Journal of Analysis and Computation (JAC)
(An International Peer Reviewed Journal), www.ijaconline.com, ISSN 0973-2861
Volume XVI, Issue IV, Jan-June 2022

A STUDY ON FOOD RECOGNITION & NUTRITION ESTIMATION

Nupur Bhave[1], Dipti Belsare[2]

[1] *PG Student, Department of MCA, MES'S Institute of Management and Career Courses Pune, India*
[2] *Assistant Professor, MES'S Institute of Management and Career Courses Pune, India*

ABSTRACT

Nowadays, maintaining a balanced diet and avoiding obesity in the human body requires a regular & standardized intake of healthy food. This study uses computer vision via the use of a mobile phone and techniques of image processing and pattern recognition to identify food items in the picture and shows their nutritional values for dietary assessment. The amount of calories that the food contains is taken into consideration here based on that nutritional values are shown. Food image recognition is a difficult task due to the nature of the images, which is why existing approaches in image recognition have achieved low classification accuracy. Deep neural networks have overcome this problem and with help of this, we present an approach to the problem of food image detection, recognition, and nutrition estimation using convolutional neural network architecture.

Keywords - Deep learning, food image recognition, dietary assessment, nutrition-estimation, Machine Learning, Convolutional Neural Network

[1] INTRODUCTION

Recently, people are becoming used to modern lifestyles since they are busy with their schedules at work and home. Obesity in adults, as well as children, is becoming a common problem. The main cause of obesity is a combination of over food consumption and lack of physical activities. Currently, the research done shows that people who are obese are most likely to have serious health problems (hypertension, heart attack, diabetes, high cholesterol, cancers, and blood pressure). On the internet, consumers can find a wide range of nutritional information and standards at their fingertips. However, this information has not prevented diet-related issues or helped people to eat healthily. In most cases, people find it difficult to examine all of the information about nutrition and dietary plans. Moreover, people fail to care about measuring or controlling their daily food intake due to a lack of nutritional knowledge, irregular eating patterns, or lack of self-control. Providing people with an effective long-term solution requires mechanisms that help them make permanent changes to their calorie intake. A system that monitors, records, and measures the number of calories consumed in a meal would be of great help not only to patients in the treatment of obesity but also to the average calorie-conscious person or fitness freak.

Our goal is to empower users with a convenient, intelligent, and accurate system that helps them become aware of their calorie intake and find the individual nutrient content in the food item. To identify the food in the system, image processing, computer vision, feature extraction, neural networks, and segmentation will be used. It will measure the volume and weight of each food item and find the number of nutrients it contains like protein, iron, carbohydrates in the food and classify them.

[2] LITERATURE REVIEW

Recently, there has been an increasing number of researchers conducting experiments and research toward the fields of food classification, leveraging machine learning/deep learning algorithms.

Image recognition and interpretation are examples of high-level processing. Statistical or deep learning approaches are frequently employed to classify the target based on the application of interest in this step. The study's findings indicate that image processing is necessary. K-nearest Neighbor (KNN), Support Vector Machine (SVM), neural network, fuzzy logic, and a genetic algorithm are examples of algorithms that can aid in the interpretation of visual data. In the food sector, neural networks and fuzzy logic methods have been successfully applied to Machine Vision Systems.

[1]. Aizawa proposed a Bayesian framework-based approach to facilitate incremental learning for both food detection and food-balance estimation. [2]. Bossard used Random Forest on the Food-101 test set achieving a classification accuracy of 50.67% by mining discriminative components. The random forest model is used for clustering the superpixels of the training dataset. Other advanced classification techniques were also applied in the work including Improved Fisher Vectors (IFV), Bag-of-Words Histogram (BOW), Randomized Clustering Forests (RCF), and Mid-Level Discriminative Superpixels (MLDS). [3]. B.Deepak proposed a system of Food recognition that can detect and recognize food items based on the input image. His model is trained on 101 categories of food items using CNN (Convolutional Neural Network).

[4]. Krizhevsky used GPUs to train the AlexNet, which enabled faster training of CNNs models. The network consists of 5 convolutional layers and 3 fully connected layers. [5]. Karen Simonyan and Andrew Zisserman proposed the VGG-16 model which achieved 92.7% top-5 test accuracy in ImageNet. The 16 in VGG16 refers to it has a total of 16 layers that have weights. [6] Introduced by Shaoqing Ren, Kaiming He, Jian Sun, and Xiangyu Zhang, the ResNet (Residual Network) model have a 34-layer plain network in the architecture that is inspired by VGG-19. [7]. DensetNet developed by Gao Huang, Zhuang Liu has n(n+1)/2 connections in total because of feed-forward fashion. [8]. Shuffle Net is an extremely efficient CNN architecture with 173 deep layers, designed for mobile devices with the computing power of 10–150 Mega FLoating-point Operations Per Seconds (MFLOPs).

