Measuring Calories and Nutrition from Food Images

Atul Kumar Tiwari
T.C.E.T.
Mumbai, India
atulktiwari310@gmail.com

Punit Savlesha
T.C.E.T.
Mumbai, India
p.savlesha@gmail.com

Yash Shetiya
T.C.E.T.
Mumbai, India
yashxlency0410@gmail.com

***Abstract*— The importance of calorie tracking is gaining wide acceptance not only in fitness sector but also in daily lives of people. It becomes important to ensure that process is simple, fast, and effective. The existing fitness applications require manual entry of food items to calculate calories intake. The advancement in field of computation and computer vision has led to rise of deep learning algorithms like CNN and platforms such as TensorFlow and Keras for effective and accurate image classification. In this paper, a fitness application that uses image processing to detect the food name and calculate the calories has been implemented. We have developed a mobile application that possess the ability to track calories on basis of food images using CNN trained neural network model.**

***Keywords*— Activation Function, Convolutional Neural Network (CNN), Deep Learning**

I. Introduction

Nutrition and calories tracking plays a significant role in health and fitness industry. The growing need of diet tracking in the world has evolved many applications that show you the calories intake a food product can provide, our application does it in a different manner. The application will take the image of the food product as the input from the user and process it to show the output i.e. the calories and nutrients in the food item. The mobile application has been developed in Dart programming and Flutter SDK. The core part of application is image processing model developed using Python's Keras and Tensorflow model. Convolutional Neural Network (CNN), a powerful deep learning algorithm for image processing and machine vision has been used to solve the image classification. The CNN model provides the capability to classify an input image into one of the pre-trained labels according to the features extracted by initial layers from input image.

II. LITERATURE SURVEY

Ch. Kavya, R. Priyadarsini and B. Madhavi in their paper titled "Calories and Nutrition Measurement from the Image of Food" designed a system to extract features like color, shape, size, and texture[1]. They used K-Nearest Neighbor (KNN) to tackle the image classification problem by training on food database. Hemalatha Reddy V., Kumari S., talked about a similar system in their paper- "Food Recognition and Calorie Measurement Using Image Processing and Machine Learning Techniques". They implemented with help of K-Means clustering [2]. I. Culjak, D. Abram's paper "A brief introduction to OpenCV," explained about practical usage of OpenCV library in image processing and machine vision . Intel introduced OpenCV library decades ago, it is an opensource library that helps users to deal with images and videos. The most peculiar thing about OpenCV is the fact that despite being an open-source tool, it is capable to giving results as good as commercial libraries developed by organizations [3].

The conventional approach of image classification has been pre-processing, feature extraction and classification by using classifiers like SVM, KNN, ANN, K-Means etc. These models have certain limitations as they are based on Machine Learning (ML) algorithms. The existing projects have managed to detect the food item and calories of naturally available food items such as fruits and vegetables, but these models fail in case of processed food. The main crux of food classification lies in combination of low-level and high-level features extracted by the neural network or ML algorithms. There are numerous fitness applications that are available in market but none of them offer a mechanism of calorie tracking through food image. The existing work have tackled the classification problem using Machine Learning approach. The ML approach suffers from limitation of feature extraction. The features to be extracted are generally hand-picked and hard coded by developer as per their understanding of domain. The process of feature extraction becomes cumbersome and relatively difficult as size of training dataset increases.

In contrast to Machine Learning, a stack of neural network layers is generally created by deep learning. The most important difference between deep learning and traditional machine learning is its performance as the scale of data increases. When the data is small, deep learning algorithms do not perform that well. This is because deep learning algorithms need a large amount of data to understand it perfectly [4].

Hokuto Kagaya's research titled "Image-based Calorie Content Estimation for Dietary Assessment," used multiple layers of CNN architecture to create a powerful model which was capable of detecting and classifying food items[5]. CNN showed tremendous success in handling complex and diverse training dataset.

