Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 174 (2020) 448–453

Procedia Computer Science

www.elsevier.com/locate/procedia

2019 International Conference on Identification, Information and Knowledge in the Internet of Things (IIKI2019)

Machine Learning Based Approach on Food Recognition and Nutrition Estimation

Zhidong Shen[1], Adnan Shehzad[1], Si Chen[1], Hui Sun[2], Jin Liu[3]

[1]*Key Laboratory of Aerospace Information Security and Trusted Computing, Ministry of Education, School of Cyber Science and Engineering, Wuhan University, Wuhan, 430079, China*
[2]*Zhongnan Hospital, Wuhan University, Wuhan 430071, China*
[3]*Computer School, Wuhan University, Wuhan430072, China*

Abstract

Nowadays, standard intake of healthy food is necessary for keeping a balanced diet to avoid obesity in the human body. In this paper, we present a novel system based on machine learning that automatically performs accurate classification of food images and estimates food attributes. This paper proposes a deep learning model consisting of a convolutional neural network that classifies food into specific categories in the training part of the prototype system. The main purpose of the proposed method is to improve the accuracy of the pre-training model. The paper designs a prototype system based on the client server model. The client sends an image detection request and processes it on the server side. The prototype system is designed with three main software components, including a pre-trained CNN model training module for classification purposes, a text data training module for attribute estimation models, and a server-side module. We experimented with a variety of food categories, each containing thousands of images, and through machine learning training to achieve higher classification accuracy.

Keywords: Food Recognition, Nutrition Estimation, Machine Learning, Deep Learning, Convolutional Neural Network

1. Introduction

Because people are very keen on measuring weight, healthy diets, and staying away from obesity, there is an increasing demand for food calorie measurement. Adult obesity is increasing at an alarming rate. The main source of obesity is the difference between dietary intake and the energy people get from the diet. High-calorie intake may be injurious and lead to various diseases. Breast, colon and prostate cancers are caused by high calorie intake. High calorie intake is the second leading cause of cancer. Dietitians have determined that the standard intake of a number of calories is required to keep the right balance of calories in the human body. As reported by the world health organization, more than 110th of the adult population in the world is obese. Obesity is a medical condition in which excess body fat has accumulated to the extent that it may have a negative effect on health [1]. If the amount of food a person takes daily is higher than the amount of energy he utilized then we can say that the respective person is becoming obese. Obesity and being overweight are interconnected to many dangerous and chronic diseases. In 2013, the American Medical Association officially declared obesity as the disease that has serious consequences on patients health and therefore requires medical treatment [2]. Therefore, daily intake measurements are important for losing weight and maintaining a healthy diet and weight for normal people. Only a timely measurement of daily food consumption can make obese people lose weight in a healthier way, and can also make healthy people better healthy. The traditional method is mainly based on the analysis of the user's record of food intake in the past 24 hours, and the clinical display has certain effects, but these methods often cause the patient's uneasiness to be forgotten by the user or the broadcast that the user does not want to use these programs [3]. As a result, researchers have replaced typical

10.1016/j.procs.2020.06.113

clinical procedures, and they have been exploring simpler and computerized possible ways to measure daily calorie estimates.

In this era, smart devices are playing an increasingly important role in daily life, and the use of smart devices for the treatment of various diseases is not uncommon[4]. To accomplish this goal, we propose a system or application to assist normal people as well as obese people in balancing their diet by measuring daily intake food attributes and ingredients through their ease. The proposed application will enable the user to figure out the content of the food item by providing the photograph of food to the system. The application will detect the food items within the photograph and recognize them using Convolution Neural Network. The system will also be able to estimate the food attributes by crawling data from the Internet.

The proposed system will allow not only the obese person but also the healthy person so that people can plan well for their daily intake calories. We will contribute to this thesis in the following ways.

- We propose a transfer learning based novel system that automatically performs the exact classification of the food image and estimates the food attributes.
- We present the dataset for evaluating current system and other deep learning-based recognition systems that will be developed in the future.
- There is no data set that contains subcontinental dishes available to the public, we created a new set of data that includes both subcontinental and other common cuisines.

