International Journal of Computer Sciences and Engineering Open Access
Review Paper Vol. 9, Issue.5, May 2021 E-ISSN: 2347-2693

Food Image Classification Using Machine Learning Techniques: A Review

Yash Baid[1*], Avinash Dhole[2]

[1, 2]Department of Computer Science & Engineering, Raipur Institute of Engineering and Technology, Raipur, India

Corresponding Author: yashbaid4500@gmail.com, Tel.: +91-8962817926

DOI: https://doi.org/10.26438/ijcse/v9i5.3136 | Available online at: www.ijcseonline.org

Received: 07/May/2021, Accepted: 17/May/2021, Published: 31/May/2021

***Abstract*—**The recognition of image is one of the most important fields in the image processing and computer vision. Image recognition has many branches but the food image classification is very unique. In today's world people are very conscious about their health. Many people around the world use some dietary assessment system for planning of their diet. In dietary assessment system people make the use of food image classification to classify the food from the image and provide the total amount of calories present in the food. The classification of food images is a very difficult task as the dataset of food images is highly non-linear. In this paper, we are going to use different types of neural network models to show, which neural network provides the best accuracy result in the recognition of food images and is most efficient to use. We are using a food image dataset (food-11) which contains 16643 images in it.

***Keywords*—** Deep Learning, CNN, RNN, Computer Vision, Image processing, DCNN

I. INTRODUCTION

In the today's world obesity and other health problem has been increasing day by day. In the paper [2], it is specify that the obesity has been doubled since 1980 in more than 70 countries. Obesity can cause many types of chronic diseases such as heart, diabetes, arthritis, etc and also decrease the immunity of the human system. Peoples have also shown their interest on reducing weight by calculating the calorie values of their food intake. Nutritional value of food should get more importance to prevent from diseases. To regularize the food habits of people they can make the use of the dietary management. Dietary management will help people by telling the information about the food they are eating. To get the calorie information, a system needs to detect the food from image and then analyze the dietary. For this paper the researchers is mainly concerned with the detection of food from the images.

Image processing and computer vision techniques are now the base of many domains. Food recognition from images is one of the domains of image processing. The recognition of food is a very difficult task as there is large similarity between food classes. Image processing also required more computation power than many text base data classification. The food recognition model should be very efficient, so people can run on a less expensive device too.

In today's world there are affordable Smartphone that have high computational power that can be used to process the high quality image data. After the comparison the model with high accuracy and speed can be implement the Smartphone for wider use.

In the next section of the paper we will elaborate some literature, in section III we will discuss about how we have proposed the method, in section IV we will discuss how we motivated towards this research and problem in food image classification and also we are going to give experimental evaluation of different existing image classification techniques, at last we will conclude our study

II. LITERATURE SURVEY

The authors proposed a system of food image recognition using Convolutional neural network. The authors use a food-11 dataset for the training and testing the neural network. The authors also use inception V3 model that is per-trained with the ImageNet. The image is also pre-processed before passing to the training. In processing phase the authors use the ZCA whitening to reduce the redundancy in the matrix of pixel images. They also changed the size of all the images to 299 x 299 x 3 to increase the processing time and also to fit in Inception V3 [1].

The authors discussed about the health effects of overweight and obesity in 195 countries that study over 25 years from 1980 to 2015. The authors have analyzed data from 68.5 million persons. In 2015, the obesity count on adult is 603.7 million & 107.7 million. Since the start of analysis in 1980, the obesity has doubled in more than 70 countries and has continuously increased in most other countries. 4.0 million Deaths globally has caused by High BMI. More than two thirds of deaths related to high BMI were due to cardiovascular [2].

The authors proposed an image analysis system for the automatic identify and quantify foods and beverages consumed at an eating occasion from images of foods and beverages captured using a mobile device. They used an approach that is based on the k-nearest neighbours and vocabulary tree. They have searched for the color, texture, and local region features for the classification of food. The color and texture is classified using the k-nearest neighbours. They used an image dataset that consist of 1453 images of eating occasions in 42 unique food categories [3].

The authors proposed a method that uses the local textural patterns and their global structure for the classification of food image. The authors used a Scale Invariant Feature Transformation (SIFT) interest point detector with the Local Binary Pattern (LBP) feature for creating a visual codebook of local textural pattern. Local texture is used to represent food image, food object is represented as the spatial distribution of the local textural structures and encoded using shape context. The authors evaluate the proposed method on the Pittsburgh fast-food image (PFI) dataset [4].

