Overview of Deep Learning in Food Image Classification for Dietary Assessment System

Conference Paper · December 2021

DOI: 10.1007/978-981-16-0730-1_18

Bhoomi Shah
The Maharaja Sayajirao University of Baroda

Hetal Bhavsar
The Maharaja Sayajirao University of Baroda

Overview of Deep Learning in Food Image Classification for Dietary Assessment System

Bhoomi Shah and Hetal Bhavsar

Abstract In modern days, people are very attentive about their health and diet. High-calorie intake food can be harmful and may lead to serious health condition. Food image identification plays a very important role in today's era. Food domain can be divided into two parts. First is to recognize food items, and second is to estimate the calorie. Accurate methods for food identification and calorie estimation can help people to fight against obesity which is the cause of overweight. So, recognition of food is first step for a successful healthy diet. Classification of food images is a very challenging task as the datasets of food images are not linear varying due to the large variations in food shape, volume, texture, color, and compositions. To recognize food items accurately, image processing can be used. Image processing techniques include image preprocessing, image segmentation, feature extraction, and image classification for recognition of food objects. Deep learning is an active research field nowadays in the field of object recognition, natural language processing, speech recognition, and many more. This paper defines the role of deep learning techniques based on a convolutional neural network for food object recognition. Many papers have been studied that have used deep learning as a tool for food object recognition and calorie estimation and achieved impressive results. This encourage us to do a detailed study in the field of food domain through deep learning. This paper analyzes, a deep learning framework, services, food items datasets, methods of segmentation and classification, and various food recognition techniques. Every method has its pros and cons as well. The main idea of this survey is to carry out a detailed study of current food item recognition techniques through deep learning. The result achieved by deep learning for recognition of food images will attract more researcher to put efforts in the field of food domain through deep learning in the future.

Keywords Computer vision · Convolutional neural network · Deep learning · Food image recognition · Object detection

---

B. Shah · H. Bhavsar (✉)
Department of Computer Science and Engineering, The Maharaja Sayajirao University of Baroda, Vadodara, Gujarat, India
e-mail: hetal.bhavsar-cse@msubaroda.ac.in

M. Paprzycki et al. (eds.), *Intelligent Systems, Technologies and Applications*,
Advances in Intelligent Systems and Computing 1353,
https://doi.org/10.1007/978-981-16-0730-1_18

1 Introduction

Food is an essential element of everyone's life. Obesity is increasing at a higher rate day by day risking the lives of many people. It is necessary to control the amount of calorie intake to prevent obesity and many other diseases. Also, analyzing food images and calorie estimation can help people to follow healthy food diet. It can also be useful for normal people to maintain their day-to-day diet [1].

In today's world when people are more conscious about their diet, food recognition and calorie estimation has gained popularity. This can be achieved in two steps. First is to recognize food image successfully, and in second step, calorie can be estimated. This paper only focuses on various approaches developed using different technologies for food image recognition to get maximum possible accuracy over a wide range of food items and cuisines.

The contents of this paper are divided in following order. Section 2 discusses the challenges in food object recognition. Section 3 gives introduction to deep learning, review of deep learning networks and review of available frameworks, and techniques in deep learning for object identification. Section 4 discusses all the datasets for food image recognition. Section 5 compares different segmentation and classification techniques used for food image recognition. Finally, Sect. 6 concludes this article and gives future direction in this area.

2 Challenges in Food Image Recognition

Classification of food images is a very challenging task as the dataset of food images are not linear. There are so many varieties of food, so it is very difficult to make such a system that can identify each type of food. Some of the challenges faced while designing such a system has been shown in Fig. 1 and can be discussed as under [2–8].

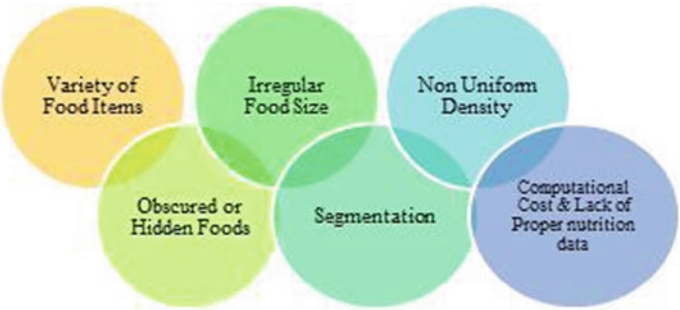

Fig. 1 Challenges for food image recognition

Fig. 2 Variety of food items

2.1 Variety of Food Items

There are so many varieties of the same food item that can be varied by different cooking method, camera quality, illumination and presentation. The varieties of rice have been shown in Fig. 2.

2.2 Obscured or Hidden Food Item

As shown in Fig. 3, here all dry fruits are not visible inside kheer and it is hard to identify the vegetables in rice.

Fig. 3 Obscured or hidden food item

Fig. 4 Irregular food size

2.3 Irregular Food Size

Food items do not have regular shapes in general. For example, the volume of a burger using a known geometric formula from a single image as shown in Fig. 4 cannot be estimated.

2.4 Segmentation

It is sometimes impossible to separate food items from the background due to the nonuniformity of the background. If the system uses a reference object for camera calibration, then segmentation is needed for that object too.

2.5 Liquid Food

Liquid food items do not have a fixed shape. To estimate liquid volume, one need to know the shape of the container holding the liquid and depth of liquid in that container.

