Automatic Fruits Calories Estimation through Convolutional Neural Networks

Ghazanfar Latif
Department of Information Technology,
Prince Mohammad bin Fahd University,
Al-Khobar, Saudi Arabia
glatif@pmu.edu.sa

Batool Alsalem, Wejdan Mubarky
Department of Computer Science,
Prince Mohammad bin Fahd University,
Al-Khobar, Saudi Arabia
{201302893, 201400818}@pmu.edu.sa

Nazeeruddin Mohammad
Cybersecurity Center
Prince Mohammad bin Fahd University,
Al-Khobar, Saudi Arabia
nmohammad@pmu.edu.sa

Jaafar Alghazo
Department of Computer Engineering,
Prince Mohammad bin Fahd University,
Al-Khobar, Saudi Arabia
jghazo@pmu.edu.sa

ABSTRACT

With the increased interest in healthy eating and diet control, the importance of keeping track of daily calorie intake is an important topic nowadays. In this paper, we propose a novel technique to automatically calculate calories based on the picture of various foods on a plate. As a first step, this paper concentrate on food plates with fruit and vegetable content only. A new dataset was developed internally for this research consisting of a total of 41,509 images. In this paper, we propose a Deep Convolutional Neural Network (DCNN) for the automatic recognition process. A custom design application was developed for capturing the image, recognition, and automatic calculation of calories. The results showed that over 92% recognition rates were achieved on most fruits and vegetables.

CCS Concepts

•Information systems → Information systems applications → Decision support systems → Expert systems.

Keywords

Calories Estimation, Convolutional Neural Networks (CNNs), Fruits Classification, Deep Learning. Vegetable Classification, Image Processing, Agriculture produce

1. INTRODUCTION

The concept of healthy eating, weight control, healthy habits has always been of great importance. The promotion of healthy living among all population ages from young to old is handled by many non-profit organizations and government agencies. Healthy living means a healthier population that is more productive. Healthy living also affects the health system in many countries and national budgets for the healthcare sector as well.

ICCTA 2020, April 14–16, 2020, Antalya, Turkey

ACM ISBN 978-1-4503-7749-2/20/04…$15.00
https://doi.org/10.1145/3397125.3397154

As part of the healthy living concept, people are required to monitor their daily intake of calories. However, the general public lacks knowledge about foods and calories in various food items. Many countries now require restaurants and food producers to write the number of calories in meals and other produce. However, this is only enforced in certain countries and still does not cover all food and food products. Thus, a methodology is required to make it easier for people to know the number of calories in the individual portions they consume. In the age of the Internet, Cloud Computing, Artificial Intelligence (AI), Internet of Things (IoT), the general public expects automatic information at their fingertips even information about the calories in a plate of food they are about to consume. Applications have been developed for calculating calories, however, they are not friendly as it requires the user to interact with the application, enter the types of food and portion sizes which is a time-consuming task and is not suitable for all ages as well. With the renewed interest in AI image processing algorithms especially CNNs, research is now underway for automatic recognition and calculation of calories based on food content on a plate.

In this paper, we propose an Application with Deep CNN Algorithm for the automatic recognition of fruit and vegetable (referred to as produce) content and automatic calculation of calorie content. A custom dataset was also created and developed for this research.

The purpose of our research is to create an application that can almost certainly give an accurate estimation of calories of vegetables and fruits specifically. The application will be available to people of all ages. The application will be developed such that it can be used both online and offline. The first step is to calculate automatically the calories for food plates containing vegetables and fruits.

The rest of the paper is organized as follows: in section 2 we detail the literature review, in section 3 proposed method is described. The details of the experimental results are discussed ion in section 4 and the article is concluded in section 5.