Additionally, researchers started to investigate which features and models are more suitable for food recognition and computed them into a food analysis system to calculate the calories. Multi-task convolutional neural networks are used for simultaneous learning of food calories, categories, and ingredients to automatically estimate the food calories from a meal image.

Though food recognition and nutrition contents analysis have been well discussed by the above work, three basic challenges remain. Firstly, most of the approaches are dealing with recognizing the food item with the help of a single picture. Secondly, Dataset becomes much larger when it comes to food images, so currently, we will be taking a finite dataset for training. Thirdly, similar types of images that are the same in terms of size and shape are difficult to recognize.

In this paper, I aim to address these issues and propose a system that identifies food from images and calculates nutritional values by increasing and improving the accuracy of the system. I aim to develop and train it on various images acquired from web searches for individual food. With this, I will try to achieve a higher classification and accuracy than most of the results presented and found.

[3] IMPLEMENTATION

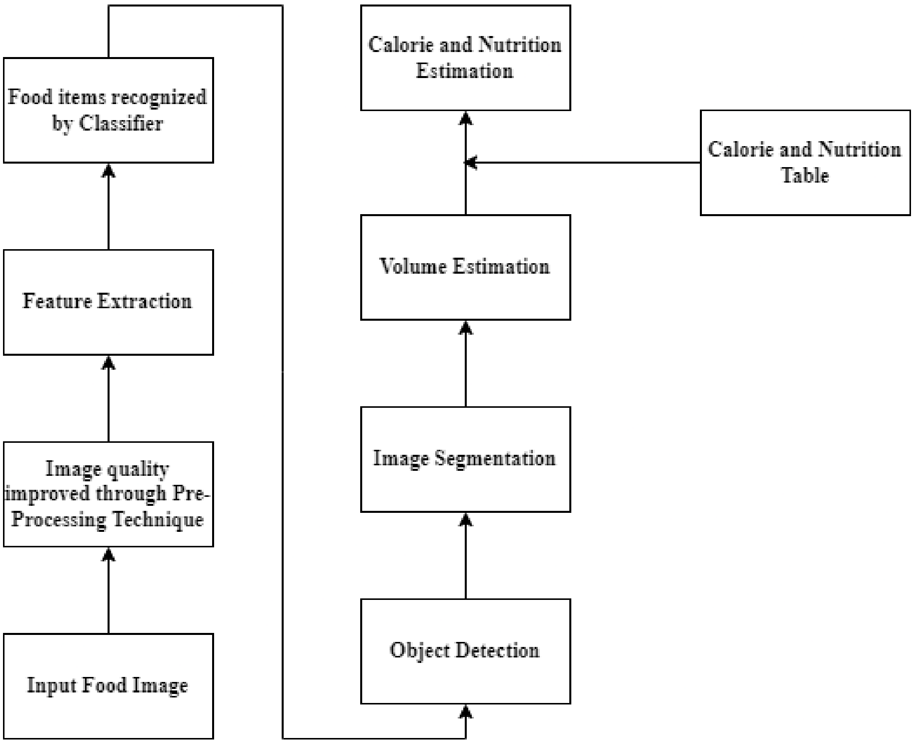

Fig.1

Steps to be followed are:

a. Pre-Processing - In Pre-processing, basically removing noise and normalizing the food image, if the image in any format needs to be converted in the specified format, resize in the specified size and remove unnecessary features from it. Pre-processing techniques include Histogram equalisation, filtering, and RGB picture to Grayscale conversion, among others.
b. Features Extraction - Feature Extraction is a crucial stage of the Face Recognition System (FRS) where the performance of recognition is dependent. It extracts the feature

vector, which is a meaningful set of data. Feature vector represents the properties of food.