2. Pre-Trained Model Selection

Here we divide our proposed methodology intro three separate parts. The first part has to deal with the transfer learning-based CNN models, the second part has to do with the text recovery from different sources while the third part has to deal with the text data training.

2.1 Pre-Trained Convolutional Neural Network Model

A pre-trained network model is used in machine learning to overcome the problem that the system gets stuck in local solution while in its training age. These models can carry out machine training to respond immediately to different data. A CNN model that we used in our suggested process of transferring learning -based food recognition and extraction attributes uses a variety of food items from our prepared dataset to get different characteristics from an object [5].

2.2 Dataset Preparing and Per-processing Phase

To obtain the needed characteristics from the images of various foods we assign for our research, we categorize each image into its corresponding class. To this end, with the help of different attributes, we distinguish each and every class. For our study, the size of the text data we receive from the internet is nearly 1.8 GB. We used two completely different frameworks to gather data. Common Crawl [6] is the first and Scrapy [7] is the second. We collected about 100 MB of data using Scrapy while using Common Crawl we collected 1.7 GB of data.

The dataset we created includes hundreds and thousands of pictures of various foods. For our research study, some images are relevant and some are not. Filtering the data set is remarkable in the preparation of a model. We use the Data Augmentation concept to improve the efficiency of training data. We perform image transformation in data augmentation [8]. To train transformation parameters, we implement Spatial Transform Network [9]. Once the training is complete, these parameters are applied to the image of the food and the image is transformed.

2.3 Textual Data Model Training

Word2Vec is a machine learning tool that helps us for the computation of vector representation of different words. Word2Vec is a two-layer neural network which is used as the substitution of the clustering algorithm because World2Vec is much powerful algorithm than the clustering algorithm [10]. During this study, we used word2vec, continuous Bag of words and skip Gram for the training of text data.

For the extraction of attributes and ingredients, first, we classify and divide the attributes according to our requirements. This classification and division are helpful when we find the distance of attributes and ingredients with respect to their corresponding classes. In order to find the distance of attributes and ingredients, the procedure that we follow is the fixation of food class and then iterate all attributes and ingredients against it. Here we refer the food is an attribute and the food item is referred to as an ingredient. In this way, we extract attributes and ingredients from the trained word2vec model

3. System Design and Implementation

The system design, system flow and implementation along with its results and evaluation are shown in this chapter. The aim of this demonstration is to throw light on the components, flows, and tools that we have used while implementing our system and their collaborative working to get the desired results and functionalities.

3.1 Architectural Overview

The system that we design for our system just used the server-side architecture. The main aim of our system is to improve the accuracy of pre-trained models so that the developers and architect can take advantage of this system by making their own web-based and Android-based application on the client side.

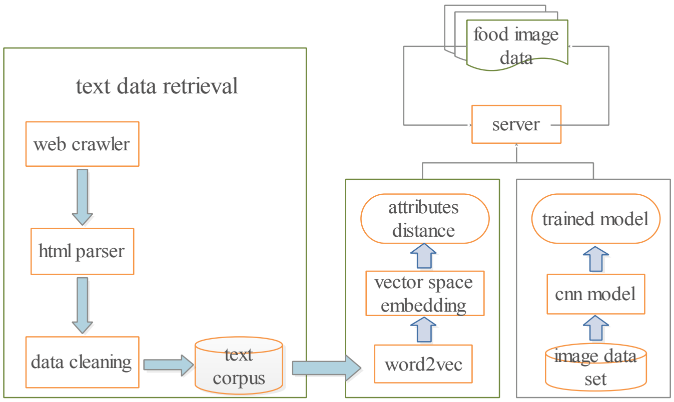

Figure 1. Block diagram of proposed system

Figure 1 explains three modules i-e Text data retrieval, Text data training and Training of our classification model using CNN. After the classification process has been completed by using a pre-trained model, the name of the food is given to the text data retrieval system. The concerned text will be extracted from Google search and through URL extraction. Afterward, HTML will be extracted from webpages and through HTML parser stop words are being removed using different python libraries. After this, it has been passed through the system of Lemmatization and Stemming. All this textual data makes a text corpus which is the input of our next module of text data training. After the retrieval of text corpus, text data will be trained using a tool named as word2vec.