2.6 Non-uniform Density

For image-based calorie estimation, it is assumed that calorie in a unit volume of food is constant. But, in real life, the density of food items can vary significantly. This variation in density can lead to some errors in calorie estimation.

2.7 Computation Cost

If automatic calorie estimation system to be deployed on a client-side device and produce output quickly, the computation cost should lie within a reasonable limit. For example, it is possible to improve the accuracy of volume estimation using 3D volume reconstruction from multiple images. In this case, computation cost can be very high.

2.8 Lack of Proper Nutrition Data

For calorie estimation from a portion of food, a large database covers every kind of food item. Since food items can be prepared in various ways, following different recipes are difficult to prepared such database. Even calorie content varies for the same food item with the different recipes. For example, calorie of a fried chicken differs from the calorie of a grilled chicken. So, this is a big challenge to accurately measure calories.

To deal with these challenges, deep learning is a recent advancement in the field of food domain and has achieved remarkable performance. The main two criteria needed for using deep learning are a large number of labeled data and a high processing machine, easily available nowadays. The main reason for the succession of deep learning is the ability to learn from extracted features directly from the images.

3 Overview of Deep Learning

Deep learning is an advanced technology for image processing, speech recognition, object detection, and food science and engineering [1, 9–12]. It works with artificial neural networks, which are designed to imitate how humans think and learn. Deep learning is a subset of machine learning. Unlike machine learning, in deep learning, basic details about the data need to be given that process through many layers and the computer trains to recognize the patterns by its own. Deep learning techniques are very successful, the reason is, availability of large number of datasets and the availability of high processing GPU [12]. We surveyed lots of recent articles in the food domain including food recognition and calories estimation. In this paper, we investigate the specific problems related to different types of food, the different food datasets used, and the preprocessing methods. Also, different networks available and its comparison, various deep learning frameworks, the segmentation and classification methods, the performance achieved in terms of Top-1 and Top-5 accuracy, and the comparison of all the CNN architecture with its advantages and disadvantages were studied. Firstly, it is very important to understand the types of learning techniques. Table 1 describes different categories of learning techniques with its definition, subcategories, methods and approach used, on which dataset it works, support which model and helpful in which situation has been defined [1, 13].

Machine learning techniques can be divided into mainly three categories.

1. Supervised learning
2. Unsupervised learning
3. Reinforcement learning

This research work is concentrated mainly on supervised learning.

Table 1 Supervised Vs. unsupervised learning Vs. reinforcement learning

Differ by	Supervised Learning	Unsupervised Learning	Reinforcement learning
Definition	The machine learns patterns by using labeled data	The machine is trained by Unlabeled data	No training dataset, so bound to learn from its experience
Data	Labeled	Unlabeled	No predefine data
Dataset	Use training dataset	Use just input dataset	Not predefine
Uses of	Prediction	Analysis	Consequence or behavior
Divided into	Classification and regression	Clustering, density estimation and dimensionality reduction	Reward-based
Approach	Maps the labeled input to the known output	Understand patterns and discovered output	Follows the trial and error methods
Models/networks	CNN, RNN, RvNN	DBN, DBM, GAN, VAE	Markov Decision process learning
Helpful in areas	Image Recognition, speech recognition, forecasting	Pre-process the data, pre-train supervised learning algorithms	Warehouses, Inventory management, aircraft control, robot motion control, chess game

3.1 Deep Learning Networks

Deep learning networks are the mathematical models that work like human brains. This mathematical model is created in form of neural network that consists of neurons. The neural network is divided into three major layers that are input layer (first layer of neural network), hidden layer (all the middle layer of neural network), and the output layer (last layer of the neural network.). The most popular deep learning networks for supervised learning can be described as follows.

3.1.1 CNN: Convolutional Neural Network

CNN is the main category to do the image recognition and image classification. CNN takes an input, process it, and classifies it under certain categories. The main advantages of CNN are parameter sharing, sparse interactions, and equivalent representations [9, 12–15].

A layer in the neural network is nothing but a collection of neurons that takes an input and provides an output. The input of each of these neurons is processed through the activation function assigned to the neurons.

3.1.2 RNN: Recurrent Neural Network

It is a very popular deep learning model that uses recursion techniques to build models. RNN saves the output of current layer which will be input to the next layer. It can memorize previous inputs due to its internal memory, because of this property it is especially used as language model. It is mostly used in natural language processing and speech processing [13, 18], text analysis and machine translation [16, 17].

3.1.3 RvNN: Recursive Neural Network

RvNN can handle the inputs of different modalities [13]. RvNN has been especially successful in NLP.

RvNN separate the image into different segment and form a syntactic tree [19].

For image classification, most of the research work has been used CNN and RNN [9–17]. Table 2 shows how convolutional neural network (CNN) differs from recursive neural network (RNN).

Many successful research works that have been done on food object recognition through CNN proved that CNN gives the best result in terms of accuracy and error rate for object recognition [9–17]. This research work is varied by the type of CNN, the dataset, segmentation, and classification techniques used.