2. LITERATURE REVIEW

Even though some food packages come with nutrition and calories facts, they still not accurate; thus, scientists decided to use machine learning algorithms and computer vision to help people determining the calories' values of their food. This problem has been stated in multiple articles and research papers. Using the help of image processing and segmenting, the researchers decided to create a program that can estimate calories based on its mass, size, shape, texture, and color. The authors propose a system that relies on food image processing and fact tables for nutrition. Using a calibration technique, they go through the steps of capturing a picture before and after food intake. The dataset consisted of 3000 images. The results of using SVM show an accuracy of 92.21%, 85%, 35%-65% [1]. Another research was conducted which discussed calorie content estimation for dietary evaluation. In [2], the authors developed a web service that had public access where users would take food images daily. In this case, the pictures uploaded are the dataset. Experts in the machine learning field analyze the images and set the calories of images using different kinds of machine learning algorithms. The dataset is 6512 food images taken by ordinary users. This Web gives an accuracy of the images up to 81%. A new idea for image recognition and calories estimation was proposed by some researchers in this review. The idea is for users to take pictures of their daily meals and upload them on the web to get the result they need, which is the calories. [2]. In [3], the authors propose a technique based on Support Vector Machine (SVM) and the feature extraction by applying deep learning on the dataset given to the machine, which provides the input based on vector features attained from feature extraction. The dataset used was collected from different online sources, it consists of 10 categories of different types of food. The total number of images in the training set is 3,200. The initial system shows an accuracy of 72.53%, but with a powerful SVM algorithm, the accuracy level increased to 90.66%. In [4], the authors proposed an easier way of training using a deep residual learning framework. This method was evaluated using the ImageNet classification dataset that contains 1000 classes of 1.8 million images. The result was 82% accuracy. In [5], the authors propose a cloud-based calorie estimation algorithm where images are uploaded to the cloud In the cloud, deep learning has been used and the result will be displayed on the user's smartphone. For image processing, SVM was used and MapReduce was used for the cloud. A high percentage of accuracy up to 97.5% was reported however the size of dataset was not reported and thus this could be attributed to a small size dataset.

Automatic recognition used CNNs and feature-based image processing techniques have been used successfully for various applications about daily life activities. In [6], the authors propose a system that uses CNN for health-related applications with the use of IoT to monitor and predict emergency health issues before they occur. The state of the art system integrates many new technologies including cloud computing as well. Accuracy was not reported due to the lack of proper dataset. In [7], the authors proposed an automatic real-time smart system for license plate recognition using deep CNNs and reported a 90.6% recognition rate. The concept of the application-based system was also used in [7]. In [8], the authors automate calories estimation by using CNN to do the work of recognizing the images and label each kind of food to its calories. The dataset was collected from Japanese and American recipes. The Japanese dataset consists of 83,000 different images whereas, the American dataset consists of 24,000 images. The result of the training for both datasets is quite similar, the accuracy of the Japanese data is 79% and the accuracy of the American data is 88%. In [9], another application-based system was proposed using feature-based artificial neural networks (ANNs) for the automatic recognition of Multilanguage handwritten digits. This is of particular interest in the fruit and vegetable calorie estimation system proposed in this paper because Multilanguage is a non-homogenous dataset similar to various fruits on a plate. In [9], the authors report a recognition rate of 96.73% using random forest along with structural features.

In [10], a study was constructed in Istanbul Sehir University has used the convolutional neural network (CNN) for recognizing the images of different food. However, in this research, they used a technique for identifying each piece of food on a plate individually and showed the result for everything on the plate separately instead of summing up the calories of all the food on the plate. The dataset contains 9,132 images and 10 classes. The dataset is trained using SVM models and the training produced an accuracy of 88%. In [11], the University of Tokyo has constructed a study for food image recognition using CNN. Deep learning was done on a Japanese dataset that consisted of 170,000 images, broken down into ten different categories of common Japanese food. The accuracy of the training on this dataset is 67%. Due to the high availability of iPhones and Android phones, it is promising to use these devices for real-time recognition. In the University of Electro-Communications, a study proposed the real-time Mobile food recognition system. SVM was used and the dataset of this study uses fifty categories with more than 100 images in each category. The accuracy gained from this experiment is 81.6% [12].