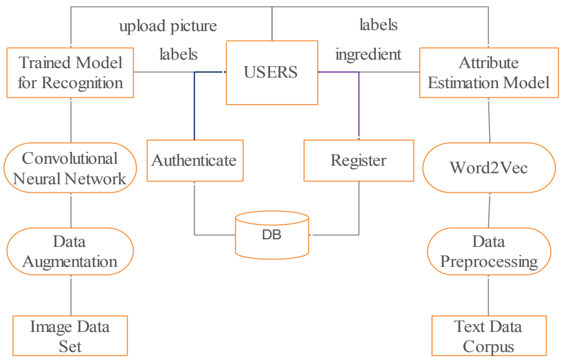

Figure 2. Dataflow diagram of proposed system

Figure 2 explains a CNN model trained on food pictures and on related features. The text data unit is composed of web scrapping of almost more than 500 URLs of each food category. The model also processes text data. After that text data collection word2vec is trained on the corpus to attain vector space embeddings on semantic similarity. For any input of an image, its features and attributes are anticipated by these models. The flow of the proposed system starts with an image and textual data preparation and ends with classification along with attributes and ingredients list. Food images collected from different sources are pre-processed and enhanced by data augmentation. Fine tuning of pre-trained CNN model and training are the next steps. After the successful training and learning phase of the system, there comes the classification phase. Classification is done after taking input from the user. The yielded label from the last step becomes the input of the attribute estimation model. The raw text starts the flow of attribute model creation. After textual data pre-processing, a tool named as “Word2Vec” produces vector embeddings comprising of distance values. These values become the basis for the attribute estimation model to produce related ingredients or attributes.

3.2 System Implementation

As we know explained above that before the classification of pictures that our dataset has, we need to train our system. For creating the environment, we are using the Linux based operating system. Before creating the Anaconda environment, we need to install the Anaconda python distribution, python 2.7 and python 3.6 version. For installing Anaconda, we follow some commands [11]. Once the Anaconda is installed, we will create two environments with python 2.7 and python 3.6 using commands. After the environment is created we will activate the environment using the command. After the environment is activated then we will install the necessary packages of Theano, Pygpu, and Keras inside the environment files [12]. After successful completion of the environment, we will now finally implement the system.

Keras library with Tensor Flow was used for model implementation in python language. Experiments with different CNN models were involved in computing features from approximately 50,000 images. Training time for each model varied but at approximately took 30 hours with varying accuracies. Inception model with its prescribed convolutional and pooling layers classified with an accuracy of 91.73 %. Basically, inception model is pre-trained on ImageNet. Inception model was fine-tuned by removing the last fully connected layer and loading pre-trained weights into the new model which in turn trained last two convolutional blocks. In order to achieve maximum accuracy, we employed techniques like data augmentation, batch normalization, dataset improvements and regularization to combat overfitting. Moreover, multi-crop evaluation, in which 10 crops or so of the image are taken during prediction was also used. Maximum value from predicted results gives the maximum accuracy.

Table 1. Comparison of models in terms of Single and Multiple Crops

Model		Top-1	Top-2	Top-3	Top-5
Inception v3	Single corpus	79.8%	87.9%	91.6%	95%
	Multiple corpus	89.12%	-	-	98.31%
Inception v4	single corpus	83.8%	89.8%	92.4%	94.7%
	multiple corpus	91.73%	-	-	98.56%
V4-101	single corpus	78.3%	85.4%	88.2%	91.2%
	multiple corpus	-	-	-	-

3.3 Results and Evaluation

We selected the CNN model based on Inception-v3 and Inception-v4 as they perform better than other models on our proposed problem domain. These models are fine-tuned on our own created datasets as well as Food-101 datasets in order to perform a comparison.