Table 2 Comparison between CNN and RNN

CNN	RNN
Suitable for spatial data such as an image. Popularly used for image	Suitable for temporal data also called sequential data. Widely used for
Considered more powerful than RNN	Less powerful than CNN
Many research works have been done using CNN for image classification	It can be used for image classification, but theoretically, only a few researchers about RNN image classifier can be found
CNNs are special for video processing and image processing	RNN works primarily on time series information on the past influence of the customer
The interconnection consumes a finite set of input and generates a finite set of outputs according to inputs	RNN can allow arbitrary input length and output length

Table 3 briefly describes types of CNN with its main features, error rate, the number of parameters used, input size, number of convolution layers and stride based on recently published papers [20, 21].

Table 3 defines the most popular types of CNN. Many researchers have successfully experiment on GoogLeNet and ResNet model [27–29]. According to research, these two networks give best classification accuracy when used with food object recognition [27–29].

3.2 Deep Learning Frameworks

Deep learning frameworks is an interface or a tool which is easy to understand and combines deep learning algorithms, pre-built models and optimized components. Instead of writing hundreds of lines of code, deep learning framework builds a model quickly which provides good community support and parallelize the process to reduce computations [30–44].

Table 4 discusses and compares some of the famous deep learning frameworks like Torch, Theano, Tensorflow, Chainer, Keras, Apache Singa, MXNet, Caffe, Microsoft Cognitive Toolkit CNTK, Deep learning 4j, Neon.

The frameworks are studied from the date when the first and stable version was released, language supported, operating system supported, type of library support and the support for graphics processing unit (GPU), central processing unit (CPU) or tensor processing unit (TPU). The list of deep learning framework is very exhaustive. Every few months, new deep learning frameworks are introduced. Table 4 contains an analysis of the famous deep learning networks introduced till 2018. All the frameworks listed in Table4 support both the CNN and RNN models.

Table 3 Classification of CNN

Year	CNN	Main features	Top 5 error rate %	No. of parameters	Input size	No. of convolution layer	Stride
1998	LeNet [22]	First popular CNN architecture originally trained to classify handwritten digits	NA	0.060 M	28 × 28	2	1
2012	Alex Net [23]	Winner of ImageNet ILSVRC-12, it is deeper as compared to LeNet	15.3	60 M	227 × 227	5	1.4
2014	GoogLeNet/Inceprion n/w [24]	Winner of ILSVRC2014 competition. It has introduced block concept	6.67	4 M	224 × 224	21	1.2
2014	VGG Net [25]	Runner up of the ILSVRC-2014 competition. Homogeneous topology Uses small size kernels	7.3	138 M	224 × 224	16	1
2015	ResNet [26]	Relu is used identify mapping-based skip connections and implement heavy batch normalization	3.6	25.6 M, 1.7 M			

4 Dataset

A strong collection of images, Dataset, is the key element to achieve best classification accuracy [45]. A dataset is created considering the number of food classes and the type of food. Table 5 describes the number of available datasets with the food content including the number of food classes and number of food images. Since

Table 4 Deep learning frameworks

Name	Initial	Stable release	Language supported	Operating system supported	Type of library support	Support for GPU/CPU/TPU
Torch [30]	2002	2017	Lua, LuaJIT, C, CUDA, C++	Linux, Android, MacOS X iOS	Deep learning and machine learning	CPU and GPU
Theano [31, 32]	2007	2019	Python,CUDA	Linux, MacOS, Windows	Machine learning	CPU and GPU
Tensorflow [33, 34]	2015	2020	Python, C++, CUDA	Linux, MacOS, Windows, Android	Machine learning	CPU and GPU and optimize for TPU
Chainer [35]	2015	2019	Python	–	Deep learning	Best for GPU
Keras [36]	2015	2019	Python	Ios, Android	Deep learning and machine learning	GPU and TPU
Apache Singa [37]	2015	2020	C++, Python, Java	Linux, MacOS, Windows, Android	Machine Learning	CPU and GPU
MXNet [38, 39]	–	2020	C, Python, R, Java, Julia, Javascript Scala Go Perl	Linux, MacOS, Windows	Deep learning and machine learning	GPU
Caffee [40, 41]	–	2017	C++	Linux, MacOS, Windows	Deep learning	CPU and GPU
Microsoft cognitive ToolKit CNTK [42]	2016	2019	C++	Windows, Linux	Deep learning and machine learning	GPU
Deeplearning4j	–	2019	Java, CUDA, C, C++	Linux, MacOS, Windows, Android	Deep learning and machine learning	CPU and GPU
Neon [44]	–	2018	Python	–	Deep learning	–

Table 5 Summary of Food image databases

Name of dataset	Year	#images	#Food classes	Food content
Food85 [2]	2010	8500	85	Japanese food
Chen [3]	2012	5000	50	Chinese food
UNIMIB2016 [4]	2016	1027	73	Pictures captured by a digital camera in dining hall
Food524DB [5]	2017	247,636	524	Merging food classes from existing database vireo, food-101, food50 and modified version of UECFOOD256
FFOcat [6]	2018	58,962	156	Selected food images Downloaded from Web
Foodx-251[7]	2019	158,000	251	Selected food items like cakes, pasta, soups, etc
FooDD [8]	2015	3000	30	Single and mixed food images including fruits
PFID [46]	2009	1098	61	Fast food items from USA
TADA [47]	2009	256 + 50 replica	–	Common food in USA
Food50 [48]	2009	5000	50	Japanese Food
UEC-Food100 [49]	2012	9060	100	Popular Japanese Food
Food101 [50]	2014	101,000	101	Popular food in USA
UECFood256 [51]	2014	31,397	256	Famous foods in Japan & other Countries
UNICT-FD889 [52]	2014	3583	899	Different Nationality dishes like Italian, Thai., Japanese, etc
Diabetes2 [53]	2014	4868	11	Selected food
New Dataset [54]	2014	5000	11	Central European Food
Menu-match [55]	2015	646	41	Food from three restaurants (Asian, Italian, Soup of 10 types)
VIREOFood-172 [56]	2016	110,241	172	Popular Chinese dishes from "go cooking" and web
ChineseFoodNet [57]	2017	185,000	280	Food images either taken from real dishes or recipe pictures or selfies

deep learning is data hungry, a large collection of food images is required to train a food-classification model. Food image datasets vary in many aspects such as, a single food image, mixed food image, non-mixed food, several food groups, liquid food image, type of cuisine and total images per food class. Table 5 summarizes of different food image databases with their respective features.