The authors proposed a method for the classification of food images using the sphere shaped support vector machine. The authors used the FCM (Fuzzy C-Means) algorithm for the segmentation of food images and sphere shaped SVM (Support Vector Machine) is used for the classification of segmented food idem. The proposed method automatically identifies the food items and also calculates their calorie value after identification [5].

The authors proposed a novel method for the classification of food images using Random Forests (RF). Random Forests allow us to mine for parts simultaneously for all classes and to share knowledge among them. For the improvement in efficiency of mining and classification, the authors are only considering the patches that are aligned with the image super pixels. The authors also created a novel and challenging dataset for the method, the dataset consist 101 food categories that have 101,100 images in it. For the labelling of food items the authors uses either the nutrition experts or Amazon Mechanical Turk [6].

The authors proposed a system for the improvement of accuracy in the dietary assessment. It can be done by analyzing the food images captured using the Smartphone. The main technique innovation of the authors is the use of deep learning-based food image recognition algorithm. It uses the Convolutional neural network based food image recognition algorithm to address the problem. The proposed system is test on the 2 real worlds i.e. UEC-256 & Food- 101 [7].

The authors proposed a system with deep Convolutional neural network (DCNN) for the recognition of food from the food photo/image. In this system the authors uses a dataset that is already fine tuned and pre-trained for training. ImageNet dataset are used because it contains 1000 food-related categories. The author also uses the food classifier in the Twitter photo data. The author also conclude that the DCNN was very suitable for large-scale image data, since it takes only 0.03 seconds to classify one food photo with GPU [8].

The authors proposed a prediction model for the classification of food images. Here the author only focuses on the Thai Fast Food images. The model used by author is based on deep learning process that is trained with natural images (GoogLeNet Dataset) and then fine-tuned to generate the predictive Thai fast food model. The authors also created a Thai Fast Food (TFF) dataset which is consist of 11 groups and those 11 groups consist of total 3960 images [9].

The authors proposed a system that is based on the Convolutional neural network. The author also compared the proposed system with the system created using deep Convolutional neural network. The author also creates a new dataset with the combination of following dataset: Food-11, FooDD, Food100 and web archives [10].

The authors proposed a method of that is used to increase the network in ways that aim at utilizing the added computation as efficiently as possible by suitably factorized convolution and aggressive regularization. The authors benchmark the system with the ILSVRC 2012 classification challenge validation set [11].

The author introduced a new database for the image recognition. The author named the database as ImageNet, this database built upon the backbone of the WordNet structure. The authors also include the analysis of the new dataset (ImageNet) in its current state: 12 sub-trees with 5247 synsets and 3.2 million images in total, the author also says that the ImageNet is much larger in scale and diversity and more accurate than the current image dataset. The authors also make the use of the Amazon Mechanical Turk for the data collection scheme. The authors also show the usefulness of ImageNet with a simple object recognition program [12].

The author proposed a method that uses the Multi-column Deep Neural Network, in this approach small (often minimal) receptive fields of Convolutional winner take all neurons yield large network depth, resulting in roughly as many sparsely connected neural layers as found in mammals between retina and visual cortex, only winner neurons are trained. Using this method several deep neural columns become experts on inputs pre-processed in different ways: predictions are averaged. It also uses graphics cards which allow the fast training of models [13].

The author proposed a mobile application that analysis and tells the calories based on the food image captured using mobile device. The app is also available for the use. The author uses a food dairy for storing the food image. The food dairy will be filled by the data entered by the user in the mobile app; here every user will have a separate food dairy [14].

The authors create a web based application for the identification of food images. The application can extract food images from other images, it also analyze the food balance, and visualize the log. The authors make the use of Food pyramid to estimate the balance of food. It is very simple and makes logging very feasible. The authors extracted the food images from the flickr. The authors also implemented an interface for correction of some value, as system not always give the correct estimation [15].

The authors create a mobile application that will be used for the accounting of daily food and nutrient intake. In creating this the authors use the image analysis tools for the identification and quantification of food that is consumed at a meal, in this application we capture the image two times first before consuming the food and second after consuming the food to estimate the amount and type of food consumed. Using two images help in determine the near to exact amount of food. The authors also convert the RGB image to YCbCr color space. The authors also make the use of the support vector machine for the identification of food items using statistical pattern recognition technique [16].

