IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 2, 2018, accepted October 17, 2018, date of current version January 7, 2019.

Digital Object Identifier 10.1109/ACCESS.2018.2879117

A Framework to Estimate the Nutritional Value of Food in Real Time Using Deep Learning Techniques

RAZA YUNUS[1], OMAR ARIF[1], HAMMAD AFZAL[2], MUHAMMAD FAISAL AMJAD[2], HAIDER ABBAS[2], HIRA NOOR BOKHARI[1], SYEDA TAZEEN HAIDER[1], NAUMAN ZAFAR[1], AND RAHEEL NAWAZ[3]

[1]School of Electrical Engineering and Computer Science, National University of Sciences and Technology, Islamabad 44000, Pakistan
[2]College of Signals, National University of Sciences and Technology, Islamabad 44000, Pakistan
[3]Department of Operations, Technology, Events and Hospitality Management, Manchester Metropolitan University, Manchester M15 6BH, U.K.

Corresponding author: Hammad Afzal (hammad.afzal@mcs.edu.pk)

ABSTRACT There has been a rapid increase in dietary ailments during the last few decades, caused by unhealthy food routine. Mobile-based dietary assessment systems that can record real-time images of the meal and analyze it for nutritional content can be very handy and improve the dietary habits and, therefore, result in a healthy life. This paper proposes a novel system to automatically estimate food attributes such as ingredients and nutritional value by classifying the input image of food. Our method employs different deep learning models for accurate food identification. In addition to image analysis, attributes and ingredients are estimated by extracting semantically related words from a huge corpus of text, collected over the Internet. We performed experiments with a dataset comprising 100 classes, averaging 1000 images for each class to acquire top 1 classification rate of up to 85%. An extension of a benchmark dataset Food-101 is also created to include sub-continental foods. Results show that our proposed system is equally efficient on the basic Food-101 dataset and its extension for sub-continental foods. The proposed system is implemented as a mobile app that has its application in the healthcare sector.

INDEX TERMS Food recognition, convolutional neural networks, vector embeddings, attribute estimation.

I. INTRODUCTION

High Calorie food intake can be harmful and result in obesity, which is a preventable medical condition that causes abnormal accumulation of fat in the body. It can result in numerous diseases such as obesity, diabetes, cholesterol, heart attacks, blood pressure, breast, colon and prostate cancers [1] and other diet-related ailments. In order to deal with such problems, people are inclined towards making a difference in their diet plans by paying more attention to what type of food they are consuming. Diet management is a key concern amongst individuals belonging to different age groups. However, one major challenge in diet management is to maintain a balance between what one eats and how one monitors his/her food consumption. The immense increase in ailments such as high cholesterol, blood pressure, strokes etc. demand for nutritional and diet management for which people resort to expensive nutrition therapies. It is a known fact that energy balance plays a pivotal role in maintaining a healthy weight and lifestyle. If people become more aware about their food intake and its nutritional value, then the diseases mentioned above and allergies can be reduced. This work aims to develop a mobile application that can record real time images of meal and analyze it for nutritional content, so that people can improve their dietary habits and lead a healthy life.

Most of the existing systems, often implemented as smart phone applications (e.g. MyFitnessPal [2], SHealth [3]) help users to keep track of their food intake. These systems assist users in achieving dietary goals such as weight gain/loss, allergy management or maintaining a healthy diet. However, they require users to manually input the food details along with the portion sizes. This can be very tedious and time consuming, resulting in users to refraining from using these applications for long periods of time. Furthermore, naive users rely on self reports of calorie intakes which often are misleading. Similarly, [4] relies on expert nutritionists to analyze images everyday. There are approaches that use mobile phone cameras to automatically recognize the food [5]–[12]. However, the task of food attributes measurement is not

Attributes		Ingredients		Nutrition	
Grill	1	Beef	1	Fat	1
Spicy	0.75	Cheese	0.9	Sodium	0.55
Juicy	0.64	Lettuce	0.8	Protein	0.44
Fry	0.54	Tomato	0.67	Cholesterol	0.24
Tasty	0.53	Potato	0.58	Carbs	0.7
Healthy	0.33	Soya	0.53	Zinc	0.3
		Egg	0.52	Manganese	0.15
		Onion	0.50		

(a)

Attributes		Ingredients		Nutrition	
Fry	1	Rice	1	Carbs	1
Spicy	0.69	Chicken	0.73	Fat	0.8
Grill	0.55	Saffron	0.63	Starch	0.49
Healthy	0.5	Chilli	0.57	Fiber	0.38
Hot	0.43	Ghee	0.55	Sodium	0.36
Healthy	0.33	Soya	0.4	Manganese	0.35
Juicy	0.39	Tomato	0.3	Zinc	0.23
Sour	0.21				

(b)

FIGURE 1. Different food attributes estimated by the proposed algorithm. Food ingredients and nutritional content are also estimated by the algorithm. (a) Burger. (b) Biryani.

carried out. In other methods [13]–[15], recognition of food is followed by food volume estimation [11] and then calorie computation. However, the volume measurement procedure is very tedious and prone to errors. Secondly, such algorithms are applied to simple food items. Crowd sourcing is also employed for the nutritional analysis of food items [16] which makes the algorithm costly and inhibits it from widespread application in daily life.

The proposed system aims to be a step towards creating awareness based on health and fitness concerns so that people can eat and live in a better way. The proposed method helps in determining the nutritional content of food automatically by making it feasible for a person to learn about what food might contain and how healthy it might be. The inherent theme is to automatically detect food items from an image of a platter and then estimate the respective food attributes such as the percentage of calcium, iron etc. along with the ingredients present in the food. Our system provides nutrition facts similar to packaged food items. The proposed system has its application in health-care industries and hospitals. Knowing about the nutritional value will further provide motivation for patients to refrain from food that can be detrimental to their health.