Other than these, Anthimopoulos et al. [54] used one visual database created with 5000 food images and organized into 11 classes reflecting the nutritional habits in central Europe in 2014. Also, in 2017, Paritosh et al. [58] proposed work on the Indian food dataset containing 100 Indian food images of 50 different classes. It is the first Indian food dataset that is available to download.

5 Segmentation and Classification

5.1 Segmentation Techniques

In computer vision, image segmentation is the process of partitioning a digital image into multiple segments (sets of pixels, also known as super pixels). The goal of segmentation is to simplify and/or change the representation of an image into something more meaningful and easier to analyze. Image segmentation is the process of assigning a label to every pixel in an image such that pixels with the same label share certain characteristics. Segmentation means to divide parts of a scene [59, 60]. Segmentation is the last step before food classification where the identification of various food images takes place.

In segmentation, the image is partitioned in equal segments. After segmentation, the boundary detection of irregular food portions becomes easy, and it gives better detection of the food portion. Matsuda et al. [61] discussed to separate the food images from the background regardless of the lighting conditions or if the food is mixed or not. According to Yang et al. [62], it is very difficult to separate those food images which does not contain any specific attribute. Ramadevi et al. [63] showed a synergy between segmentation and object recognition is done using the EM algorithm, OSTU, and genetic algorithm. They have also discussed the difference between region-based and edge-based segmentation. There are different types of segmentation techniques, and it can be divided mainly into two categories: layer-based segmentation methods and block-based segmentation methods. Block-based segmentation can be partitioned into region-based segmentation and edge-based segmentation. Also, other types are segmentation based on weakly supervised learning in CNN, threshold segmentation methods, segmentation based on clustering, etc. Yan Hao [64] have evaluated and recap different image segmentation algorithms and compared them with pros and cons. A similar approach has been used by Shakuntala and Surendra [65], after studying different types of image segmentation, they have concluded that "all the works done in the field of image segmentation are needed to be monitored manually there is no such method which can detect the objects with precision and without any database, which obviously takes time to get build." Another similar approach has also been found by Vairaprakash and Subbu [66]. W. Liming et al. [67] used an object detection method which is a combination of top-down recognition with bottom-up image segmentation has been developed. On the other hand, Wataru and Keiji [68] have proposed neither a traditional proposal approach nor the fully convolutional approach, but a middle approach. The approach combined fully convolutional networks and back-propagation-based approach. However, Bryan et al. [69] computed multiple segmentations of images and then learned the object classes to choose the correct segmentation. A similar approach can also be found by Zhu et al. [70], They have combined two concepts: A set of segmented objects can be partitioned into classes based on global and local features; and perceptually, similar object classes can be used to assess the accuracy of image segmentation. They have shown improvement in the accuracy of segmenting food images using a segmentation compared to normalized cut without classifier feedback when there is no prior information about the scene. This idea has improved the overall accuracy of classification. Several methods for segmentation have been studied with the approach used, and performance of each has been summarized in Table 6.

Apart from this, Sandhya et al. [76] used different approaches like traditional methods and modern deep learning techniques for food image recognition which have been discussed, and the proposed work has been implemented on Indian cuisine, whereas Adnan et al. [77] discussed different calorie estimation techniques with pros and cons. Zhao et al. [78] proposed an algorithm on deep learning which can correctly identify food and analyze nutrition. Doyen et al. [79] developed a deep learning-based system known as "FoodAI" for recognition of food smartly, and the system has been tasted on commonly consumed food in Singapore. Ashutosh et al. [29] experiment classification of food or non-food item based on deep convolution neural network. They have used GoogLeNet model and achieved 99.2% classification and 83.6% in recognition of food. Simon and Barbara [80] proposed a novel approach, based on deep convolution neural network for food and drink image recognition, known as "NutriNet," and Achieved 86.72% classification accuracy. Md et al. [81] also used convolution neural network for classifying food images and achieved 92.86% accuracy.

Since there is no general solution to the image segmentation problem, these techniques often have to be combined with domain knowledge in order to effectively solve an image segmentation problem for a problem domain [11, 82].

5.2 Image Classification Techniques

Food image recognition can be achieved with machine learning-based approach or deep learning-based approach [10]. Recently, deep learning has become very famous as it gives impressive result due to the feature that it can learn features automatically just like a human does [83]. It is probably the best approach in the cases when we do not have enough pre-defined features. As this research work focuses on a deep learning approach, this paper discusses all of the work done till now for food image recognition through deep learning. Detailed studies have been shown in Table 7.