The authors present a Pittsburgh Fast-food Image Dataset (PFID), a collection of visual data to facilitate research in automated food recognition. PFID contains data of 101 fast food categories from 11 popular fast food chains, and capturing images and videos in both restaurant and a controlled lab setting. The authors benchmark the dataset using two standard approaches, color histogram and bag of SIFT [17].

The authors present a system that can recognized the foods from videos, the videos is directly recorded in the restaurant using web camera. After the recognition result, the system also estimates and tells the food calories of intake. The authors perform evaluation of the system on a dataset of 101 foods from 9 food restaurants in USA [18].

The authors present a calorie and nutrition measurement system that helps the patient and dieticians in measuring and managing the daily food intake. The detection of food is done using feature extraction [19].

The authors focus on Thai food image classification. The author's method creates feature vector using many features about texture and color, then recognize the food using SVM. The training of the system is done by grouping the food images by it type and the amount of calories [20].

The authors present mobile food recognition system. The system will estimate the calorie and nutritious of foods and also records the eating habit. This system is not needed to send any information to the server as image recognition program is already built into it. In this system the user needs to draw a boundary wall by touching on the screen this helps in identifying only that area that is needed to be scan, this reduces the overhead of scanning all the area in search of food [21].

The authors present a system that is based on a large deep Convolutional neural network to classify the 1.2 million high resolution images in the ImageNet LSVRC-2010. The authors also make the use of the GPU to increase the training and evolution speed of the system. The authors also develop a regularization method called "dropout" that is used to reduce the over fitting in the fully connected layers. The Convolutional neural network is consist of fewer connections and parameters then the standard feed forward neural network, so they are easy to train, while their theoretically-best performance is likely to be only slightly worse [22].

III. METHODOLOGY

In this approach, we are going to use different types of neural network. A neural network is a network or circuit of neurons, or in a modern sense, an artificial neural network made up of artificial neurons or nodes. Thus a neural network is either biological neural network or artificial neural network. The biological neural network is composed of real biological neurons, while the artificial neural network made up of artificial neurons that are used to solve the artificial intelligence problem. The main use of neural network is in the field of machine learning and artificial intelligence.

There are many types of neural networks but we are only focusing on the neural network that is mostly used in image processing, they are:-

1) **Modular Neural Network: -** In this type of neural network, many independent networks contribute to the result collectively. Many subtask performed and constructed by each of these neural network. It provides the set of input that are unique, when compared with other neural network. The complexity of a problem is easily reduced while solving problem by these modular network because they completely break down the sizeable computational process into small components. The total time of processing will depend on the involvement of neurons in the computation of results. Modular Neural Network is one of the fastest-growing areas of artificial intelligence.

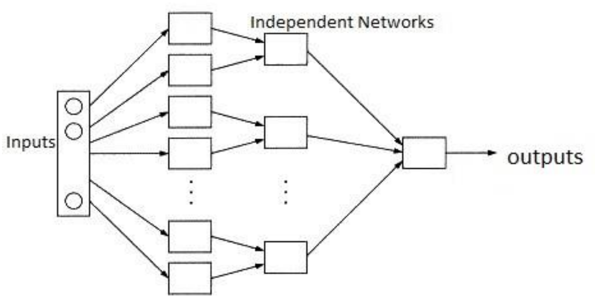

Figure 1:- Modular Neural Network

2) **Feed Forward Neural Network: -** In feed forward neural network, the information travels only in one direction and is the purest form of an artificial neural network. This type of neural network contains hidden layers and data enter through input nodes and exit through output nodes. Classify activation function is used in this neural network. There is no back propagation, and only the front propagation wave is allowed. The feed forward neural network can be used in speech recognition and computer vision. It is very easy to maintain this type of neural network.

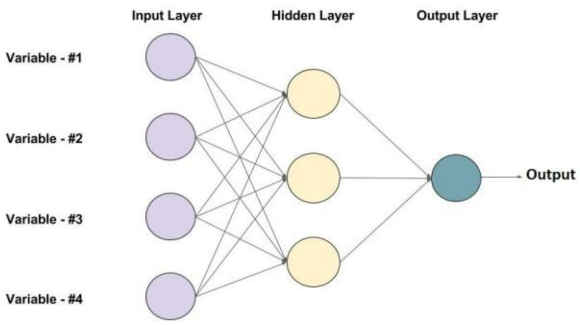

Figure 2:- Feed Forward Neural Network

3) **Recurrent Neural Network: -** The principle of the RNN is to feedback the output of the layer back to input again. This principle helps to predict the outcome of the layer. In the computation process, each neuron will act as a memory cell. The neuron will retain some information as it's go to the next step. The data to be used later will be remembered and work for the next step will go on the process. The prediction will be improved by error correction. Some changes are made to create right predicate output. The rate of how fast the network can make the correct prediction from the wrong prediction is known as learning rate. The main application of Recurrent Neural Network is to convert the text to speech. The recurrent neural network was designed for supervised learning.