c. Classification - Classification applies the feature vector on training and testing images. It is used for the result of the outcome in recognition. There are different classifiers like Support Vector Machine (SVM), Deep Learning Neural Network, Artificial Neural Network (ANN), and Convolutional Neural Network (CNN).
d. Object Detection - Calibration objects are detected by an object detection method called [9]. Faster R-CNN where we put an image with RGB channels as input and get a series of bounding boxes. For each bounding box created by Faster R-CNN, its class is judged.
e. Image Segmentation - After detection, we need to segment each bounding box. [10]. GrabCut is an image segmentation approach based on optimization by graph cuts. For each bounding box, we get a precise contour after applying the GrabCut algorithm. After segmentation, detecting the boundaries of irregular food portions becomes easier, and food portion detection improves. After that, we estimate each food's volume and calorie.
f. Volume Estimation - We need to calculate scale factors based on the calibration objects. For this, we can use 1 Yuan coin (basic monetary unit of the People's Republic of China) as a reference (diameter-2.50cm) and calculate the side view's scale factor, top view's scale factor and select the required volume estimation formula.
g. Calorie Estimation - After getting food volume, a food's calorie is calculated by searching its density in the food density table [11] and energy in the nutrition table.

[4] EXPERIMENTAL IMPLEMENTATION

Convolutional Neural Networks (CNN):

A deep learning neural network called a convolutional neural network (CNN) is a type of deep learning neural network. It's a machine learning method that takes an image as input, assigns significance to various aspects/objects in the image, and can distinguish between them. It works by extracting features from the images. CNN consists of the following:
1. The input layer is a grayscale image.
2. The Output layer which is a binary or multi-class labels
3. Hidden layers consist of convolution layers, ReLU (rectified linear unit) layers, the pooling layers, and a fully connected Neural Network.

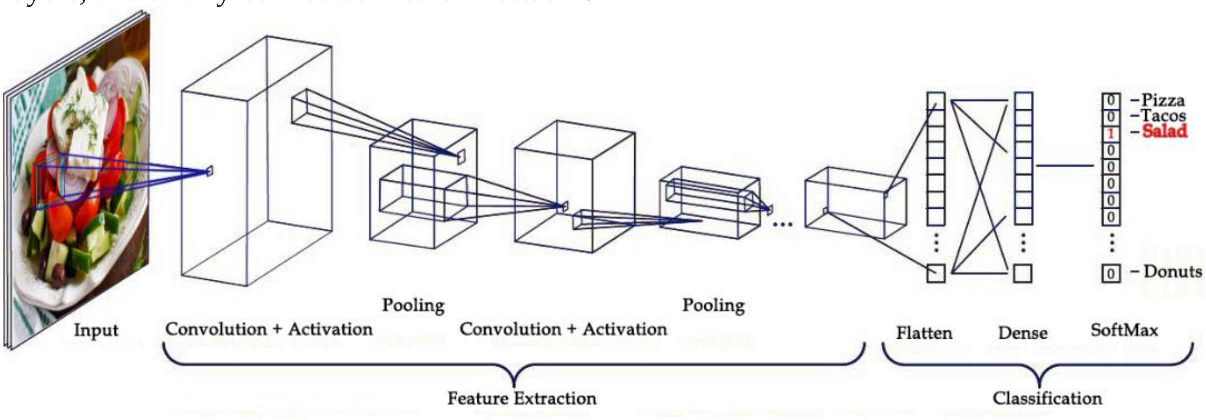

Fig.2 - Simple CNN Architecture [12]

TensorFlow & Keras:

In the past few years, various techniques have started to become available to the broader software development community. Industrial strength packages such as TensorFlow have given us the building blocks that Google uses to write deep learning applications for embedded/mobile devices to scalable clusters in the cloud without having to write code for GPU matrix operations, partial derivative gradients, and stochastic optimizers that make efficient applications possible. On top of all of this, there are user-friendly APIs such as Keras that abstract away some of the lower-level details and allow us to focus on rapidly prototyping a deep learning computation graph. We can mix and match to get the desired result. Keras and other Deep Learning libraries provide pre-trained models. There are deep neural networks with efficient architectures (like VGG, Inception, ResNet) that are already trained on datasets like ImageNet. Using these pre-trained models, we can already use the learned weights and add a few layers on top to finetune the model to our new data. This helps to save time and computation when compared to other models trained from scratch.

ReLU:

The rectified linear unit, or ReLU, is a method of increasing the non-linearity of a network without changing the receptive fields of convolution layers by using an activation function. ReLU enables faster data training, whereas Leaky ReLU can manage the problem of vanishing gradient. [13] Sigmoid function, Softmax function, Hyperbolic Tangent(tanh), Leaky Relu, Parameterized Relu, Exponential Linear Unit (ELU), Softplus, Softsign, Scaled Exponential Linear Unit (SELU), Linear Action function are some of the additional activation functions.