This section studied all the recent articles on food image classification. Many research works have been done on food image recognition that has used convolution neural network with different combined approaches and different datasets as shown

Table 6 Segmentation methods

Reference	Year	Approach	Performance
[15]	2015	Have proposed a DCNN-based food image segmentation which requires no pixel-wise annotation	It has been found that the proposed method gives better food region detection as compared to other RCNN
[61]	2012	Have used mixed techniques such as whole image, circle detector and JSEG segmentation	Top 1 accuracy is 21%, and top 5 is 45%
[71]	2017	Introduced a new efficient image segmentation method which consider various feature vectors such as color, texture, etc.	It has been found that the suggested method is an innovative segmentation method that needs No other information about images 70% of tested images gives better result as compared to ACT algorithm
[72]	2009	Have suggested a region-based approach which is a combination of object detection and image segmentation	Have proposed a modular energy function that is easy to evaluate and enhances computer vision technologies. The author has also proposed a number of future works
[73]	2016	Have collaborated image segmentation with machine learning techniques	The problem can be best solved with the ensemble model of random forest, ANOVA RBF SVM, and generalized linear model, giving the accuracy of 93.11111%
[74]	2013	Have proposeda mobile food recognition system which uses methods such as Grub Cut, SURF based, bag-of-features, SVM and fast χ^2 kernel	The suggested method gives 81.55% of top five accuracy
[75]	2005	Have suggested a segmentation algorithm that can be used for color food images having background	The assessed segmentation performance computed from the area under the receiver operation characteristic (ROC) curve was 0.9982

in Table 7. It has been observed from the studies that convolution neural network is a hot topic in food object recognition as it reduced the complex problem of food image recognition. Many researchers have tried to combined method based on convolution neural network with architectures such as Inception and GoogLeNet. It has also been proved that classification accuracy has been increased as the model gets deeper.

Table 7 Food classification

Reference	Year	Approach	Dataset	Top 1	Top 5
[27]	2016	Used deep learning approach for the classification and fine-tuned the image recognition architecture Inception	UEC Food-100	81.5	97.3
			UEC-Food-256	76.2	92.6
			Food-101	88.3	96.9
[28]	2016	Proposed a food recognition method based on CNN, and the architecture is based on Inception-ResNet and InceptionV3 model	Food-101	72.55%	91.31%
[54]	2014	Two ANN models were used. One is without hidden layer and another is with one hidden layer. Classification was done using three supervised methods SVM, ANN, and random forests (RF)	Diabetes	75.0%	–
[58]	2017	Constructed a multilayered deep convolutional neural network (CNN) architecture, also proposed new Indian Food dataset	Indian Food database	73.50	94.40
			ETH Food-101	72.12	91.61
[59]	2016	Proposed deep architecture, namely Arch-D, that defines relationship between food and ingredients label through multitask learning	UEC-Food-100	82.1	97.3
			VIREO	82.1	95.9
[84]	2014	Used the approach called random forest discriminant components (RFDC) and compared it with various other methods, also have introduced Food-101 dataset having 101,000 images	Food-101	56.4%	–
[85]	2015	Used deep convolutional neural network for food photograph recognition task in the ImageNet and have implemented this combination on Twitter photograph data. Achieved high level of accuracy proving DCNN gives best result on large-scale image data	UECFood-100	78.7%	–
			UEC-Food-256	67.57	89.0

(continued)

Table 7 (continued)

Reference	Year	Approach	Dataset	Top 1	Top 5
[86]	2016	Proposed new algorithm based on CNN and achieved impressive result on two datasets, namely Food-101 and UEC-FOOD-256	Food-101	77.4%	93.7%
			UEC-Food-256	54.7	81.5
[87]	2015	food size and labels ad apply this method to a dataset of images from 23 different restaurants	Food-101	79.0%	–
			Food201 segmented	76.0	–
			Menu-Match	81.4	–
[88]	2015	Used CNN-based classifier to estimate, used a deep convolution neural network with sic layers, used to classify food image patches. Experiments have achieved attractive result	Own database with 573 food items	84.90	–
[89]	2016	Used the graph cut method and uses deep convolution neural network to classify food images and have achieved remarkable accuracy for a single food image	Own database with 10,000 high-resolution images	99.0	–
[90]	2014	Used deep convolutional neural network with Fisher vector with HOG and color patches	UEC-Food-100	72.3	92.0
[91]	2016	Proposed a new dataset on which food recognition algorithms can be tasted	UNIMINB2016	78.3	–
[92]	2019	Developed a model with five-layer CNN, and first ever combining bag-of-features model with support vector machine to achieve high level of accuracy	ImageNet	74	–
[93]	2017	Gives overview on different approaches of food detection and model used for comparison is based on GoogLeNet architecture, principle component analysis and support vector machine	FCD	98.81%	–
			Ragusa DS	95.78%	

(continued)

Table 7 (continued)

Reference	Year	Approach	Dataset	Top 1	Top 5
[94]	2017	combination of multiple classifiers based on different convolutional models that complement each other hence improved performance	Food-101 Food-11	–	–
[95]	2018	Efficiently combined visual content, context and external knowledge and tested on different datasets	Different dataset	–	–
[96]	2018	Proposed a segmentation algorithm based on random forest and has used boundary detection and filling methods. Also, compared the proposed algorithm with three existing methods	Food-101	90.5	–