In visual object recognition tasks, Convolutional Neural Networks (CNN) [17], [18] have found great success and therefore CNNs are also employed for recognizing food items present in an image [9], [10], [12]. In this work, we employ CNNs to acquire top 1 recognition accuracy rates of 85%. Another challenge is in the accurate computation of ingredients and nutritional value of the food [19]. Our goal is to minimize the user input and automate this task as much as possible. We employ deep neural networks for estimating the ingredients and the attributes of the food. Our focus not only revolves around attributes like protein, calcium or vitamins etc. but also includes ingredients present in the food items. Our system automatically estimates food attributes, food ingredients and nutritional content.

An output of the proposed system is implemented as a mobile app ''Rate Your Plate'', where the mobile phone takes an image of food, recognizes it and displays the ingredients and attributes automatically. Figure 1 shows food attributes automatically estimated by the proposed algorithm.

The proposed system consists of two components. The first component uses CNNs to recognize the food item in an image (Section III-B). The second component estimates food attributes using text retrieval from internet archives as well as scrapping of data from nutritional and recipe websites for ingredients and nutrient counts. This data is trained on a two layer neural network, from which we can compute probabilities of existing ingredients in a particular food item. Each food item typically has a standard serving size against which calories and nutritional content can be calculated. The system uses deep learning algorithms, a server with a trained model to recognize food images and estimate its attributes along

with ingredients, and a conventional mobile application. The classifier requires a large dataset containing multiple images against every category of food item for training purposes. This required assistance from publicly available datasets such as Food-101 dataset [7] and Image-net. We augmented the Food-101 with subcontinental cuisine since we did not find any dataset that included subcontinental dishes.

The major contributions of the presented work are listed below.

- A food recognition engine that is trained using Convolutional Neural Networks.
- An extension of the Food-101 dataset that covers the subcontinental cuisine, involving well-defined training and validation classes.
- A real time food attribute estimation using vector space embedding [20]. This module is trained on data scrapped from internet archives of various nutritional and recipe websites.

The rest of the paper is organized as follows: In Section II, we give an overview of the related work. The details of our methodology are presented in Section III, and results from our experimentation are illustrated in Section IV. Section V concludes the paper.

II. RELATED WORK

Mobile devices are evolving rapidly. Every season, new generation of mobile devices are release that are more capable and computationally powerful than the previous generation. Along with the rapid growth of wireless internet technologies that promise high data rate and massive device connectivity, mobile multimedia services and applications can transform the healthcare sector. Numerous studies have been conducted to study the impact of mobile applications in healthcare processes [21]–[23]. Similarly, the use of social media for health-related purposes has also been research upon [24], [25]. Personal health apps are also driving a mobile revolution in health care. In this section, we briefly review the different methods for measuring food intake.

Methods for measuring food intake range from manual dietary assessments [2], [3], [26] to automatic sensing methods. In this section, we briefly review the automatic imaging based methods. Pouladzadeh *et al.* [13] proposed a system that involves capturing an image of the food and processing it through predefined steps, which follow a pipe line architecture. These steps include food image segmentation and food portion recognition. Calorie measurement is done using nutritional fact tables. The system often fails to detect various food portions in mixed food; it also fails to segment them properly. The area measurement technique proposed is based on a depth estimation technique. However, their system uses a dataset that is too simplistic, consisting of food items placed on white plates with smooth texture. Chen *et al.* [27] use a depth camera such as Kinect to estimate the volume of food for calorie measurement. However, dependency of their system on Kinect makes the algorithm unsuitable for normal use. Model-based measurement of food portion size is proposed in [28]. The method consists of three stages i.e. base plane localization, food segmentation and volume estimation. A 2D-3D model to image registration scheme is used for volume estimation. The algorithm does not perform accurately in cases of shadows, reflection, complex food, ingredients and motion blurring. Similarly, Fang *et al.* [29] use special fiducial markers placed in the scenes to estimate the food portion size. Im2calorie [30] estimates food categories, ingredients and volume of individual dishes and calories. However, the calorie annotated dataset used is not sufficient [31]. The main approach for calorie estimation in the above mentioned methods is to start off by recognizing the food category, followed by food portion size estimation and finally calorie estimation using standard nutritional fact tables.

There are other approaches that directly estimate the calories from food images [31], [32]. Ege and Yanai [31] directly estimate food calories from photos of food by simultaneously learning about food categories, ingredients and cooking directions. They argue that simultaneous learning of categories, ingredients and calories will boost performance as there exists a correlation between them.

Various approaches have also been proposed for food recognition only. Ahmed and Ozeki [8] propose two methods to recognize food. These methods include Speeded up Robust Features (SURF) and Spatial Pyramid Matching (SPM). The former method (SURF) requires a dictionary of code words, and histograms are generated against those code words using a linear kernel classification scheme. The latter (SPM) accounts for spatial information by dividing and subdividing the given image into increasingly smaller sub regions and computing histograms in each. Kawano and Yanai [6] propose a real time mobile food recognition system, which continuously acquires frames of the image from the camera device. The user draws boxes around the food items on the screen and food recognition is carried out within the boxes. The graph cut based segmentation algorithm GrabCut is used for accurate food segmentation. Recognitions is performed using the linear kernel SVM (support vector machine). Camera position and viewing direction need to be maintained to obtain more reliable SVM classifications. Convolutional neural network have also been employed for the recognition task [30], [31], [33]–[35] and as a result the recognition accuracy has improved significantly.

Availability of large a dataset is crucial for machine learning based food recognition algorithms. Food-101 [7] is a large publicly available dataset of food images. It contains 101 classes of food items with 1000 images for each class. Similarly, UEC Food 100 [36] contains 100 categories of food images. VIREO Food-172 [37] contains 110,241 food images from 172 categories. Each food image is manually annotated with 300 ingredients. Calorie annotated datasets include [31] and [38]. There is no publicly available dataset that contains subcontinental dishes. Therefore we created a new dataset containing both subcontinental and other common cuisines.