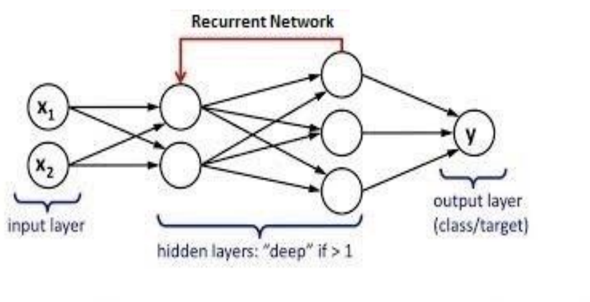

Figure 3:- Recurrent Neural Network.

4) **Convolution Neural Network: -** In this type of neural network, learn-able biases and weights are given to the neurons initially. CNN was designed to map image data to an output variable. They are proven so effective that they are the go-to method for any type of prediction problem involving image data as an input. CNN works well with data that has a spatial relation. The images are remembered in parts to help the network in computing operation. The photos are recognized by taking the input feature batch-wise. In the computing process, image is converted to Grayscale from HSI to RGB scale. The classification of images done into various categories after the image is transformed. Convolutional Neural Networks have a very high level of accuracy. That is also the reason why convolution neural networks are dominating the computer vision technique. It is mainly used in Image data application and classification problem.

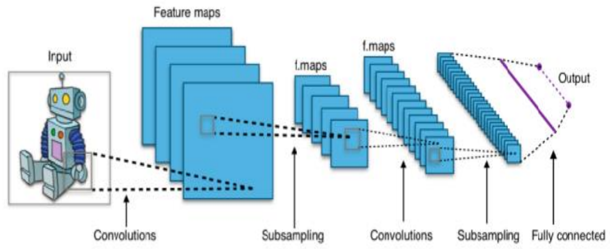

Figure 4:- Convolutional Neural Network [1]

5) **Multilayer Perceptrons: -** It is a classical type of neural network. It is comprised of one or more layers of neurons. Data is fed to the input layer, there may one or more hidden layers providing the level of abstraction, and prediction are made on the output layer also called the visible layer. MLP are suitable for classification prediction problem where inputs are assigned a class or label. Data is often provided in a tabular format, such as you would see in a CSV or excel.

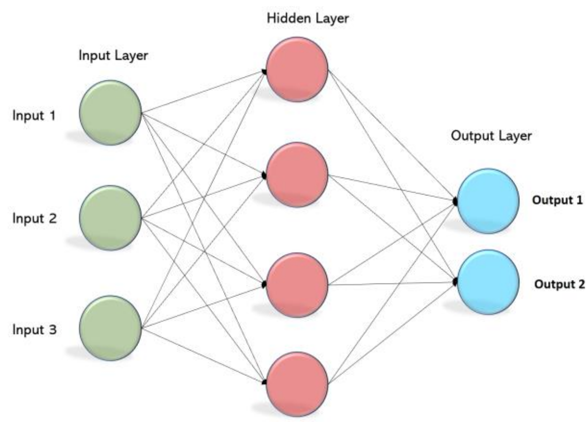

Figure 5:- Multilayer Perceptrons

IV. RESULTS AND DISCUSSION

There are some problems in food image recognition:-

1. The exhibited methodologies have overlooked any other neural network for the recognition of food images.
2. In most of the exhibited methodologies they haven't tried to implement the method on the devices.

For discussing we have analysis the results different exhibited methods. The analysis is:-

The authors obtained an accuracy of 92.86% using inception V3 model, while compare to the other**s** Fine Tuned methods like Alexnet [10] has an accuracy of 82.23% and the accuracy of Caffenet [10] has 80.12%. The transfer learning uses the knowledge earned from previous learning in new dataset to classify images that is why the approach used by the authors has better accuracy than all the others [1].