6 Conclusion and Future Work

This paper mainly focuses on how deep learning techniques can be used for food object recognition. It has been surveyed a wide range of articles designed for automated food image recognition. The paper covers popular networks for deep learning, mostly used frameworks, famous services available for object detection and Standard datasets. Different segmentation and classification techniques have also been discussed. The challenges in food object recognition have also been studied. According to the authors' knowledge, this is the first article that has fully focused on food object recognition through deep learning and has studied every phase of food object recognition. Also, it has been cleared from the challenges and studies that there is no such model exists that can recognize each and every type of food. Every model has its own limitation due to the nonlinear nature of food. It has been concluded that food image recognition is a hot topic in computer vision, and the use of convolution neural network has improved the result accuracy of food image recognition. It has been observed that there are variety of food domains that has not been touched yet as not much work done on recognition of liquid food items, and also, there is very less work done till now for European food and Indian cuisine. In fact, there are very limited number of dataset which exist for Indian food and European food. So, in the future, more work in this area needs to be focused. The result achieved by deep learning for recognition of food images will attract more researchers to put efforts in the field of food domain through deep learning in the future.

References

1. G. William, Y. Wei, A survey of deep learning: platforms, applications and emerging research trends. Special section on human-centered smart systems and technologies. IEEE Access **6** (2018)
2. H. Hajime, J. Taichi, Y. Keiji, Image recognition of 85 food categories by feature fusion, in *IEEE International Symposium on Multimedia* (2010), pp. 296–330
3. M. Chen, H. Yung, J. Chia, H. Shih, M. Shane, C. Eugene, H. Che, O. Ming, Automatic Chinese food identification and quantity estimation, in *SIGGRAPH 2012*, vol. 29 (Springer, Asia, 2012), pp. 1–4
4. C. Gianluigi, N. Paolo, S. Raimondo, Food recognition: a new dataset, experiments, and results. IEEE J. Biomed. Health Inform. **21**(3), 588–598 (2017)
5. C. Gianluigi, N. Paolo, S. Raimondo, Learning CNN-based features for retrieval of food images, in *ICIAP Workshops* (2017), pp. 426–434
6. D. Ivan, D. Mauro, Ontology-driven food category classification in images, in ICIAP (2019), pp. 607–617
7. K. Parneet, S. Karan, W. Weijun, B. Serge, D. Ajay, foodx-251: a dataset for fine-grained food classification, in *Computer Vision and Pattern Recognition* (CVPR19) (2019)
8. P. Parisa, Y. Abdulsalam, Foodd: food detection dataset for calorie measurement using food images, new trends in image analysis and processing, in *ICIAP Workshops* (2015), pp. 441–448
9. P. Ajeet, P. Manjusha, R. Siddharth, Application of deep learning for object detection, in *International Conference on Computational Intelligence and Data Science* (Elsevier, Amsterdam, 2018)
10. V. Bruno, S. Resende, C. Juan, A survey on automated food monitoring and dietary management systems. J. Health Med. Inform. (2017)
11. D. Li, A tutorial survey of architectures, algorithms,and applications for deep learning. APSIPA Trans. Signal Inform. Process. (2014)
12. Z. Lei, Z. Chu, L. Fei, Q. Zhengjun, H. Yong, Application of deep learning in food: a review. Comprehensive Rev. Food Sci. Food Safety (2019)
13. P. Samira, T. Haiman, A survey on deep learning: algorithms, techniques and applications. ACM Comput. Surv. **51**(5) (2018)
14. G. Ian, B. Yoshua, C. Aaron, *Deep Learning* (MIT Press, Cambridge, 2016)
15. S. Wataru, Y. Keiji, CNN-based food image segmentation without pixel-wise annotation, in *ICIAP* (Springer, Berlin, 2015), pp. 449–457
16. L. Pengfei, Q. Xipeng, H. Xuanjing, Recurrent neural network for text classification with multi-task learning, in *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence* (2016)
17. S. Ilya, M. James, H. Geoffrey, Generating text with recurrent neural networks, in *Proceedings of the 28th Inter-national Conference on International Conference on Machine Learning (ICML'11)* (2011), pp. 1017–1024
18. C. Kyunghyun, M. Bart, G. Caglar, B. Dzmitry, B. Fethi, S. Holger, B. Yoshua, Learning phrase representations using RNN encoder-decoder for statistical machine translation, in *The Conference on Empirical Methods in Natural Language Processing* (2014), pp. 1724–1734
19. S. Richard, C. Cliff, C. Lin, Y. Andrew, Parsing natural scenes and natural language with recursive neural networks, in *International Conference on Machine Learning* (Omnipress, 2011), pp. 129–136
20. K. Asifullah, S. Anabia, Z. Umme, Q. Aqsa, A survey of the recent architectures of deep convolutional neural networks, in *Computer Vision and Pattern Recognition* (2019)
21. A. Zahangir, M. Tarek, Y. Christopher, W. Stefan, S. Paheding, N. Shamima, C. Brian, A. Abdul, K. Vijayan, The history began from alexnet: a comprehensive survey on deep learning approaches, in *Computer Vision and Pattern Recognition* (2018)
22. K. Fukushima, Neocognitron: A hierarchical neural network capable of visual pattern recognition. Neural Netw. **1**(2), 119–130 (1988)