For food attribute estimation, we take a completely different approach and use vector space representation of words from a large dataset using Word2Vec. To get accurate and relevant results from vector space embeddings of words, we collected a large amount of text data from the internet, mostly from food and nutritional and recipe web sites. Semantically related words such as milk and yogurt will appear adjacent in the vector space embedding as compared to milk and apple. The idea is to use distance measure in the vector space to compute food attributes.

III. METHODOLOGY

This section describes the modules comprising the proposed system. The proposed system consists of two major modules:

- Food Recognition:
  Recognizing food items from images
- Attribute Estimation:
  Estimating food attributes of the recognized food item using textual corpus

The dataset used in this research is also described here.

A. DATASET

Our goal is to make a dataset that contains common food items, augmented with subcontinental dishes. We started by experimenting on the publicly available dataset of food images, i.e. Food-101 [7]. It contains 101 classes of food items with 1000 images for each class. Food-101 is designed specifically for multi-class classification. There are other datasets as well that have been used for food recognition previously; one such dataset is Food-5k. Food-5k contains 5000 images, out of which 2500 are of food and 2500 of non-food. However, this dataset can be used only for binary classification to discriminate food items from non-food items and therefore, is not suitable for our task. Moreover, Food-101 does not include food items or classes from the subcontinental cuisine which makes a large portion of the food that people intake in the subcontinental region. Some subcontinental dishes exhibit low inter-class variation and are very similar to each other, so collecting high quality data for accurate classification of different categories is a big challenge. The results returned by Google search engine against textual search queries for food images are quite relevant with very low noise content. Based on such results from Google search engine, our new dataset is created by querying Google against each label of our dataset. The newly formed dataset has classes from Food-101, that are common and eaten everywhere. Thus, the final dataset contains all the food items from Food-101 dataset and 100 additional subcontinental food classes. The dataset is split into training and validation images. Each class contains around 800 training images and 200 validation images.

B. FOOD RECOGNITION

Food Recognition deals with recognition of food item when given an image. Owing to the great success of CNNs, we experimented with top performing pre-trained models to train our dataset using transfer learning. Based on their performance in other domains, we selected several pre-trained CNN models such as VGG-16, VGG-19 [18], Inception-v3 [39], Inception-v4 [40] and ResNet [41]. These models are pre-trained on the Imagenet dataset. Transfer learning is used to train these models on our dataset. The last fully connected layer is removed and appended with dropout, ReLU activations and softmax layers. Fine tuning the model on our dataset took about 15 hours on a single Titan X GPU with 12GB of memory. Models based on Inception-v3 and Inception-v4 gave better performances and were used as the recognition engine in the rest of the paper.

TABLE 1. Parameters used and their values for data augmentation.

Parameter	Value
Width Shift	0.2
Height Shift	0.2
Rotation	90
Rescale	1/255
Shear	0.2
Zoom	0.2
Horizontal Flip	True
Vertical Flip	True

1) FINE TUNING

After initial filtering, we selected Inception-v3 and Inception-v4 as the appropriate model based on their performance on our dataset and proceeded on improving the validation accuracy of the model on our dataset. For this, we employed various techniques. First, intensive data augmentation is performed so that the model is robust to affine variations as much as possible and the images are efficiently trained. In every epoch, various transformations, with random parameters specified by parametric range, are applied on each image of the dataset to produce copies that are transformed from the original image. These include translations, rotations, shearing, zooming and flipping. The optimized results are obtained by setting parameters for data augmentation as mentioned in Table 1. Images are also re-scaled, considering the large values of RGB coefficients for the model to process. Re-scaling involved multiplication of the RGB value by 1/255 factor, so target values are normalized and lie between 0 and 1. This makes the image robust to variations in illumination and makes processing data faster.

Other steps for improving the accuracy include *batch normalization* and *regularization*, to tackle over-fitting, and multi-crop evaluation. In multi-crop evaluation, at the time of testing an image, multiple crops of an image are taken from different regions of the image and each crop is tested individually. The most frequently occurring class in the list of resulting predicted labels is considered. In this work, four crops are obtained by equally dividing the image into four squares and a center crop is also taken. The image is then flipped and the same process is applied. So in total, we get 10 crops for each image. Moreover, learning rates and decay

