education and research. It's a Jupyter notebook environment that requires no setup to use. It is a free research tool provided by Google that facilitates to run codes that require high performance GPU.

. The transfer learning models which are used here are Google InceptionV3, VGG16, VGG19 and ResNet. By pre-trained model, the model is already built over the other dataset such as the ImageNet and we add layers on top of those models, as per the requirements of our classification. i.e., the number of classes.

Obtained result:

The results for all the models are tabulated below. The parameters used here are validation accuracy and loss rate.

TABLE I
COMPARISON OF ACCURACY AND LOSS RATE OF DIFFERENT MODELS

MODEL	ACCURACY	LOSS RATE
Inception v3	0.879	0.5893
VGG19	0.789	0.7725
VGG16	0.782	1.1035
ResNet	0.6991	1.0804

InceptionV3 has produced highest accuracy of 87.9% and least error rate of 0.5893 followed by VGG19, VGG16 and

ResNet. VGG19 which is 19 layers deep performed well against VGG16 which is 16 layers deep. Resnet with 1.0804 has less validation loss rate compared to VGG16 as it uses residual modules.

V. CONCLUSION AND FUTURE WORK

In this research study, the Convolutional Neural Network, a Deep learning technique is used to classify the food images in to their respective classes. The dataset considered is the Indian food dataset and we were able to achieve accuracy of 87.9% in case of the inception V3 model compared to other models such as the VGG19 that produced 78.9%. The VGG16 model and the ResNet model were able to produce accuracy of 78.2% and 69.91% respectively.

As far as the future enhancement is concerned, the task of classification can be improved by removing noise from the dataset. The same research can be carried out on larger dataset with more number of classes and more number of images in each class, as larger dataset improves the accuracy by learning more features and reduces the loss rate. The weights of the model can be saved and used to design a web app or mobile app for image classification and further calories extraction of the classified food.

ACKNOWLEDGMENT

This experiment was carried out in R.V. College of engineering, Bengaluru, Department of Computer Science and Engineering, under the guidance of Dr. Shobha G. We also thank our Head of the Department Dr. Ramakanth and the principal of the institution, Dr. K N Subramanya for providing us good lab facilities and suitable environment to successfully complete our experiment.

VI. REFERENCES

[1] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016).Rethinking the inception architecture for computer vision. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2818-2826)

[2] J. D. A. Berg and L. Fei-Fei, "Large scale visual recognition challenge 2010," http://image-net.org/download, 2010, [Online; accessed 29-Jan- 2018].

[3] A Deep Convolutional Neural Network for Food Detection and Recognition by Mohammed A. Subhi and Sawal Md.Ali

[4] Food Classification from Images Using Convolutional Neural Networks David J. ttokaren, Ian G. Fernandes, A. Sriram, Y.V. Srinivasa Murthy, and Shashidhar G. Koolagudi

[5] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh,S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, et al. Imagenet large scale visual recognition challenge.2014. 1, 8

[6] Thai Fast Food Image Classification Using Deep Learning by Narit Hnoohom and Sumeth Yuenyong

[7] Comparison of Convolutional Neural Network Models for Food Image Classification by Gözde ÖZSERT YİĞİT and Buse Melis ÖZYILDIRIM

[8] Deep Residual Learning for Image Recognition,Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun

[9] Food Calorie Measurement Using Deep Learning Neural Network by Lukas Bossard and Matthieu Guillaumin and Luc Van Gool