23. A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, in *NIPS* (2012), pp. 1106–1114
24. S. Cristian, L. Wei, J. Yangqing, S. Pierre, R. Scott, A. Dragomir, E. Dumitru, V. Vincent, R. Andrew, Going deeper with convolutions, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2015), pp. 1–9
25. H. Tang, W. Wang, K. Gimpel, K. Livescu, Discriminative segmental cascades for feature-rich phone recognition, in *ASRU* (2015)
26. H. Kaiming, Z. Xiangyu, R. Shaoqing, S. Jian, Deep residual learning for image recognition, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2016)
27. H. Hassannejad, G. Matrella, P. Ciampolini, I. Munari, M. Mordonini, Food image recognition using very deep convolu-tional networks. Intl Work Multime Ass Die Mana **1**, 41–49 (2016)
28. Y. Qian, M. Dongyuan, W. Jingfan, Deep learning based food recognition (2016)
29. S. Ashutosh, Y. Lin, E. Touradj, Food/non-food image classification and food categorization using pre-trained Google Net model, in *Conference, the 2nd International Workshop* (2016)
30. Torch, https://torch.ch/
31. Theano, https://deeplearning.net/software/theano/
32. The Theano Development Team: Theano: A Python framework for fast computation of mathematical expressions (2016)
33. Tensorflow, https://www.tensorflow.org/
34. A. Martin, A. Ashish, B. Paul, B. Eugene, C. Zhifeng, C. Craig, Tensorflow: large scale machine learning on heterogeneous distributed systems (2016)
35. Chaier, https://chainer.org/
36. Keras, https://keras.io/
37. Apache Singa, https://singa.incubator.apache.org/
38. C. Tianqi, L. Mu, L. Yutian, L. Min, W. Naiyan, W. Minjie, Mxnet: a flexible and efficient machine learning library for heterogeneous distributed systems (2015)
39. Mxnet, https://mxnet.io/
40. Caffe, https://caffe.berkeleyvision.org
41. J. Yangqing, S. Evan, D. Jeff, K. Sergey, L. Jonathan, G. Ross, G. Sergio, D. Trevor, Caffe: convolutional architecture for fast feature embedding, in *Proceedings of the 22nd ACM International Conference on Multimedia* (2014), pp. 675–678
42. Microsoft Cognitive Toolkit CNTK, https://www.microsoft.com/en-us/research/product/cognitive-toolkit/
43. Deeplearning4j, https://deeplearning4j.org
44. Neon, https://neon.nerva-nasys.com/docs/latest
45. A. Krizhevsky, I. Sutskever, H. Ge. ImageNet classification with deep convolutional neural networks (2012), pp. 1097–1105
46. C. Mei, D. Kapil, W. Wen, Y. Lei, S. Rahul, Y. Jie, PFID: Pittsburgh fast-food image dataset, in *ICIP* (2009), pp. 289–292
47. M. Anand, B. Marc, Z. Fengqing, J. Carol, A. Deborah, S. David, J. Edward, Personal dietary assessment using mobile devices. Comput. Imaging, **7246** (2009)
48. J. Taichi, Y. Keiji, A food image recognition system with Multiple Kernel Learning, in *ICIP* (2009), pp. 285–288
49. K. Yoshiyuki, Y. Keiji, Real-time mobile food recognition system, in *CVPR Workshops* (2013), pp. 1–7
50. B. Lukas, G. Matthieu, J. Luc, G. Van, Food-101 - Mining discriminative components with random forests. ECCV **6**, 446–461 (2014)
51. K. Yoshiyuki, Y. Keiji, Foodcam-256: a large-scale real-time mobile food recognition system employing high-dimensional features and compression of classifier weights. ACM Multimedia 761–762 (2014)
52. Giovanni, M., Farinella, D., Filippo, S.: A Benchmark Dataset to Study the Representation of Food Images. ECCV Workshops, pp.584–599, (2014).
53. D. Joachim, A. Marios, S. Sergey, G. Stavroula, Two-view 3D reconstruction for food volume estimation. IEEE Trans. Multimedia **19**(5), 1090–1099 (2017)