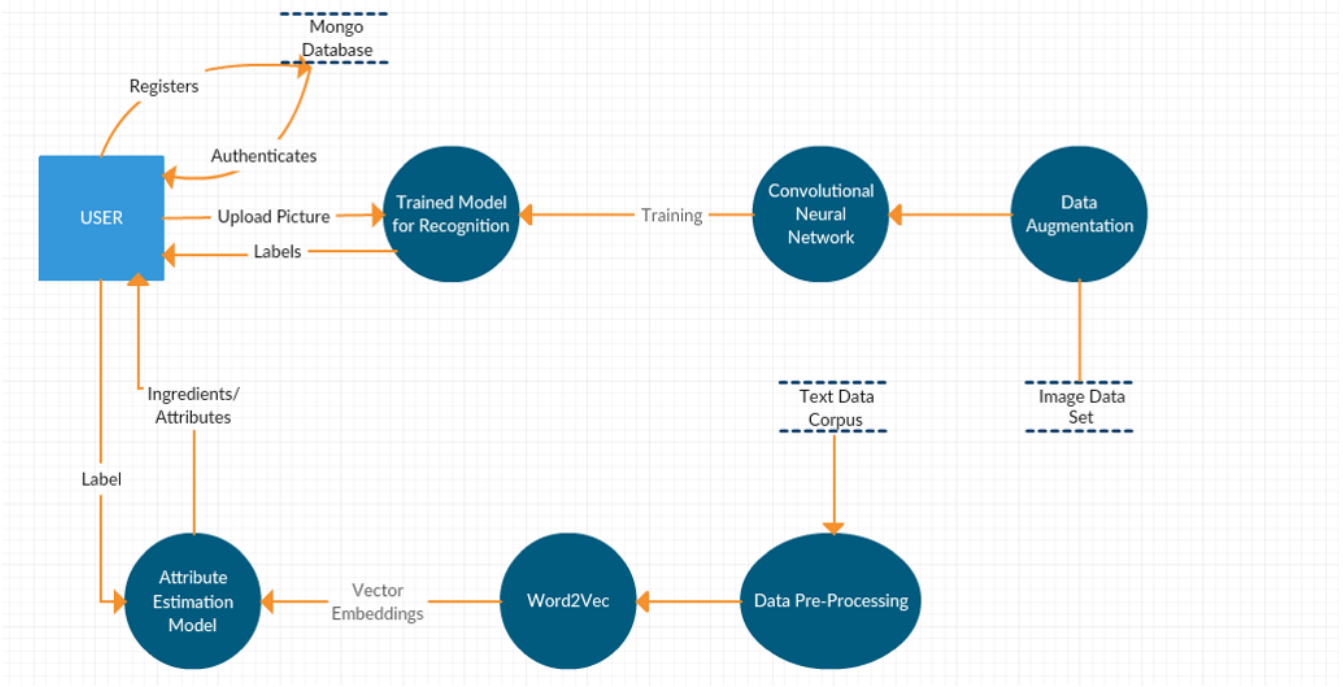


FIGURE 2. The System Flow of the Proposed system.

values are also optimized to get more accurate results. Early stopping saves time as it stops training if the model does not show improvement in the validation accuracy for a set amount of continuous epochs. The last two convolution blocks of the Inception model are also made trainable, along with the final fully connected layer, so that more high level features specific to our dataset are learned.

C. ATTRIBUTE ESTIMATION

The next task after food recognition is to compute the food attributes including the ingredients and their nutritional value. As mentioned in Section II, there are generally two approaches used. In the first approach, the food portion size is estimated and then standard nutritional tables are used to compute the attributes [13], [27], [30]. In the second approach, the food attributes are learned directly from the image [31], [32]. We take a completely different approach and use vector space representation of words from a large dataset [20]. To get accurate and relevant results from vector space embeddings of words, we proceeded to collect a large amount of text data from the internet, mostly from food and nutrition sites. The collected data is then trained using Word2Vec [20], which produces word embeddings. Syntactically and semantically similar words are adjacent or have smaller distance in the vector space as compared to words that are not similar. The motive is to find food attributes by measuring the distance between the food item and the ingredients in the learned vector space. Small distance between the food item and an attribute means they occur closely in the original text and the probability of that attribute appearing in the food item is high. The overall flow is illustrated in Figure 3.

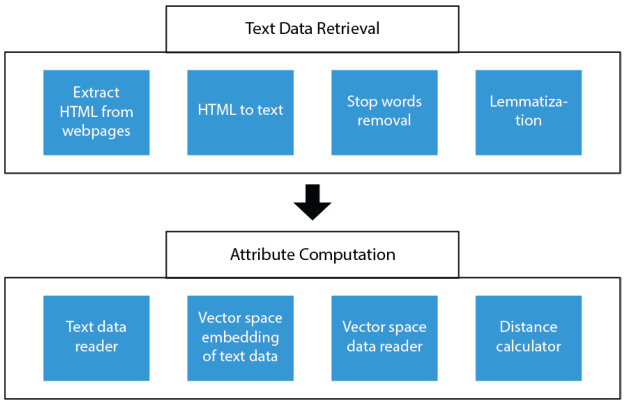


FIGURE 3. Attribute Estimation using Vector Space Embeddings of Textual Data for Ingredients retrieved from Internet.

1) TEXT DATA RETRIEVAL AND PRE-PROCESSING

There are two sources from which text data is obtained for this task. First is Common-Crawl, which is an archive hosted on an Amazon S3 bucket. It is an open repository of web crawl data, which contains data of thousands of web pages, easily accessible through an index and API. From this repository, we collect thousands of web pages of different recipe and nutrition websites. This constituted the main chunk of our text data. The second source is Google Search. To collect data from Google Search results, the web crawler Scrappy is used. We search against each food category, food attribute and ingredient on Google and from the resulting pages, Scrapy is used to retrieve raw text data from the first 500 pages for each search query. The text collected from Google search

TABLE 2. Configuration for training of word embeddings using Word2Vec.

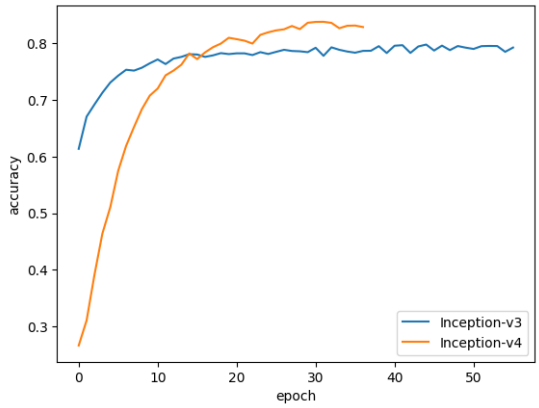
Parameter	Value
Window Size	8
Vector Dimensionality	200
Sample	1e-4
Negative	25
Iterations	15

is more relevant as it return pages corresponding to precise labels while the text from Common-Crawl is more generic.