Convolutional neural networks (CNN) are one of the most popular and powerful paradigms available in deep learning for solving image classification and other machine vision related problems. They make use of huge amount of low-level and high-level features to create a feature map and perform complex classification task [6]. Convolutional neural networks consist of two major layers, namely convolutional layers, and pooling layers. [7].

III. IMPLEMENTATION

The Image processing model was implemented using Convolutional Neural Network model through Keras and TensorFlow in python. The basic idea of CNN is to extract low level and high-level features from training dataset, assign random weights and biases and eventually fine-tune and learn these values. These weights and biases along with suitable activation functions are stored in the model which is generated by batchwise training over train dataset. The model can classify for an unseen input data, generally referred as test data because of these values.

We created our own CNN architecture using Keras. The model type was Sequential, which allows to build a model layer by layer. Four convolutional layers along with pooling, flattening and dense layer. Gradient, color, dimensions, edges are few low-level features that are extracted by convolutional layer. The addition of more layers enables the model to capture high level features and hence a proper understanding of each image is established. Convolutional layer is accompanied by Pooling layer. There are two types of Pooling operations – increment in dimension and decrement in dimensions. A holistic and effective understanding of image dataset is achieved because pooling layer has the capability to learn high level rotational features [8].

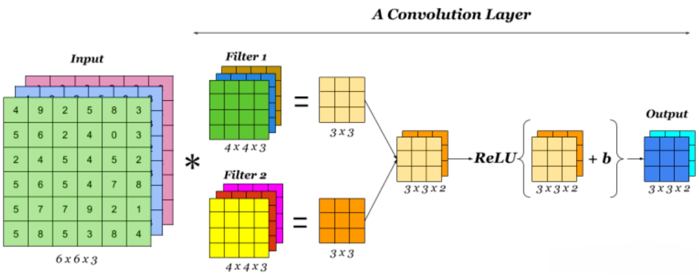

Fig. 1: Convolution Operation

The max pooling operations results in maximum value from the product of filter and area of image covered by the filter. In contrast to this, average pooling method returns the mean of all the values obtained by convolution operation in each portion of image. The above process ensures that model has successfully learnt all features of training dataset. Now, the feature map is flattened and feed to a regular Artificial Neural Network (ANN) to perform classification operation. The ANN employs backpropagation technique for each epoch. Eventually, the algorithm is capable enough to distinguish between features of input images and hence perform classification task for unseen data. The next process involved compilation of model using loss function, metrics, and optimizer. We selected 'adam' optimizer. The purpose of optimizer is to constantly adjust the learning rate throughout the training. The learning rate of model impacts the values of weights and biases, which eventually impacts the accuracy of model.

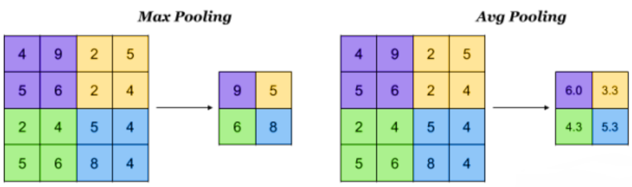

Fig. 2: Pooling Operation

Loss functions are widely used in machine learning to quantify model's performance. There are several loss functions such as L1 loss, L2 loss, tanmito loss, logistic, hinge, Chebyshev, cross entropy etc. We used 'categorical cross entropy'. It denotes a number which model aims to minimize by learning over training dataset [9]. We can understand loss function as optimization parameter.

The model was trained on 1276 data images of 17 classes. The dataset consisted of multiple images with different angle, color, texture, etc. We also implemented data augmentation methods such as mirroring, rotation, random cropping, and color shifting to create a robust dataset for training. The model was trained over 40 epochs. Epochs stands for a counter, which denotes number of times algorithm will iterate over training dataset. It denotes number of times algorithm will learn the features and adjust its weights and biases by going through whole training dataset [10]. It determines the number of times the model will iterate over the complete dataset. We found that there was drastic increase in training accuracy and validation accuracy of model with increase in number of epochs, till a certain limit. After that negligible increase in accuracy is observed after increasing epochs.