In [13], the authors propose a calorie content measurement technique of a food sample. The technique involves a system that includes an estimating unit to determine the fat content and water content of a food sample. The data is collected using a calorie measurement module and it represents and enables estimation of both contents of a food sample. A computing device takes the stored data collected by the estimating unit and processes it. Many food items eaten are not as accurately described by an index value but are variable when it comes to their densities. For additionally required accuracy, extra parameters are added such as volume and temperature can be measured to calibrate the estimating unit.

In [14], the authors propose a food detection dataset for calorie measurement. A system that uses Vision-Based Measurement (VBM) to improve the accuracy of food intake reporting, and the dataset presented in this paper aids in different types of food recognition and learning methods. The dataset involves 3000 photos of various food types taken from different cameras with unique lightings. Since the researchers included single and mixed food portions, the food recognition technique that trains the dataset can improve accuracy when showing results of combined and separated food. In [15], the authors propose a different technique for food categorization. The proposed system measures food calorie and nutrition via mobile to help manage daily food intake. It mainly focuses on the recognition of food as well as the training phase in which the classification algorithm is used. The technique is known as cloud-based SVM which is better used for classifying objects in a cluster. Based on the results of the usage of this technique and keeping the database up-to-date, the accuracy of the recognition phase increases in single food portions, as well as mixed and non-mixed food portions. The accuracy of the LIBSVM approach is around 20%. In [16], the authors propose a technique based on SVM. The dataset consisted of pictures from Corel stock photos, and they were used as a classification to evaluate heavy-tailed radial basis function (RBF) kernels, to surpass traditional polynomial or Gaussian RBF kernels in performance. The authors also discovered that a simple remapping can have an impact on the performance of the SVM's that run linearly to such an extent that, for such a problem, they become a valid RBF kernel alternative. In [17], the authors propose a smart system that enables a photo of the food sample to be taken prior and after intake to estimate the consumption of nutrient elements and calories based on the chosen food.

In [18], the authors proposed the use of deep learning neural networks to measure food calorie. The system is available on smartphones where the user can capture a photo of a food sample and do an automatic calorie intake measurement. In order to take advantage of images from different sources, the dataset was decided based on what images users will take and share publicly through a Web service. With 6512 images in a dictionary, the approach shows low accuracy. However, the results of the recognition phase of single food segments show a 99% accuracy. In [19], the authors propose a cloud-based virtualization model with computational power to run an e-health mobile application efficiently and provide it with the flexibility of using cloud resources. By using virtual swapping, the process can be expedited. The dataset was applied on 5 categories, pomegranate with an orange, a single bread slice, 2 slices of bread, 2 bananas, and a single banana. For time accuracy, it improved by 42% and for the calorie estimation method by 20.5%.

3. PROPOSED METHOD

The proposed Produce Calorie mobile application is designed to be user friendly. The user captures the image of the plate containing the produce and submits the image. The image is then transmitted over the cloud to a cloud based algorithm for the recognition and calorie estimation algorithm. The results are transmitted back to the user via the application in fairly short time due to the fast processing speed of the algorithms. Figure 1 shows the workflow described for the produce calorie mobile application.

Figure 1. Workflow of the proposed system

3.1 Dataset for Experiments

A customized dataset was developed for the purpose of this study. The dataset used in this research consists of 20 diverse kinds of fruits and 20 kinds of vegetables. The data is divided into 16 different classes; 8 classes of fruits and 8 classes of vegetables. Each class has a different combination of 5 different fruits/vegetables. However, there are some classes that consist of a combination of 4 fruits/vegetables. The fruits and vegetables were captured on different plates' sizes, shapes, and colors. The total number of images in the dataset is more than 41,509 images with around 4000 images in each class. Table 1 details the dataset images and numbers.

The images were taken from different angles (top-view, side-view and all other possible angles). Also, the images were taken using different types of cameras that have different resolutions. 6 cameras were used to collect this dataset. Some of these cameras were the ones that are built-in in smartphones such as Galaxy Note 8, iPhone 10, Huawei Mate 10 lite, iPhone 6 plus, and others. Apart from the cameras and different angles, the images were also taken in different places to ensure that the backgrounds of plates are different. Images in every class have almost 12 different backgrounds in order to have a better and accurate training set to be trained in the machine. Having all the dataset captured with the same background will make the machine bias and will not be able to recognize any plate that put in a different background. The lights were also considered when taking the images, images were taken under different light sources to maintain the accuracy of the dataset under every possible light. . Figure 2 shows the sample of fruit images from the newly build dataset.