The author used two food classification baseline models those are: - Baseline 1: Color Histogram + SVM Classifier & Baseline 2: Bag of SIFT Features + SVM Classifier. The result of the base line experiments, one can simply see the SIFT often outperforms the color histogram baseline even in the case of salads. The author separate the training and test data samples so that no instance of a food item appears in both training and test. The accuracy of proposed method was 78% in the case of salad category, which is 12% lower than the baseline 2 but 12% higher than the baseline 1, for meat category has an accuracy of 52% while baseline 2 has 56% and baseline 1 has 47%. With the remaining food categories (hamburger, miscellaneous, sandwich, subs, wraps) has outperform both baseline 1 and baseline 2 [4].

The author uses evaluation metrics such as segmentation, accuracy, true and false positive rate and classification rate. The accuracy is an important assess for classification purpose, the better classification attains when the classifier gives higher accuracy rate. The true positive rate is the measure of positive proportion which is correctly identified by the classifier. Classification rate is one of the most important concerns for the dietary assessment system. The proposed method has an accuracy of 95% which provide better classification performance [5]. The evaluation function of CR is:-

$$CR = \frac{TP + TN}{TP + TN + FP + FN}$$

The result obtained is 78.77% as Top-I accuracy in UEC-FOOD100 and 67.57% for UEC-FOOD256, both of which were the best result so far. The author also applies the food classifier with the combination of DCNN and the twitter photo data. The DCNN was suitable only for the large-scale image data, since it takes only 0.03 seconds to classify one photo in a GPU. The UEC-FOOD100 and UEC-FOOD256 has been a Japan dataset [8].

The author performed the experiment on a notebook computer which has i7-4700, 8GB of memory and Windows 8.1 enterprise as the operating system. The author uses the 3,300 images as a training dataset, in which 300 image per group. For the testing purpose the author reserved 660 images, which is divided into 60 images per group. Using this author is able to achieve an accuracy of 88.33% on overall dataset. The author has achieved 100% accuracy in the barbecued red pork in sauce with rice group. The prediction model can classified 93.33% accuracy with an omelette on rice, 70% with a rice topped with stir-fried chicken and basil, 93.33% accuracy with barbecued red pork in sauce with rice, 93.33% accuracy with a stewed pork leg on rice, 83.33% accuracy with a Thai fried noodle, 88.33% accuracy with a rice with curried chicken, 83.33% accuracy with a steamed chicken with rice, 93.33% accuracy with a shrimp-paste fried rice, 76.67% accuracy fried noodle with pork in soy sauce and vegetables, and 90% accuracy with a wide rice noodles with vegetables and meat [9].

The authors get the accuracy of 70.12%. The authors run the proposed system on a collected dataset with caffenet framework which can run on a GPU. The authors need to run the test 5 times to remove the random effect since the weight and bias values generated normally. The author also shows the learning pattern, it is learning rate is 0.3, Momentum is 0.7, and Maximum iteration is 40000 on a Food-11 dataset. The proposed system of author is compared with 2 other system Alexnet & caffenet, the Alexnet get accuracy of 80.51, 82.07 respectively which is much higher than the authors proposed model. It is clear from the result that the fine-tuned data provide better result [10].

V. CONCLUSION

In this paper, we have reviewed some of image recognition technique that is based on deep learning & Convolutional neural networks. After studying all the research paper in the reference, we are able to identify some enhancements that can be done to increase the accuracy of the model. In this survey we found that most of the authors uses Convolutional neural network as there base neural network for the recognition of food images, they don't consider any other neural network. The most efficient method of food identification can be implementing in a real device system. In this paper we can conclude that other neural network models should be consider to get the higher accuracy in the identification of food images.

REFERENCES

[1] Md Tohidul Islam, Sagidur Rahman, B.M. Nafiz Karim Siddique, Taskeed Jabid, "*Image Recognition with Deep Learning*," 2018 International Conference on Intelligent Informatics and Biomedical, Bangkok, Thailand**, pp. 106-110, 2018.**

[2] "*Health effects of overweight and obesity in 195 countries over 25 years*,", New England Journal of Medicine 377, **no. 1, 13–27, PMID: 28604169, 2017**.

[3] Y. He, C. Xu, N. Khanna, C. J. Boushey and E. J. Delp, "*Analysis of food images: Features and classification*," IEEE International Conference on Image Processing (ICIP), Paris, **pp. 2744-2748, 2014**.

