Food Classification Using Deep Learning Algorithm

R. V. Jamnekar, R. R. Keole, S. W. Mohod, T. R. Mahore, and Sagar Pande

Abstract Monitoring of food plays a significant role in leading health-related issues and tasks. With its multiple applications and features, image processing emerges to be an interesting field in the process of identifying food items. In this paper, a technique has been presented for classifying the food image using the You Only Look Once (YOLO) algorithm. Unlike the conventional artificial neural networks, the YOLO algorithm has more efficiency, and it has been trained on a loss function that corresponds straight to detection, and the complete model is trained with 6000 epochs. Due to the high variance in the alike domain of food images, food classification becomes a difficult task but it has a significant role in lives at the present time as it can be utilized by numerous sources. In this paper, a comparison of the working of the YOLO algorithm with other techniques that are used in image processing such as ResNet-50, VGG-16, ImageNet, and Inception has been elaborated. In this work, the famous dataset from Kaggle is used for implementation purposes. The dataset consists of 4000 Indian Food Image 80 different categories or classes. The proposed model is giving 99% accuracy for classifying the food.

Keywords Image processing · Food · Classification · YOLO algorithm · Detection · Image pre-processing · Convolution neural networks

R. V. Jamnekar · S. W. Mohod · T. R. Mahore
Computer Science & Engineering, DRGIT&R, Amravati, Maharashtra, India

R. R. Keole
Information Technology, HVPM, Amravati, Maharashtra, India

S. Pande (✉)
Intelligent System, School of Computer Science and Engineering, LPU, Phagwara, Punjab, India
e-mail: sagarpande30@gmail.com; sagar.pande@vitap.ac.in

Assistant Professor Senior Grade I, School of Computer Science and Engineering VIT-AP, Amaravati Andhra Pradesh, India

1   Introduction

In this present time, people are more considerate about their health and diseases and it compels them to be more conscious about the everyday food and the diet. Not only about the good food and diet but people are considerate about the nutritional values that are contained within a food. Technology has touched almost every aspect of human lives with its efficient applications and techniques and with the exponential growth of technology and evolution of technology, the traditional method to classify food has been replaced with applications that automatically detect the food and recognize their nutritional details from the pictures captured using different machine learning algorithms and computer vision. Applications can automatically scan the diets of individuals and help in numerous aspects.

Overeating is concerning these days because people are overeating and it makes them less active. Considering the busy schedules and stressed lives of people, the importance of proper classification of food is vital and will play a significant role in the lives of people.

Over the past few years, a fair amount of research and development have been carried out in the field of calorie analysis and visual-based diet, and still, the efficient and structured extraction of information from the food clicks remains an exigent issue. Few of the techniques that are currently in use for dietary assessment included manually recording instruments and self-reporting and doing it manually makes it a tedious task to execute. To overcome this, enhancements to the present techniques are a necessity. One of the possible potential solutions to overcome this challenge is the mobile cloud computing system.

In this paper, the YOLO algorithm has been used to classify the food images. The main praise worthy feature of this algorithm is its remarkable speed. It is outstanding when it comes to execution and speed and processes 45 frames per second. With the capability to acknowledge generalized object representation, YOLO algorithms stand to be the best algorithm for detection of objects. The architecture of this algorithm is more like fully convolutional neural network (FCNN). Full images are trained by YOLO, and it precisely optimizes the detection performance.

1.1   Organization of the Paper

Section 2 that is literature survey presents the related works and techniques that are used in classifying the food through the processing of images. Section 3 details the proposed methodology using the YOLO algorithm, and further in the paper Sect. 4, results and analysis have been explained. Conclusion and future scope have been presented in the Sect. 5 of the paper, and Sect. 5 marks the end of the paper.

1.2 Contribution of the Paper

- In dataset used in this work consists of 80 different categories or classes of Indian Food Images consisting of 4000 instances.
- The paper explains the YOLO algorithm in detail along with its use in developing the food classification model.
- Comparison based on efficiency and working methodology has been done among several algorithms and techniques that are used in developing food classification and presented.
- Recent work of many different researchers focusing on food classification has been explained in the literature survey of the paper.
- The paper also provides an insight on the future work that could be done in order to enhance the performance of the model that is presented in the paper.

2 Literature Survey

In this literature survey, multiple papers targeting food classification using image processing techniques and many different algorithms have been reviewed, and the information extracted after reviewing the papers has been mentioned below.

To develop the model for classifying food using the food images, in [1], the dataset that was used contained 101,000 images and 101 categories. To make the system realistic, this dataset was considered. In the dataset, each food category contained 750 clips for training and 250 clips of testing. To train the huge dataset that contains multimedia data, CNN requires high-performance computing machines. After training the system properly, it was able to produce results in an efficient time.