Table 1. Summary of our newly developed dataset

Set of Fruits	Class	# of Produce	Images
Mango, Grape, Plum, Kiwi, Pear	1	5	4,363
Apple, Orange, Banana, Pomegranate, Strawberry	2	5	4,766
Pineapple, Fig, Peach, Apricot, Avocado	3	5	4,878
Summer Squash, Lemon, Lime, Guava, Raspberry	4	5	4,082
Banana, Pomegranate, Orange, Grape	5	4	505
Guava, Pear, Lime, Apricot	6	4	1,055
Apple, Strawberry, Plum, Kiwi	7	4	513
Avocado, Lemon, Raspberry, Fig	8	4	1,054
Broccoli, Cauliflower, Chili pepper, Bell pepper, Eggplant	9	5	4,255
Garlic, Corn, Zucchini, Radish, Okra	10	5	4,046
Tomato, Cucumber, Onion, Potato, Carrot	11	5	4,047
Spinach, Green beans, Lettuce, Turmeric, Scallion	12	5	4,885
Tomato, Lettuce, Onion, Potato	13	4	502
Garlic, Turmeric, Corn, Green beans	14	4	1,003
Cucumber, Carrot, Chili pepper, Cauliflower	15	4	522
Eggplant, Bell pepper, Scallion, Radish	16	4	1,033

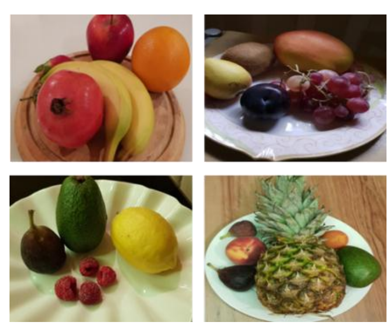

Figure 2. Samples of fruit images from the new dataset

CNN test has also been constructed on an available dataset imported from Kaggle [20]. This dataset is called Fruit-360 and consists of segmented fruit images only. Fruit-360 has 81 fruit types, but we minimized it to 18 fruits only in order to fit with the goal of this research. The whole dataset contains 22,3411 images, 16,731 images for training and 5,610 images for testing.

3.2 Convolutional Neural Networks (CNN)

The Convolutional Neural Networks is a specific type of Artificial Neural Networks which is commonly used in Machine Learning to help computers recognize and classify the objects of an image. Moreover, CNNs have a different architecture than the regular Artificial Neural Networks where each layer is fully connected to all neurons of the previous layer. The CNN architecture consists of two main parts, the hidden layers, and the classification part. In the hidden part and the feature extraction part, the network will implement a sequence of pooling and convolutions operations during the detection of the features. If you have an image of a zebra for example, here is where the machine will realize that it has stripes, four legs, and two ears. In the classification part, the fully connected layers will be delivered as a classifier on top of all the extracted features and assigns the probability of how the algorithm predicts each object in an image.

Additionally, CNN works better with image processing and image classification because the other classifiers with pixel vector might lose a huge number of spatial interactions between pixels in an image. However, CNN uses adjacent pixel information effectively to downsample the image by convolving it and gives the predictions of the objects in the last layer. There are many important terms used in CNN Architecture like the batch size and the number of epochs. Batch size is a hyper-parameter used in deep learning to define samples of the dataset that will be transmitted through the network in each iteration. The larger the batch value, the faster and higher accuracy the training becomes. One epoch is one pass of the whole dataset through the network; one pass means a forward pass and a backward pass. Since an epoch is too large to be given to the computer at once, it is divided into several, smaller batches. The dataset will be passed to the network multiple times using an iterative algorithm called Gradient Descent.