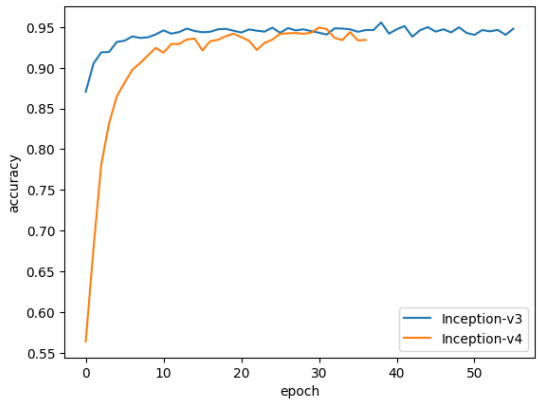
The raw text data obtained from Google and Common Crawl is in HTML format. It is preprocessed to remove HTML tags, CSS, JavaScript code and comments. Such meta-data is removed from the text and tokenized to get individual words. This makes the data usable and individual words can be processed. For semantic-based text processing, the data must be free of auxiliary and irrelevant words such as 'this', 'are', 'an' etc. These words have the highest occurrence probabilities and therefore need to be removed from the text data so that they do not provide hindrance in the learning of relevant data. Such stopwords are removed. Furthermore, text can exist in different forms like past tense, future tense, plural etc. This distributes the probabilities of semantically same words into different forms. To solve this issue, the text is lemmatized. Lemmatization is the mapping of a word to its corresponding root word in the dictionary. For example, words like 'eating' and 'tomatoes' are lemmatized to 'eat' and 'tomato'. The word corpus thus obtained is then used for training purposes.

2) TRAINING AND VECTOR EMBEDDINGS

Word2Vec implements deep learning techniques to efficiently compute vector representation of words in a multi-dimensional space. Each word in the text corpus is considered as an input to the log-linear classifier which learns words that appear within a certain range of the input, considering the fact that words that occur further away from the input might have low probability of being semantically similar. It is based on two architectures, Continuous Bag of Words (CBOW) and Skip-Gram Model (SGM) for computing vector representations from the corpora. We have used the CBOW approach in our work. The embedded vector space allows the proposed method to determine semantic relationships, using a cosine distance metric between word vectors. We have used the Google implementation of Word2Vec in our work. The training of Word2Vec has multiple options such as the type of architecture to use, the dimensionality of vector space, size of window in which to learn occurrence of words, the training algorithm such as softmax or negative sampling, threshold for down-sampling etc. After experimenting with different configurations, the most relevant results are achieved using the values described in Table 2. Word2Vec takes the whole text corpus as an input, creates a vocabulary of words used in that text, learns the distances between those words and returns a binary file containing learned vector embeddings.



(a)



(b)

FIGURE 4. Accuracy vs. Epoch Graph on our dataset with Inception-v3 and Inception-v4. (a) Top-1. (b) Top-5.

3) ATTRIBUTE EXTRACTION

To measure the probability of occurrence of a food attribute in the food item, we compute the distance between the food attribute and the food label in the vector space learned using Word2Vec. A user can be interested in ingredients, nutrition values or characteristics based on the application area. A static list of these attributes is created. We tried to encompass all possible and relevant attributes that can occur in a food item in the lists. After recognizing the food item, the predicted food label along with the food attributes are presented to the Word2Vec, which computes the cosine distance of the vector space representations of the predicted label and the food attributes. The Similarity function of the Word2Vec module returns the probability of one word occurring within the window size of another word based on cosine distance. After finding out the similarity of each attribute with the predicted label, we get its approximate probability of occurring besides the predicted label in the text, i.e. its probability of existing in the food item. Ingredients, nutrition and characteristics occur in text at different scales. Ingredients have the highest frequency, so they are the most accurate as their sample size is large. Nutrition values are a little sparse in

TABLE 3. Accuracies of food recognition process on our dataset and Food101 dataset.

Models		Accuracy							
		Top1		Top2		Top3		Top5	
		Own DS	Food101 DS	Own DS	Food101 DS	Own DS	Food101 DS	Own DS	Food101 DS
Inception-v3	Single Crop	79.8	80	87.9	80.9	91.6	84.8	95	89.6
Inception-v4	Multi Crop	81.81	72.25	-	-	-	-	95.51	89.53
Inception-v3	Single Crop	83.8	78.3	89.8	85.2	92.4	87.9	94.7	90.5
Inception-v4	Multi Crop	85	79.22	-	-	-	-	95.67	90.63

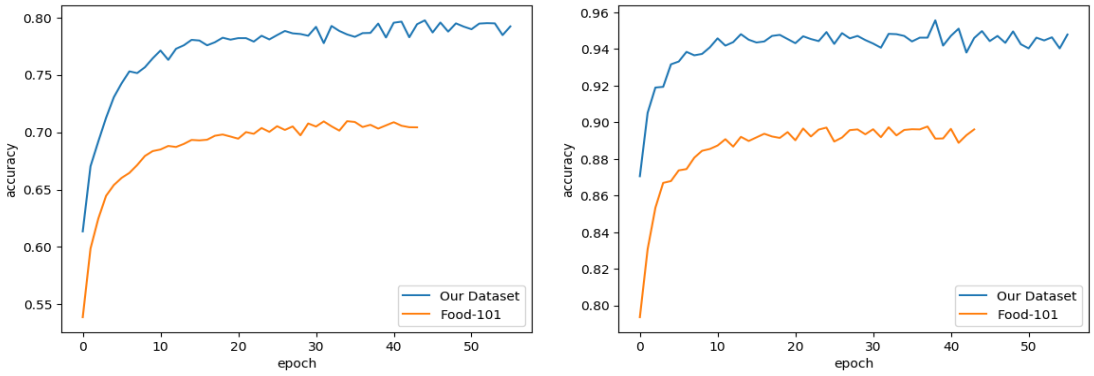


FIGURE 5. Performance of model based on Inception-v3 on our Dataset and Food-101 Dataset. (a) Top-1. (b) Top-5.