The model proposed in [2] is divided into three contrasting parts. The first part is pre-trained convolutional neural network nodel, the second is dataset preparing and pre-processing phase, and the third and last part is textual data model training. Information's like the type of the food and its attributes such as nutritional value and caloric value is provided by the system proposed in [2]. Image of the food is taken by the system, and the image then is classified. After the classification of the image, the system details the attributes of the food. Further, the result is enhanced utilizing multi-crop, data augmentation, and similar technologies like these. The model proposed in [2] achieved the exactness quite well, and an accuracy rate of 85% was achieved.

According to [3], the dataset that was used for building their system was the publicly available Food 101 dataset which has 100 images of 101 classes. Further, for the classification of these images, SVM was used. Average accuracy was reported after performing fourfold cross validation. In the system that is proposed in [4], although the dataset consisted of 101 classes but only 50 classes were used in the actual work. To store the missing information, BDF and GPCA were used. To extract the feature, LBP and NRLBP were used. They were fed into SVM classifiers for

identifying food images. The accuracy obtained for the proposed model was not mentioned in [5].

In the paper [6], personalized classifiers are expanded on a large scale for daily food image identification in the real world. The architecture of the model comprises a NCM classifier, and the other classifier which is used in the architecture is NN for each user and a model of food distribution which is time independent has been used in order to achieve better performance and exactness in the result.

According to [7], the model has used convolution neural networks to train the dataset, and at the end, the accuracy of 61.4% and top accuracy of 85.2% have been achieved. The dataset used was Food 101 dataset and was trained from scratch. ImageNet weights were used to pre-train the models. The model that outperformed all the other models was pre-trained InceptionV3 model whose top layers were unfrozen in stages.

3 Proposed Methodology

The algorithm used to train the model is the YOLO algorithm. Image processing using the YOLO algorithm is considered uncomplicated and straightforward. The You Only Use Look Once (YOLO) algorithm is capable of training on full images, and it directly optimizes the detection performance. It has numerous benefits over the regular traditional methods. The design of YOLO algorithm permits end-to-end training and real-time speeds and maintains high-average precision. The YOLO algorithm is based on regression, it does not select any particular part of the image it rather predicts the bounding boxes and classes for the full image in a single run of the algorithm.

Instead of searching for the interested regions in the image which is being inputted and could contain an object, YOLO algorithm splits the input image into numerous cells and each cell becomes responsible for prediction of K bounding boxes. YOLO signifies the probability that the cell holds a particular class. The equation for the very same is

$$\text{SCORE}\, m, n = Pm * Mn$$

Probability of presence of an object of certain class 'm'.

YOLO is said to be a clever and convolutional neural network (CNN) and is known for doing object detection in real-time. Single neural networks are applied by YOLO algorithm to the full image, and then, the image is divided into regions and it predicts probabilities for each region and bounding boxes. Predicted probabilities weigh the bounding boxes. The General Yolo-based detection system is depicted in Fig. 1.

The dataset that is used in the following model is a self-prepared database, and it consists of 4000 different images and 80 different types of food as depicted in Fig. 2. Sample images used for training the model are depicted in Fig. 3.
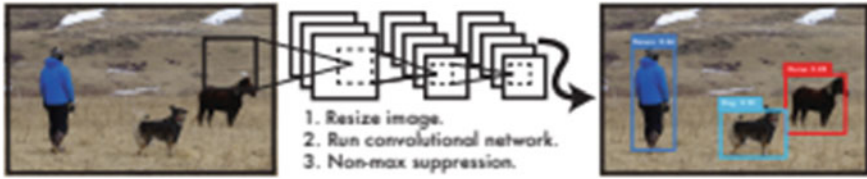
Fig. 1 YOLO detection system [4]

Fig. 2 Names of 80 different classes available in the dataset
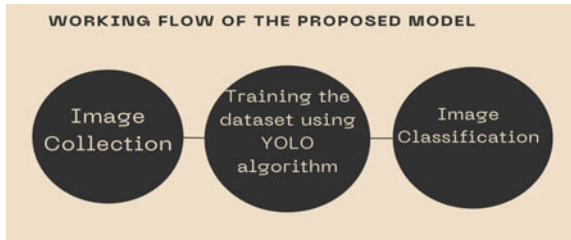

Fig. 3 Sample images from the dataset [8]

The model that is proposed for food classification in this work is depicted in Fig. 4 and elaborated as follows:

- 4000 images of forty different types of food were captured.

Fig. 4 Steps involved in developing the model

Table 1 Comparative
analysis of the results
obtained using various
techniques

Classifiers	Accuracy %
ResNet-50	78.1
VGG-16	67.2
ImageNet	58.4
Inception	77.2
YOLO	99.1

- After using the dataset that contained 4 k different images of food, image encoding was applied for all 80 classes for which LabelImg tool was used.
- Dataset was divided in 70–30 ratio for training and testing the model.
- Finally, to train the model, YOLO algorithm was used with 6000 epochs.