The weight changes as the number of epochs increases, making the curve go from the under-fitting graph to the optimum graph to the over-fitting. The size of the images of the training model must be resized to improve the time and the performance; one way of resizing images is by squaring them. The dropout layer is another important layer in CNN. Dropout is used to improve overfit on neural networks. It is recommended to use dropout along with other techniques like L2 Regularization. When looking at classification errors, we notice that the testing error gets smaller while using Dropout.

Three CNN models are proposed with a different set of layers and experiments are performed using the newly build dataset as well as the online Kaggle dataset. The first CNN model contains 9 layers, the second model contains 11 layers and the third model contains 15 layers.

4. EXPIREMENTAL RESULTS

Table 2 shows the results obtained in this paper. For correctly and accurately recognitions different produce individually. Aside from Oranges, the rest of the produce achieved a recognition rate higher than 90% and reached a maximum of 96% accuracy.

Table 3 shows the results of the proposed algorithm with the three models. Among the 3 CNN models, Model 3 achieves the highest accuracy of the Fruits-360 dataset with an accuracy of 95%.

Table 2. Best case experimental results for different fruits using different CNN models

Class	CNN Model	Training		Testing	
		Acc. %	Loss	Acc. %	Loss
Mango	3	93	0.42	92	0.16
Grape	1	96	0.26	92	0.14
Plum	1	91	0.43	92	0.13
Kiwi	2	81	0.44	92	0.14
Pear	1	94	0.24	92	0.14
Apple	2	93	0.37	92	0.15
Orange	1	87	0.42	91	0.16
Banana	1	93	0.43	92	0.14
Pomegranate	2	92	0.32	92	0.15
Strawberry	3	96	0.27	92	0.15
Pineapple	1	92	0.40	92	0.13
Fig	2	85	0.71	91	0.16
Peach	1	81	0.92	92	0.14
Apricot	1	91	0.48	92	0.13
Avocado	2	91	0.37	92	0.14
S. Squash	3	93	0.28	92	0.14
Lemon	1	93	0.40	92	0.14
Lime	2	95	0.41	92	0.15
Guava	1	94	0.29	92	0.14
Raspberry	1	93	0.35	92	0.13

Table 3. Experimental results for Fruits-360 dataset

CNN Model	Test Loss	Test Accuracy
1	0.72	74%
2	0.42	89%
3	0.17	95%

Table 4 shows the results obtained from the proposed algorithm in comparison with similar methods mentioned in previous literature. The proposed method achieves a recognition accuracy of 92% on the new dataset and recognition accuracy of 95% on the fruit-360 dataset. The comparison in table 4 shows that the proposed algorithm outperforms all similar previously proposed algorithms in the literature.

Table 4. Comparison of experimental results with the latest existing methods

Source	Method	Dataset (Images)	Accuracy
Proposed Method	CNN	41,509 (new dataset)	92%
Proposed Method	CNN	22,3411 (Kaggle)	95%
Chapelle et al. [13]	SVM	2670	84%-89%
Zhang et al. [4]	ResNet	1.28 million	82%
Shirmo et al. [7]	SVM	9,132	88%
Baghel et al. [3]	SVM	3,200	90.66%
Aizawa et al. [5]	CNN	83,000 Japanese, 24,000 for American	79% for Japanese and 88% for American

5. CONCLUSION

With the help of deep learning and CNN, we were able to achieve the goal of this research by training models on a large dataset of produce that was specifically developed for this study. The proposed method achieved high recognition accuracy that outperformed previously proposed algorithms in the literature. The CNN code was tested on a single image and produces values that are sent to the android studio to be multiplied by calories according to the quantity of each item on the plate. Our research could serve many people who care about their health and diet and it could also help old people to stay fit and healthy because it is very easy to use since it only requires a click of a button. In the near future, this application will be published for public use after some improvements to include other types of food. For future work, the dataset will be developed further and published online for the benefit of other researchers in this area. Future work will also continue developing algorithms for the classification and recognition of other food items on a plate.

6. REFERENCES

[1] Pouladzadeh, Parisa, Shervin Shirmohammadi, and Rana Al-Maghrabi. "Measuring calorie and nutrition from food image." IEEE Transactions on Instrumentation and Measurement 63.8 (2014): 1947-1956.