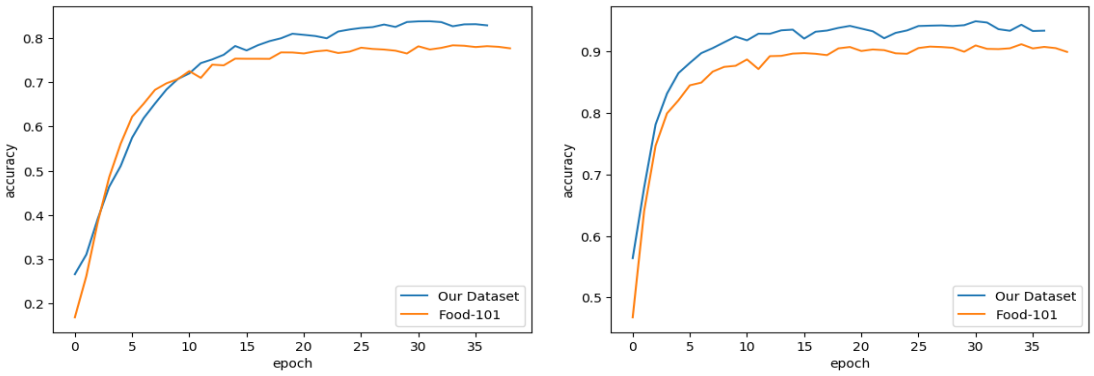


FIGURE 6. Performance of model based on Inception-v4 on our Dataset and Food-101 Dataset. (a) Top-1. (b) Top-5.

the text while characteristics occur very infrequently, so these are less accurate. The probabilities of each of these categories are normalized by dividing the probabilities of each category with the highest probability of that category. This localization gives us a more accurate representation.

IV. RESULTS AND DISCUSSIONS

A. FOOD RECOGNITION

As mentioned in Section III-B, CNN models based on Inception-v3 and Inception-v4 are selected as they perform better than the other models tested. These models are fine tuned on our dataset, as well as for Food-101, for comparison. Recognition accuracy results on our dataset and the Food-101 dataset are shown in table 3. The dataset consists of 200 food categories including subcontinental food classes. The model is trained on 68705 training images and evaluated on the 5284 validation images, disregarding data augmentation. Table 3 reports Top-1, Top-3 and Top-5 accuracies as well as single crop and multiple crop accuracies. Top-k accuracy means that top k predicted labels contain the true class. Multiple crop means that the multiple crops of the image are presented to the recognition model for prediction.

The model is trained for 40 epochs, after which there is no further improvement observed in accuracy as can be seen from Figure 4.

Table 3 shows the accuracy of our system on the Food-101 Dataset. Food-101 consists of 101 food classes, that are most popular on the food sharing website, Foodspotting. The models are trained on the 90900 training images, and

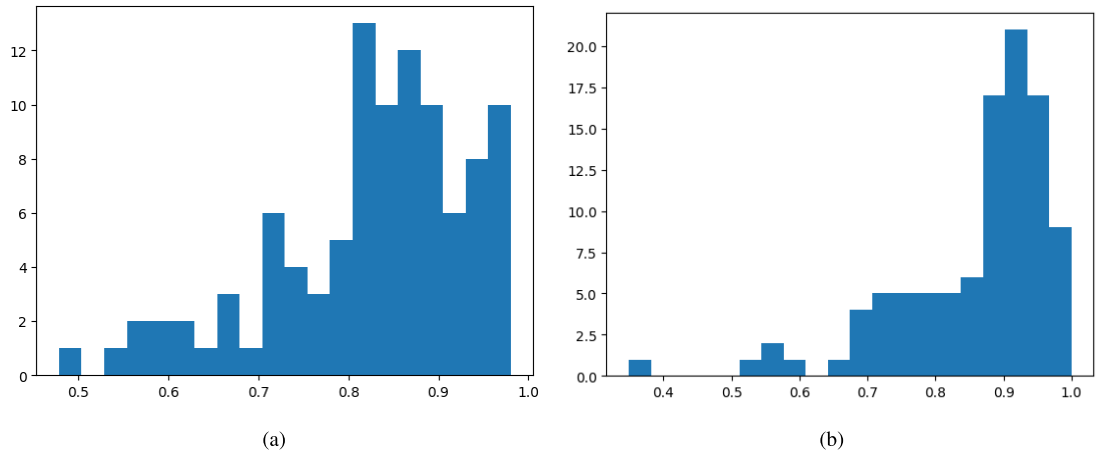


FIGURE 7. Histogram of class accuracies on our Dataset. The *x* axis represents the accuracy bins whereas the *y* axis shows the frequency of occurrence of the classes. The shifting of the peaks towards the right means more classes exhibit high accuracy in Inception-v4 model. (a) Inception-v3 model. (b) Inception-v4 model.