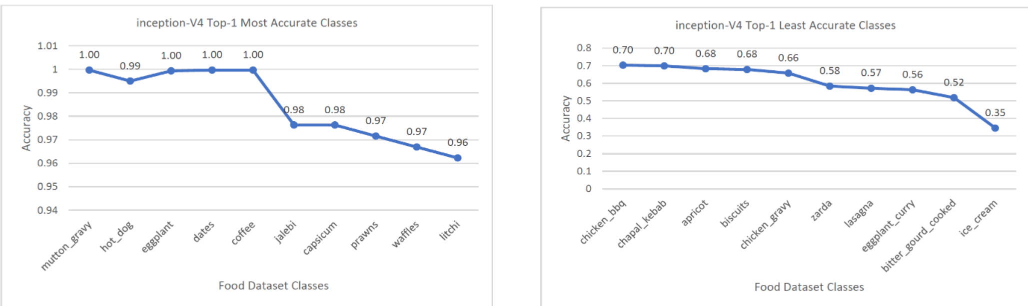

Figure 3. (a) Top-1 most accurate of inception-v4 model; (b) Top-1 least accurate class

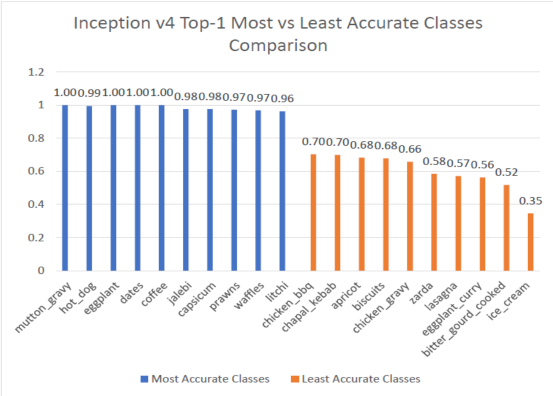

Figure 4. Comparison of top-1 most vs least accurate classes

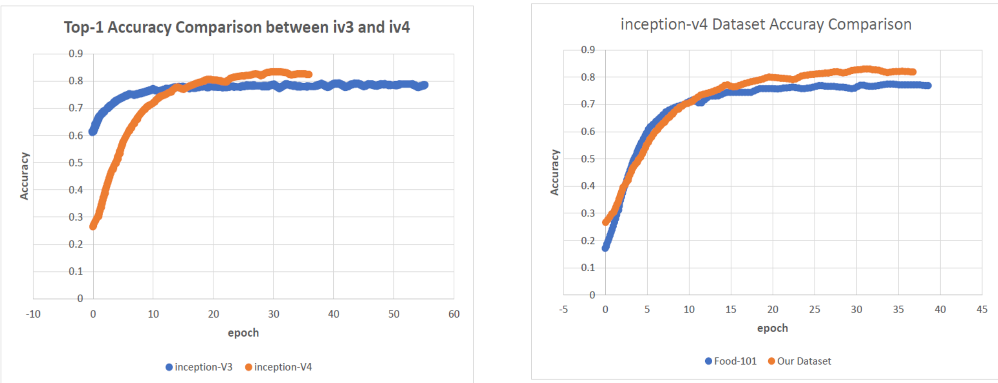

Figure 5. (a) Top-1 accuracy comparison; (b) Inception-v4 model based performance

4. Drawbacks and Improvements

In this section, we will explain the drawbacks current and improvements as future work.

4.1 Recognition and Detection of various food

Current systems do not adequately identify and process mixed physical images. They do not involve cooking foods, liquid foods, and composite foods such as salads and sandwiches. In future research, a mixed food image and a cooking-like physical image are processed by combining image segmentation techniques to solve the phenomenon that the image has oblique edges or each other causes the recognition detection to fail.