[2] Miyazaki, Tatsuya, Gamhewage C. de Silva, and KiyoharuAizawa. "Image-based calorie content estimation for dietary assessment." Multimedia (ISM), 2011 IEEE International Symposium on. IEEE, 2011.

[3] Calorie Estimation from Fast Food Images - Machine learning. (n.d.). Retrieved from http://cs229.stanford.edu/proj2015/151_report.pdf

[4] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). doi:10.1109/cvpr.2016.90

[5] Kuhad, P., Yassine, A., & Shimohammadi, S. (2015). Using distance estimation and deep learning to simplify calibration in food calorie measurement. 2015 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA). doi:10.1109/civemsa.2015.7158594.

[6] Latif, G., Shankar, A., Alghazo, J. M., Kalyanasundaram, V., Boopathi, C. S., & Jaffar, M. A. (2019). I-CARES: advancing health diagnosis and medication through IoT. Wireless Networks, 1-15.

[7] Alzubaidi, L., Latif, G., & Alghazo, J. (2019, October). Affordable and Portable Realtime Saudi License Plate Recognition using SoC. In 2019 2nd International Conference on new Trends in Computing Sciences (ICTCS) (pp. 1-5). IEEE.

[8] Miyazaki, T., Silva, G. C., &Aizawa, K. (2011). Image-based Calorie Content Estimation for Dietary Assessment. 2011 IEEE International Symposium on Multimedia. doi:10.1109/ism.2011.66

[9] Alghazo, J. M., Latif, G., Alzubaidi, L., & Elhassan, A. (2019). Multi-Language Handwritten Digits Recognition based on Novel Structural Features. Journal of Imaging Science and Technology, 63(2), 20502-1.

[10] Pouladzadeh, P., & Shirmohammadi, S. (2017). Mobile Multi-Food Recognition Using Deep Learning. ACM Transactions on Multimedia Computing, Communications, and Applications,13(3s), 1-21. doi:10.1145/3063592

[11] Subhi, M. A., & Ali, S. M. (2018). A Deep Convolutional Neural Network for Food Detection and Recognition. 2018 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES). doi:10.1109/iecbes.2018.8626720

[12] Kawano, Y., & Yanai, K. (2013). Real-Time Mobile Food Recognition System. 2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops. doi:10.1109/cvprw.2013.5

[13] Webster, Jack Mathew, and Vasile Bogdan Neculaes. "System and method for measuring calorie content of a food sample." U.S. Patent Application No. 12/873,067.

[14] Pouladzadeh, Parisa, Abdulsalam Yassine, and Shervin Shirmohammadi. "Foodd: Food detection dataset for calorie measurement using food images." International Conference on Image Analysis and Processing. Springer, Cham, 2015.

[15] Pouladzadeh, Parisa, et al. "Cloud-based SVM for food categorization." Multimedia Tools and Applications 74.14 (2015): 5243-5260.

[16] Chapelle, Olivier, Patrick Haffner, and Vladimir N. Vapnik. "Support vector machines for histogram-based image classification." IEEE transactions on Neural Networks 10.5 (1999): 1055-1064.

[17] Villalobos, Gregorio, et al. "An image processing approach for calorie intake measurement." Medical Measurements and Applications Proceedings (MeMeA), 2012 IEEE International Symposium on. IEEE, 2012.

[18] Pouladzadeh, Parisa, et al. "Food calorie measurement using deep learning neural network." IEEE International Instrumentation and Measurement Technology Conference Proceedings (I2MTC). 2016.

[19] Peddi, Sri Vijay Bharat, Abdulsalam Yassine, and Shervin Shirmohammadi. "Cloud based virtualization for a calorie measurement e-health mobile application." 2015 IEEE International Conference on Multimedia & Expo Workshops (ICMEW). IEEE, 2015.

[20] Fruits 360 Dataset on Kaggle. Downloaded from: https://www.kaggle.com/moltean/fruits