[5] DATA MANAGEMENT AND ANALYSIS METHODS:

Food Image Dataset:

For the experimental needs of the system, we will use a dataset named Food-101[14]. This dataset has 101000 images in total. It's a food dataset with 101 categories(multiclass). There are 750 training samples and 250 test samples for each food category. The training images were not cleaned on purpose, so there is still some noise in them. This comes mostly in the form of intense colours and sometimes wrong labels. All of the images were resized to a maximum of 512 pixels on each side. The entire dataset is 5GB in size.

Evaluation:

To evaluate the performance of the food recognition system, we visualize random images from each of the 101 classes and then split the image data into train and test using train.txt and test.txt. However, working on the entire data set with 101 classes takes a lot of time and processing to experiment and try other designs. To continue with my research, I'm going to make train min and test mini, which will limit the dataset to three classes.

Implementation:

The initial step is to import libraries like TensorFlow, NumPy, os, and pandas. Then, if you desire to adopt a GPU for training the deep learning model, please install CUDA with version

11.0 because that version holds TensorFlow with version 2.4.1. After that, we need to prepare data by producing a dataframe with columns and set parameters for the image augmentations technique. After we've created the batches, we can use the transfer learning technique to train the model because we don't need to build the CNN architecture from scratch. We will apply [15] VGG-16, ResNet50, Inceptionv3, EfficientNet architectures to our model. By using these 4 pre-trained models we will compare their results based on the accuracy of the test achieved.

REFERENCES

[1]. K. Aizawa, Y. Maruyama, H. Li, and C. Morikawa, ‘‘Food balance estimation by using personal dietary tendencies in a multimedia food log,’’ IEEE Trans. Multimedia, vol. 15, no. 8, pp. 2176–2185, Dec. 2013.

[2]. L. Bossard, M. Guillaumin, and L.Van Gool, “Food-101–mining discriminative components with random forests,” in Computer Vision– ECCV 2014. Springer, 2014, pp. 446–461

[3]. S.Suma, M.Bharathi, G.Harikumar, B.Deepak, “Food Recognition And Calorie Estimation Using Image Processing,” Vol-6 Issue-1 2020

[4] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in Advances in neural information processing systems, pp. 1097–1105, 2012.

[5] Simonyan, Karen & Zisserman, Andrew. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv 1409.1556.

[6]. Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun: Deep Residual Learning for Image Recognition, Dec 2015

[7] G. Huang, Z. Liu, L. Van Der Maaten and K. Q. Weinberger, “Densely Connected Convolutional Networks,” 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017, pp. 2261–2269, DOI: 10.1109/CVPR.2017.243.

[8] X. Zhang, X. Zhou, M. Lin, and J. Sun, “ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices,” 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, 2018, pp. 6848–6856, DOI: 10.1109/CVPR.2018.00716.

[9] Girshick, Ross. "Fast R-CNN." Proceedings of the IEEE international conference on computer vision. 2015.

[10] C. Rother, V. Kolmogorov, A. Blake, Grabcut: Interactive foreground extraction using iterated graph cuts, in ACM transactions on Graphics (TOG), Vol. 23, ACM, 2004, pp. 309–314.

[11] U. Ruth Charrondiere, David Haytowitz, and Barbara Stadlmayr, “FAO/INFOODS Density Database Version 2.0 (2012)”, E-ISBN 978-92-5-107346-9.

[12] Chairi Kiourt, George Pavlidis, Stella Markantonatou, “Deep learning approaches in food recognition”, July 2020, pp. 83-108, DOI:10.1007/978-3-030-49724-8_4.

[13] François Chollet. “Layer activation functions.”, Keras.io, Retrieved 2016-09-18, https://keras.io/api/layers/activations/#layer-activation-functions .

[14] Bossard L., Guillaumin M., Van Gool L. (2014) Food-101 – Mining Discriminative Components with Random Forests. In: Fleet D., Pajdla T., Schiele B., Tuytelaars T. (eds) Computer Vision – ECCV 2014. ECCV 2014. Lecture Notes in Computer Science, vol 8694. Springer, Cham. https://doi.org/10.1007/978-3-319-10599-4_29.

[15] François Chollet, “Keras Applications.”, Keras.io, Retrieved 2016-09-18, https://keras.io/api/applications/