The model was tested on 565 images belonging to same classes and model was saved in h5 format. We created an interface in form of Android and iOS mobile application that provided feature to read an image from gallery. The mobile application was written in Dart programming language using Flutter SDK. It consists of various screens – welcome screen, login, home page and profile. The mobile application leverages TensorFlow lite technology to store the pre-trained neural network model and predict for new data. The home page contains menu widget that enables user to upload image from gallery, whenever a user inputs the food image model is loaded and prediction of food name is made. The application's source code contains a text file that stores calories, proteins and carbohydrate content associated with food item. These values are also fetched and shown as output to the user.

IV. RESULTS AND DISCUSSION

The methodology of CNN is quite like that of traditional ML algorithms – input, feature extraction, feature map, and classification. However, unlike ML algorithms the features are learned automatically by CNN network. The initial convolutional layers are responsible for feature extraction. The size of filter provided to Keras functions decided the dimensions of filter matrix in each convolution layer. The filter performs convolution operation by dragging the kernel matrix over same size portion of image and calculates product of matrix multiplication. The advantage lies in the fact that unlike traditional ML approach, we don't have to hard-code the feature extractors. They are automatically assigned by the CNN algorithm We found out that features like corner, texture, shape, and edge were extracted by neural network. These features are automatically selected and updated. We also found that CNN used canny edge detection method to detect edges of food images.

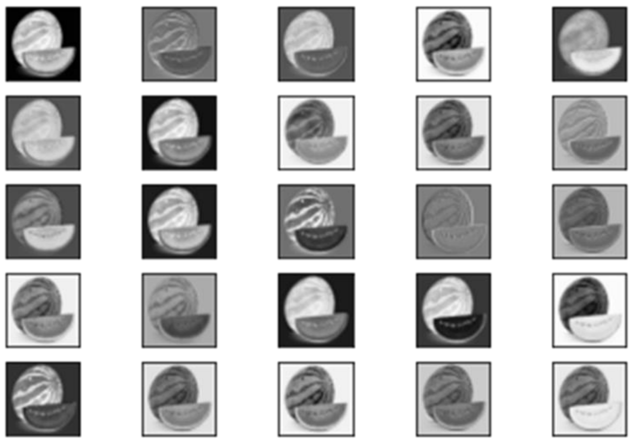

Fig. 3: A sample feature map created by CNN

Although CNN algorithm has capability to tackle highly complex problems, but it becomes important that parameters are selected appropriately. We performed hit and trail approach to select appropriate values for model. Various combination of number of filters in each layer, number of epochs and different activation functions were used in hyper-tuning the model in order to achieve an accurate and effective model.

Training Accuracy and Training Loss are metrics associated with model's performance with training dataset i.e., seen data, whereas Validation accuracy and validation loss refers to performance of model w.r.t. unseen data. Hence, we give more priority to validation accuracy and validation loss while deciding value of any parameter. The results were recorded in below tables.

Table 1: Performance of different activation functions

Sr. No.	Activation Functions	No of Epochs	Training Accuracy	Training Loss	Validation Accuracy	Validation Loss
1	Sigmoid	10	0.7543	0.8978	0.778	0.8035
2	ReLu	10	0.8425	0.4626	0.9618	0.2190
3	TanH	10	0.6071	0.5144	0.851	0.5865

Table 2: Number of filters in Hidden Layers of Convolutional Neural Network

Sr. No.	No of filters in Hidden Layers of Neural Network				Training Accuracy	Training Loss	Validation Accuracy	Validation Loss
	1	2	3	4				
1	8	8	8	8	0.5654	1.1468	0.6125	0.9834
2	16	16	16	16	0.5743	1.0226	0.6912	0.9541
3	8	16	32	64	0.7339	0.7960	0.8576	0.4392
4	32	64	128	256	0.8315	0.5117	0.9349	0.3176

Table 3: Performance of model with different epochs size

Sr. No.	Number of Epochs	Training Accuracy	Training Loss	Validation Accuracy	Validation Loss
1	10	0.8425	0.4626	0.9618	0.2190
2	20	0.9071	0.2780	0.9101	0.1210
3	30	0.9134	0.2202	0.9067	0.15494
4	40	0.9449	0.1888	0.9234	0.1204

Activation functions play significant role in neural networks by introducing the non-linearity property. It's absence simply equates to weighted sum of input and bias. They play pivotal role in decision making regarding activation of neurons in the architecture. Sigmoid, ReLu, TanH etc. are most widely used activation functions for CNN paradigm. We found that ReLu is suitable for this training dataset as it gave high validation accuracy and low validation loss. It also reduced computation time.