4 Result and Analysis

The proposed system in the paper yields a noticeable accuracy rate of 99%. The YOLO algorithm which is used to train the model uses a totally different approach. The extremely fast speed of the algorithm makes it more popular, and the additional benefit that comes along with this algorithm is its capability to run in real-time.

The proposed model has innumerable benefits over other methods and that is as follows:

- The YOLO algorithm is extremely fast.
- It looks uses the encoding of the image for training and testing.
- It is comparatively easier in implementation.
- It outperforms various other detection methods.

The above-mentioned attributes of the YOLO algorithm used in training the model helped in achieving a decent and remarkable accuracy rate. The results obtained after using various techniques which were used for training the models are given in Table 1. Algorithms like ResNet-50, VGG-16, ImageNet, Inception, and YOLO were used for training the model. The dataset was divided into 70–30 ratio for training and testing purpose, respectively, as depicted in Fig. 5. For generating efficient training and testing results of the proposed model high-configuration architecture consisting of AMD RYZEN 9 4000 Series processor, 64-bit Windows 10 Operating System, and 32 GB of RAM, NVIDIA GeForce GTX 960 was used.

5 Conclusion and Future Scope

With the emerging need of classification of food based on their nutrition and various other parameters, traditional methods prove to be extremely inefficient and

Fig. 5 Dataset specifications after splitting the data
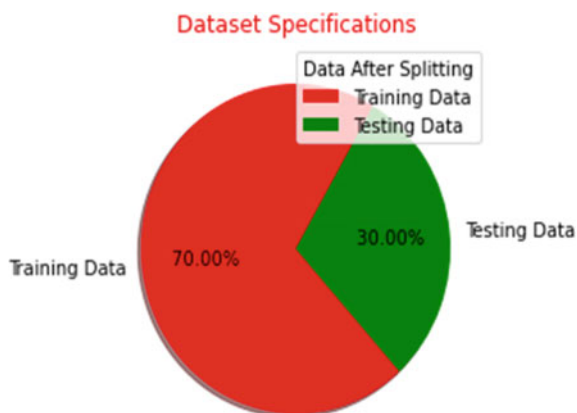
a time taking process. With the evolution of technology over a long period of time, researchers have found various methods to classify food in a more efficient way. In this paper, the YOLO algorithm has been explained in detail, and the paper also details the methodology to build a system that classifies food using image processing techniques which uses the YOLO algorithm to train the models. The system proposed in this paper gives a remarkable accuracy rate of 99%. Numerous papers that focus on the image processing technologies have been reviewed and elaborated in the literature survey of the paper. We can achieve more accuracy if the dataset is precise and contains more unique images of types of food. Impurity of the training images needs to be removed for getting more enhanced performance. Also, the number of epochs used for training the model can be improvised for getting more promising results.

References

1. Attokaren DJ, Fernandes IG, Sriram A, Murthy YS, Koolagudi SG (2017) Food classification from images using convolutional neural networks. In: TENCON 2017-2017 IEEE region 10 conference, pp 2801–2806. https://doi.org/10.1109/TENCON.2017.8228338
2. Shen Z, Shehzad A, Chen S, Sun H, Liu J (2020) Machine learning based approach on food recognition and nutrition estimation. Procedia Computer Science 174:448–453, ISSN 1877–0509. https://doi.org/10.1016/j.procs.2020.06.113
3. Inunganbi S, Seal A, Khanna P (2018) Classification of food images through interactive image segmentation. In: Nguyen N, Hoang D, Hong TP, Pham H, Trawiński B (eds) Intelligent information and database systems. ACIIDS 2018. Lecture notes in computer science, vol 10752. Springer, Cham. https://doi.org/10.1007/978-3-319-75420-8_49
4. Redmon J, Divvala S, Girshick R, Farhadi A (2016) You only look once: unified, real-time object detection. IEEE Conf Comput Vision Pattern Recogn (CVPR) 2016:779–788. https://doi.org/10.1109/CVPR.2016.91
5. Pimple KM, Likhitkar PP, Pande S (2022) Convolutional neural networks for malaria image classification. In: Gupta D, Polkowski Z, Khanna A, Bhattacharyya S, Castillo O (eds) Proceedings of data analytics and management. Lecture notes on data engineering and communications technologies, vol 91. Springer, Singapore. https://doi.org/10.1007/978-981-16-6285-0_37

6. Yu Q, Anzawa M, Amano S, Ogawa M, Aizawa K (2018) Food image recognition by personalized classifier. In: 2018 25th IEEE international conference on image processing (ICIP), pp 171–175. https://doi.org/10.1109/ICIP.2018.8451422
7. Islam MT, Karim BMNS, Rahman S, Jabid T (2018) Food image classification with convolutional neural network. In: 2018 international conference on intelligent informatics and biomedical sciences (ICIIBMS), pp 257–262. https://doi.org/10.1109/ICIIBMS.2018.8550005
8. https://www.kaggle.com/iamsouravbanerjee/indian-food-images-dataset