[4] Z. Zong, D. T. Nguyen, P. Ogunbona and W. Li, "*On the Combination of Local Texture and Global Structure for Food Classification*," IEEE International Symposium on Multimedia, Taichung, **pp. 204- 211, 2010**.

[5] S. J. Minija and W. R. S. Emmanuel, "*Food image classification using sphere shaped — Support vector machine*," International Conference on Inventive Computing and Informatics (ICICI), Coimbatore , **pp. 109-113, 2017**.

[6] Lukas Bossard, Matthieu Guillaumin, and Luc Van Gool, *"Food-101 – mining discriminative components with random forests,"* Computer Vision – ECCV (Cham) (David Fleet,

Tomas Pajdla, Bernt Schiele, and Tinne Tuytelaars, eds.), Springer International Publishing, **pp. 446–461, 2014**.

[7] Chang Liu, Yu Cao, Yan Luo, Guanling Chen, Vinod Vokkarane, and Yunsheng Ma, "*Deepfood: Deep learning-based food image recognition for computer-aided dietary assessment,*" **CoRR abs/1606.05675, 2016**.

[8] K. Yanai and Y. Kawano, "*Food image recognition using deep convolutional network with pre-training and fine-tuning,*" IEEE International Conference on Multimedia & Expo Workshops (ICMEW), Turin, **pp. 1-6, 2015**.

[9] N. Hnoohom and S. Yuenyong, "*Thai fast food image classification using deep learning,*" International ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI-NCON), Chiang Rai, **pp. 116-119, 2018**.

[10] G. Özsert Yiğit and B. M. Özyildirim, "*Comparison of convolutional neural network models for food image classification*," IEEE International Conference on INnovations in Intelligent SysTems and Applications (INISTA), Gdynia, **pp. 349-353, 2017**.

[11] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, and Zbigniew Wojna, *"Rethinking the inception architecture for computer vision,"* **CoRR abs/1512.00567 ,2015**.

[12] J. Deng, W. Dong, R. Socher, L. J. Li, Kai Li and Li Fei-Fei, "*ImageNet: A large-scale hierarchical image database,*" IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, **pp. 248- 255**, **2009**.

[13] D. Ciregan, U. Meier and J. Schmidhuber, "*Multi-column deep neural networks for image classification*," IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, **pp. 3642-3649, 2012**.

[14] Kiyoharu Aizawa, Makoto Ogawa, "*FoodLog: Smartphone based Multimedia Food Recording Tool*" International Conference on Artificial Reality and Telexistence (ICAT), Tokyo, Japan, **pp.143-144, 2013**.

[15] Keigo Kitamura, Toshihiko Yamasaki, Kiyoharu Aizawa, "*FoodLog: Capture, Analysis and Retrieval of Personal Food Images via Web*" CEA, Beijing, China, **pp. 23-29, 2009**.

[16] Fengqing Zhu, Marc Bosch, Insoo Woo, SungYe Kim, Carol J. Boushey, David S. Ebert & Edward J. Del, "*The Use of Mobile Devices in Aiding Dietary Assessment & Evaluation*", IEEE Journal of Selected Topics In Signal Processing, **pp. 756-766, 2010**.

[17] Mei Chen, Kapil Dhingra, Wen Wu, Lei Yang, Rahul Sukthankar, Jie Yang, "*PFID: PITTSBURGH FAST-FOOD IMAGE DATASET*", ICIP, **pp. 289-292, 2009**.

[18] Wen Wu, Jie Yang, "*FAST FOOD RECOGNITION FROM VIDEOS OF EATING FOR CALORIE ESTIMATION*", ICME, **pp.1210-1213, 2009**.

[19] Parisa Pouladzadeh, Shervin Shirmohammadi, Rana Al-Maghrabi, "*Measuring Calorie and Nutrition from Food Image*", IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT, **pp. 1-10, 2014**.

[20] Natta Tammachat, Natapon Pantuwong "*Calories Analysis of Food Intake Using Image Recognition,*" International Conference on Information Technology and Electrical Engineering (ICITEE), Yogyakarta, Indonesia, **2014**

[21] Yoshiyuki Kawano and Keiji Yanai, "*Real-time Mobile Food Recognition System*", IEEE Conference on Computer Vision and Pattern Recognition Workshops, **pp. 1-7, 2013**.

[22] Alex Krizhevsky, Iiya Sutskever, Geoffrey E. Hinton, "*ImageNet Classification with Deep Convolutional Neural Networks*", ImageNet LSVRC-2010, **pp.1-9, 2010**.