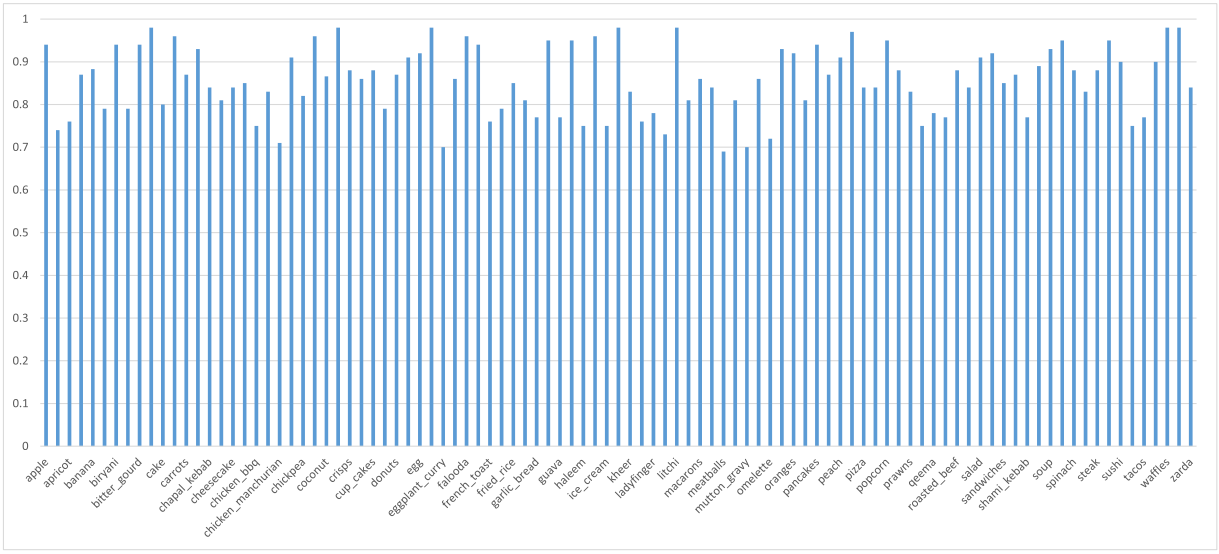


FIGURE 8. Class wise accuracy of Inception-v4 model trained on our dataset.

evaluated on the 10100 validation images, again without data augmentation. The model accuracy leveled out after 40 epoch as can be seen in Figure 4.

Figures 5 and 6 show that the accuracies of the model based on Inception-v3 and Inception-v4 are higher on our dataset than Food-101. This resulted from rigorous filtering based on detailed statistics, selection of high quality images and intense data augmentation so that the model generalizes better on the dataset.

Figure 7 depicts histograms of individual class accuracies for Inception-v4 and Inception-v3. These histogram plots are for the classification accuracies and their corresponding frequencies. The *x* axis represents the accuracy bins whereas the *y* axis shows the frequency of occurrence. The graphs show a comparison of accuracies achieved after training on the Inception v3 and v4 models. The shifting of peaks towards the right means more classes exhibit high accuracy in Inception-v4 than in Inception-v3. Inception v4 being more extensive gave much better accuracy of 85%. In addition to that the Food-101 dataset was also trained on Inception v4 to see how its dataset quality differs from our dataset. According to the results, our dataset gave an accuracy of 85% in comparison to Food-101 which gave 78% accuracy. Class wise accuracy using inception-v4 model is shown in Figure 8. Some of the classes in our data set look very similar to each other, due to which these classes did not exhibit good accuracy. For example, the algorithm was able to rightly classify meatball only 69% of the times, because the meatball looked very similar to shami kabab, pakora and guava.

B. ANALYSIS OF ATTRIBUTE ESTIMATION

Estimated probabilities of attributes in a food item can only be analyzed on an individual basis and cannot be compared with a standard. Each food item has its specific attributes.

We gather text data from two sources, Common-Crawl and Google Search. For comparison, we learn two vector spaces. The first is learned only from the text collected from Google Search; the second is learned from Google Search as well as the Common-Crawl archive. In vector space, the keywords that appear closer occur more closely in the text. Sometimes, other food items of the same cuisine appear beside a food item in the text, so they tend to be close in the vector space. Similarly, when the vector space surrounding an ingredient is computed, other ingredients which are present in the dish appear closer. The proposed method for estimating attributes achieved encouraging results and will be further improved in future work.

V. CONCLUSIONS

This paper presents a system that exploits the extensive use of mobile devices to provide health information about the food we eat. The mobile-based app takes the image of the meal and presents approximate ingredients and nutritional values in food. A dataset is created that consists of common and subcontinental food items. We employ a fine tuned Inception model to recognize food items and propose a method to estimate attributes of the recognized food item. The results are improved via data augmentation, multicrop evaluation, regularization and other similar techniques. 85% accuracy is achieved on our dataset. Our proposed method for estimating attributes also achieved encouraging results. Future endeavors in this domain can include the practical application of this work and more improvements in the android application with advanced features to make it a complete guide for everyday meals.

REFERENCES

[1] M. F. Munsell, B. L. Sprague, D. A. Berry, G. Chisholm, and A. Trentham-Dietz, "Body mass index and breast cancer risk according to post-menopausal estrogen-progestin use and hormone receptor status," *Epidemiol. Rev.*, vol. 36, no. 1, pp. 114–136, 2014.

[2] *Myfitnesspal*. Accessed: Oct. 1, 2018. [Online]. Available: https://www.myfitnesspal.com/

[3] *Samsung Health*. Accessed: Oct. 1, 2018. [Online]. Available: https://health.apps.samsung.com/

[4] *Rise*. Accessed: Oct. 1, 2018. [Online]. Available: https://www.rise.us/

[5] M. Puri, Z. Zhu, Q. Yu, A. Divakaran, and H. Sawhney, "Recognition and volume estimation, of food intake using a mobile device," in *Proc. Workshop Appl. Comput. Vis. (WACV)*, 2009, pp. 1–8.

[6] Y. Kawano and K. Yanai, "Real-time mobile food recognition system," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2013, pp. 1–7.

[7] L. Bossard, M. Guillaumin, and L. Van Gool, "Food-101—Mining discriminative components with random forests," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Zürich, Switzerland: Springer, 2014, pp. 446–461.

[8] A. Ahmed and T. Ozeki, "Food image recognition by using bag-of-surf features and hog features," in *Proc. 3rd Int. Conf. Hum.-Agent Interact.*, 2015, pp. 179–180.

[9] H. Hassannejad, G. Matrella, P. Ciampolini, I. De Munari, M. Mordonini, and S. Cagnoni, "Food image recognition using very deep convolutional networks," in *Proc. 2nd Int. Workshop Multimedia Assist. Dietary Manage.*, 2016, pp. 41–49.

[10] N. Martinel, G. L. Foresti, and C. Micheloni. (2016). "Wide-slice residual networks for food recognition." [Online]. Available: https://arxiv.org/abs/1612.06543

[11] S. Fang, F. Zhu, C. Jiang, S. Zhang, C. J. Boushey, and E. J. Delp, "A comparison of food portion size estimation using geometric models and depth images," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 26–30.