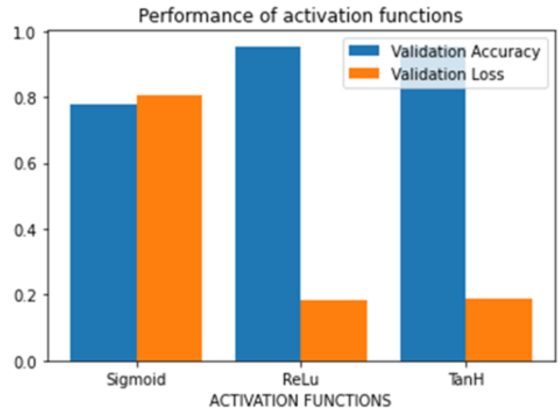

Fig. 4: Performance of activation functions

A CNN model with four convolution layer was created with various filter size. The reptation of same kernel over input image in a map of activations gives rise to feature map. We found that providing different number of filters in each hidden layers gives better result than same number of filters in every layer. We decided to use configuration of 32 – 64 – 128 – 256 for four hidden layers.

An accuracy metric provides a quantitative approach to measure the performance of model. It denotes the ratio of sum of true positive and true negative to sum of all predictions. Machine Learning makes use of variety of loss functions to learn optimum and effective weights and biases for particular dataset. It gives a quantitative metric that also denotes the performance of trained model on test/validation dataset. We aim for minimum loss model. We obtained an accuracy of 92.34% and loss of 12.04% with validation data.

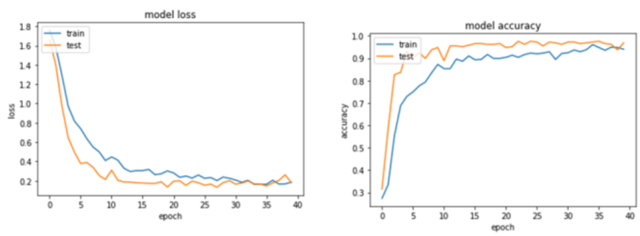

Fig. 5: Model loss and accuracy against epochs

We created an interface in form of Android and iOS mobile application that provided feature to read an image from gallery. The mobile application was written in Dart programming language using Flutter SDK. It consists of various screens – welcome screen, login, home page and profile. The mobile application leverages TensorFlow lite technology to store the pre-trained neural network model and predict for new data. The home page contains menu widget that enables user to upload image from gallery, whenever a user inputs the food image model is loaded and prediction of food name is made. The application's source code contains a text file that stores calories, proteins and carbohydrate content associated with food item. These values are also fetched and shown as output to the user.

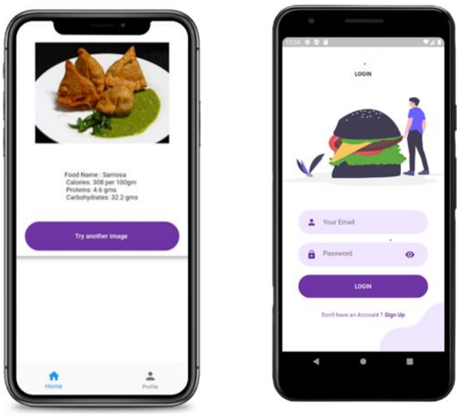

Fig. 6: User Interface of mobile app

V. CONCLUSION

We have developed an app which can classify 17 classes of different food items and display its class name along with its calories. The image processing model was developed using CNN algorithm. We also performed a comparative study of various activation functions, impact of filter size. Along with these, comparisons were also made based on number of epochs, and we concluded that 40 epochs are the best option for keeping the computational efficiency as well as the achieving high training and validation accuracy. In addition to displaying calories, the app also displays nutritional values such as carbohydrates, proteins, and fats of that food item.