4.2 Enhancement of Systems and Datasets

Data sets and features have a great impact on the detection results. Existing data sets are not sufficient and contain limited parameters such as different lighting conditions, camera angles, different backgrounds, etc. In future research, better review techniques [13] should be used to review various types of data sets. In addition, the system and application are optimized architecturally, and a database for storing calculated values, food labels, and other parameters is combined with a faster lookup technique to process the image

4.3 Calories Awareness and Nutrition aware

It is important to understand calorie calculations and their importance. Literature [13] describes the problems in the field based on the small fast food questionnaire, and the literature [14] uses game methods to obtain more informational foods and calorie values. In order to better nurture the basic awareness of calorie calculations between users, it can be extended by introducing new calories to measure nutritional characteristics and combing with deep learning techniques.

5. Conclusion

Currently, fatness is a major issue in human life. Curiosity is found among people to measure their heaviness and healthy eating in order to avoid overweightness. So, this paper presents a novel system that tells us the information about the type of food we eat and its attributes. This system takes the image of the food from the user and after correct classification, the system will tell us about the attributes of the food. A dataset that consists of a common meal of Food-101 and our subcontinental food has been used in our system. We have fine tune the Inception V-3 and V-4 model to recognize the food items and also proposed a method to measure the attributes of the food using the attribute estimation model. The results are enhanced via data augmentation, multi-crop, and similar techniques. Our proposed method for classification as well as for the extraction of attributes achieves a considerably high accuracy of 85 %.

We have also described the possible improvements and the future work to enhance the usability and accuracy of the system.

6. Acknowledge

This work was supported by the National Key R&D Program of China (Grant No.2018YFC1604000/2018YFC1604002).

References

[1] ″wikepedia.” [Online]. Available: https://en.wikipedia.org/wiki/Obesity#cite_note-WHO2015-1.

[2] The American Medical Association Annual Meeting (June 18, “No Title.” [Online].)Available: http://www.ama-assn.org/ama/pub/news/news/2013/2013-06-18-new-ama-policies-annual-meeting.page.

[3] P. Pouladzadeh, P. Kuhad, S. V. B. Peddi, A. Yassine, and S. Shirmohammadi, “Food calorie measurement using deep learning neural network,” in Conference Record - IEEE Instrumentation and Measurement Technology Conference, 2016.

[4] W. Jia, R. Zhao, N. Yao, J. D. Fernstrom, M. H. Fernstrom, R. J. Sclabassi and M. Sun, “A food portion size measurement system for image-based dietary assessment,” in IEEE 35th Annual Northeast conference on Bioengineering, 2009.

[5] ″CNN Stanford University.” [Online]. Available: http://cs231n.github.io/convolutional-networks/.

[6] ″Center of Science in Public Interest.” [Online]. Available: https://cspinet.org/eating-healthy/why-good-nutrition-important.

[7] P. Pouladzadeh, S. Shirmohammadi, and R. Al-Maghrabi, “Measuring calorie and nutrition from food image,” IEEE Trans. Instrum. Meas., 2014.

[8] ″Data Augmentation Keras.” [Online]. Available: https://keras.io/preprocessing/image/.

[9] ″Spatial Transformer Networks.” [Online]. Available: https://arxiv.org/abs/1506.02025.

[10] ″SVM Wikipedia.” [Online]. Available: https://en.wikipedia.org/wiki/Support_vector_machine.

[11] ″Installing Anaconda on Linux.” [Online]. Available: https://www.digitalocean.com/community/tutorials/how-to-install-anaconda-on-ubuntu-18-04-quickstart

[12] ″Installation of Theano, Pygpu and Keras.” [Online]. Available: https://medium.com/@xuweimdm/how-to-install-theano-in-your-anaconda-environment-7fd78769c05f.

[13] P. Pouladzadeh, A. Yassine, and S. Shirmohammadi, “FooDD: Food detection dataset for calorie measurement using food images,” in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2015.

[14] G. M. Farinella, M. Moltisanti, and S. Battiato, “Classifying food images represented as Bag of Textons,” in 2014 IEEE International Conference on Image Processing, ICIP 2014, 2014.