[12] P. Pouladzadeh and S. Shirmohammadi, "Mobile multi-food recognition using deep learning," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 13, no. 3s, p. 36, 2017.

[13] P. Pouladzadeh, S. Shirmohammadi, and R. Al-Maghrabi, "Measuring calorie and nutrition from food image," *IEEE Trans. Instrum. Meas.*, vol. 63, no. 8, pp. 1947–1956, Aug. 2014.

[14] F. Zhu *et al.*, "The use of mobile devices in aiding dietary assessment and evaluation," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 4, pp. 756–766, Aug. 2010.

[15] W. Zhang, Q. Yu, B. Siddiquie, A. Divakaran, and H. Sawhney, "snap-n-eat'food Recognit. nutrition estimation a smartphone," *J. Diabetes Sci. Technol.*, vol. 9, no. 3, pp. 525–533, 2015.

[16] J. Noronha, E. Hysen, H. Zhang, and K. Z. Gajos, "Platemate: Crowdsourcing nutritional analysis from food photographs," in *Proc. 24th Annu. ACM Symp. User Interface Softw. Technol.*, 2011, pp. 1–12.

[17] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[18] K. Simonyan and A. Zisserman. (2014). "Very deep convolutional networks for large-scale image recognition." [Online]. Available: https://arxiv.org/abs/1409.1556

[19] F. Cordeiro, E. Bales, E. Cherry, and J. Fogarty, "Rethinking the mobile food journal: Exploring opportunities for lightweight photo-based capture," in *Proc. 33rd Annu. ACM Conf. Hum. Factors Comput. Syst.*, 2015, pp. 3207–3216.

[20] T. Mikolov, K. Chen, G. Corrado, and J. Dean. (2013). "Efficient estimation of word representations in vector space." [Online]. Available: https://arxiv.org/abs/1301.3781

[21] V. Vodopivec-Jamsek, T. de Jongh, I. Gurol-Urganci, R. Atun, and J. Car, "Mobile phone messaging for preventive health care," *Cochrane Database Syst. Rev.*, vol. 12, p. CD007457, Oct. 2012.

[22] G. Nasi, M. Cucciniello, and C. Guerrazzi, "The role of mobile technologies in health care processes: The case of cancer supportive care," *J. Med. Internet Res.*, vol. 17, no. 2, 2015.

[23] M. S. Marcolino, J. A. Q. Oliveira, M. D'Agostino, A. L. Ribeiro, M. B. M. Alkmim, and D. Novillo-Ortiz, "The impact of mhealth interventions: Systematic review of systematic reviews," *JMIR mHealth uHealth*, vol. 6, no. 1, 2018.

[24] S. A. Moorhead, D. E. Hazlett, L. Harrison, J. K. Carroll, A. Irwin, and C. Hoving, "A new dimension of health care: Systematic review of the uses, benefits, and limitations of social media for health communication," *J. Med. Internet Res.*, vol. 15, no. 4, 2013.

[25] E. Hagg, V. S. Dahinten, and L. M. Currie, "The emerging use of social media for health-related purposes in low and middle-income countries: A scoping review," *Int. J. Med. Informat.*, vol. 115, pp. 92–105, Jul. 2018.

[26] J.-E. Lee, S. Song, J. S. Ahn, Y. Kim, and J. E. Lee, "Use of a mobile application for self-monitoring dietary intake: Feasibility test and an intervention study," *Nutrients*, vol. 9, no. 7, p. 748, 2017.

[27] M.-Y. Chen *et al.*, "Automatic Chinese food identification and quantity estimation," in *Proc. Tech. Briefs SIGGRAPH Asia*, 2012, p. 29.

[28] H.-C. Chen *et al.*, "Model-based measurement of food portion size for image-based dietary assessment using 3D/2D registration," *Meas. Sci. Technol.*, vol. 24, no. 10, p. 105701, 2013.

[29] S. Fang, F. Zhu, C. J. Boushey, and E. J. Delp, "The use of co-occurrence patterns in single image based food portion estimation," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Nov. 2017, pp. 462–466.

[30] A. Meyers *et al.*, "Im2Calories: Towards an automated mobile vision food diary," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 1233–1241.

[31] T. Ege and K. Yanai, "Image-based food calorie estimation using knowledge on food categories, ingredients and cooking directions," in *Proc. Thematic Workshops*, 2017, pp. 367–375.

[32] T. Miyazaki, G. C. de Silva, and K. Aizawa, "Image-based calorie content estimation for dietary assessment," in *Proc. IEEE Int. Symp. Multimedia (ISM)*, Dec. 2011, pp. 363–368.

[33] K. Yanai and Y. Kawano, "Food image recognition using deep convolutional network with pre-training and fine-tuning," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops (ICMEW)*, Jun./Jul. 2015, pp. 1–6.

[34] Y. Cui, F. Zhou, J. Wang, X. Liu, Y. Lin, and S. Belongie, "Kernel pooling for convolutional neural networks," in *Proc. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 3049–3058.

[35] C. Liu *et al.*, "A new deep learning-based food recognition system for dietary assessment on an edge computing service infrastructure," *IEEE Trans. Services Comput.*, vol. 11, no. 2, pp. 249–261, Mar. 2018.
[36] Y. Matsuda, H. Hoashi, and K. Yanai, "Recognition of multiple-food images by detecting candidate regions," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2012, pp. 25–30.
[37] J. Chen and C.-W. Ngo, "Deep-based ingredient recognition for cooking recipe retrieval," in *Proc. ACM Multimedia Conf.*, 2016, pp. 32–41.
[38] *Allrecipes*. Accessed: Oct. 1, 2018. [Online]. Available: http://allrecipes.com/
[39] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 2818–2826.
[40] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Proc. AAAI*, 2017, pp. 4278–4284.
[41] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.

Authors' photographs and biographies not available at the time of publication.