The application has been developed on Model View Template (MVC) pattern of software engineering hence separation of concerns can be achieved. The application can be further integrated with backend database and APIs for development of industry ready product. We received suggestions for development of internal features like macro-based nutrient tracking, water intake tracker and workout tracker to enhance the application into a holistic fitness related product. We intend to expand the number of food labels on which neural network has been trained in future.

VI. ACKNOWLEDGMENT

It would be unfair if we do not acknowledge the help and support given by Professors, students, friends etc. We want to to express our sincere gratitude towards Ms. Purnima Chandrasekar, our project guide for her valuable feedback and support at every stage. We want to thank Thakur College of Engineering and Technology for giving us the opportunity to work on this project.

VII. REFERENCES

[1] Ch. Kavya, R. Priyadarsini and B. Madhavi, “Calories and Nutrition Measurement from the Image of Food” International Journal of Advance Research, Ideas and Innovations in Technology, https://www.ijariit.com/manuscripts/v3i3/V3I3-1205.pdf

[2] Hemalatha Reddy V., Kumari S., Muralidharan V., Gigoo K., Thakare B.S. (2020) Literature Survey—Food Recognition and Calorie Measurement Using Image Processing and Machine Learning Techniques. In: Kumar A., Mozar S. (eds) ICCCE 2019. Lecture Notes in Electrical Engineering, vol 570. Springer, Singapore

[3] I. Culjak, D. Abram, T. Pribanic, H. Dzapo and M. Cifrek, "A brief introduction to OpenCV," 2012 Proceedings of the 35th International Convention MIPRO, Opatija, 2012, pp. 1725-1730

[4] P. Pouladzadeh, S. Shirmohammadi and R. Al-Maghrabi, "Measuring Calorie and Nutrition From Food Image," in IEEE Transactions on Instrumentation and Measurement, vol. 63, no. 8, pp. 1947-1956, Aug. 2014, doi: 10.1109/TIM.2014.2303533.

[5] Hokuto Kagaya, Kiyoharu Aizawa, Makoto Ogawa ," Image-based Calorie Content Estimation for Dietary Assessment," "MM '14: Proceedings of the 22nd ACM international conference on Multimedia November 2014," 2018 25th IEEE International Conference on Image Processing (ICIP), Athens, 2018, pp. 171-175, doi: 10.1109/ICIP.2018.8451422.

[6] W. Zhang, D. Zhao, W. Gong, Z. Li, Q. Lu and S. Yang, "Food Image Recognition with Convolutional Neural Networks," 2015 IEEE 12th and Its Associated Workshops (UIC-ATC-ScalCom), Beijing, 2015, pp. 690-693, doi: 10.1109/UIC-ATCScalCom-CBDCom-IoP.2015.139.

[7] Ms. Ankita A. Podutwar , Prof. Pragati D. Pawar , Prof. Abhijeet V. Shinde,” A Food Recognition System for Calorie Measurement”, International Journal of Advanced Research in Computer and Communication Engineering ISO 3297:2007 Certified Vol. 6, Issue 1, January 2017, ISSN (Online) 2278-1021 ISSN (Print) 2319 5940.

[8] K Scott Mader, “Food Images (Food 101)’, Available: https://www.www.kaggle.com/kmader/food41

[9] Yamashita, R., Nishio, M., Do, R.K.G. et al. Convolutional neural networks: an overview and application in radiology. Insights Imaging 9, 611–629 (2018). https://doi.org/10.1007/s13244-018-0639-9

[10] S. Albawi, T. A. Mohammed and S. Al-Zawi, "Understanding of a convolutional neural network," 2017 International Conference on Engineering and Technology (ICET), Antalya, 2017, pp. 1-6, doi: 10.1109/ICEngTechnol.2017.8308186.