54. M. Anthimopoulos, L. Gianola, L. Scarnato, P. Diem, S. Mougiakakou, A food recognition system for diabetic patients based on an optimized bag-of-features model. J. Biomed. Health Inform. 1261–1271 (2014)
55. O. Beijbom, N. Joshi, D. Morris, S. Saponas, S. Khullar, Menu-match: restaurant-specific food logging from images in 2015. IEEE **25**, 844–851 (2015)
56. C. Jingjing, W. Chong, Deep-based ingredient recognition for cooking recipe retrieval. ACM Multimedia 32–41 (2016)
57. C. Xin, Z. Yu, Z. Hua, D. Liang, W. Dongyan, ChineseFoodNet: a large-scale image dataset for chinese food recognition, in *Computer Vision and Pattern Recognition* (2017)
58. P. Paritosh, D. Akella, T. Bappadi, N. Puhan, FoodNet: recognizing foods using ensemble of deep networks. IEEE Signal Process. Lett. **24**(12) (2017)
59. J. Chen, C. Ngo, *Deep-Based Ingredient Recognition for Cooking Recipe Retrieval* (ACM, 2016), pp. 32–41
60. G. Farinella, M. Moltisanti, S. Battiato, Classifying Food Images Represented as Bag of Textons (IEEE, 2014), pp. 5212–5216
61. Y. Matsuda, H. Hoashi, K. Yanai, Recognition of multiple-food images by detecting candidate regions, in *IEEE International Conference on Multimedia and Expo* (2012), pp. 25–30
62. S. Yang, M. Chen, D. Pomerleau, R. Sukthankar, Food recognition using statistics of pairwise local features. IEEE **1**, 2249–2256 (2010)
63. Y. Ramadevi, B. Kalyani, T. Sridevi, Synergy between object recognition and image segmentation. Int. J. Comput. Sci. Eng. **2**(8), 2767–2772 (2010)
64. Y. Song, H. Yan, Image segmentation algorithms overview, in *Computer Vision and Pattern Recognition* (2017)
65. S. Shakuntala, K. Surendra, A review paper on image segmentation and object recognition procedures. Int. J. Sci. Eng. Technol. **4**(1) (2016)
66. G. Vairaprakash, K. Subbu, Review on image segmentation techniques. Int. J. Sci. Mod. Eng. **1**(8), 2319–6386 (2014)
67. W. Liming, S. Jianbo, S. Gang, S. Fan, Object detection combining recognition and segmentation, in *Asian Conference on Computer Vision* (Springer, Berlin, 2007), pp. 189–199
68. S. Wataru, Y. Keiji, Foodness proposal for multiple food detection by training of single food images, in *Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management* (2016), pp. 13–21
69. C. Bryan, A. Russell, S. Josef, T. William, A. Freeman, Using multiple segmentations to discover objects and their extent in image collections, in *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (2006)
70. Z. Fengqing, B. Marc, K. Nitin, B. Carol, J. Edward, Multiple hypotheses image segmentation and classification with application to dietary assessment. IEEE J. Biomed. Health Inform. **19**(1)
71. A. Roohollah, Q. Kazem, R. Mohammad, An efficient evolutionary based method for image segmentation, in *Computer Vision and Pattern Recognition* (2017)
72. G. Stephen, G. Tianshi, K. Daphne, Region-based segmentation and object detection, in *Proceedings of the 22nd International Conference on Neural Information Processing Systems* (2009), pp. 655–663
73. S. Ashima, S. Sahil, Image segmentation and object recognition using machine learning, in *International Conference on Smart Trends for Information Technology and Computer Communications* (Springer, Berlin, 2016), pp. 204–210
74. K. Yoshiyuki, Y. Keiji, Real-time mobile food recognition system, in *Proceedings of the 2013 IEEE Conference on Com-puter Vision and Pattern Recognition* (2013), pp. 1–7
75. M. Domingo, P. Franco, Segmentation of colour food images using a robust algorithm. J. Food Eng. 353–360 (2005)
76. A. Sandhya, C. Gauri, C. Devangi, D. Eesha, J. Shevi, Survey of different approaches used for food recognition, in *Information and Communication Technology for Competitive Strategies* (Springer, Berlin, 2018), pp. 551–560
77. S. Adnan, Z. Nauman, H. Mir, Food item recognition and intake measurement techniques, in *11th International Conference on Machine Learning and Computing* (2019)

78. Y. Zhao, J. Ming, Y. Chen, C. Cao, C. Forde, N. Wah, S. Tat, Food photo recognition for dietary tracking: system and experiment, in *International Conference on Multimedia Modeling* (Springer, Berlin, 2018), pp. 129–141
79. S. Doyen, H. Wang, K. Shu, X. Wu, L. Hung, A. Palakorn, P. Ee, C. Steven, FoodAI: food image recognition via deep learning for smart food logging, in *Conference on Knowledge Discovery and Data Mining* (2019)
80. M. Simon, K. Barbara, NutriNet: a deep learning food and drink image recognition system for dietary assessment. Nutrient (2017)
81. T. Md, B. Islam, K. Nafiz, S. Siddique, T. Rahman, Image recognition with deep learning, in *International Conference on Intelligent Informatics and Biomedical Sciences* (IEEE, 2018)
82. M. Nida, M. Zaitouna, Survey on image segmentation techniques, in *International Conference on Communication, Management and Information Technology* (Elsevier, Amsterdam, 2015), pp. 797–806
83. C. Vishwanath, C. Priyadarshini, Food image segmentation using edge adaptive based deep-CNNs. Int. J. Intell. Unmanned Syst. 2049–6427 (2019)
84. L. Bossard, L., M. Guillaumin, Lv. Gool, Food-101-mining discriminative components with random forests in computer vision, in *ECCV* (2014), pp. 446–461
85. K. Yanai, Y. Kawano, *Food Image Recognition Using Deep Convolutional Network with Pre-training and Fine-Tuning* (IEEE, 2015), pp. 1–6
86. Yan, L., Guanling, C., Vinod, V., Yunsheng, M., Chang, L.,Yu, C.: Deep food: deep learning-based food image recognition for computer-aided dietary assessment., pp.37–48, (2016).
87. A. Meyers, Im2calories: towards an automated mobile vision food diary, in *IEEE International Conference Computer Vision* (2015), pp. 1233–1241
88. S. Christodoulidis, M. Anthimopoulos, S. Mougiakakou, Food recognition for dietary assessment using deep convolutional neural networks, in *ICIAP* (2015), pp. 458–465
89. P. Pouladzadeh, P. Kuhad, S. Peddi, A. Yassine, S. Shirmohammadi, *Food Calorie Measurement Using Deep Learning Neural Network* (IEEE, 2016), pp. 1–6
90. Y. Kawano, K. Yanai, *Food Image Recognition with Deep Convolutional Features* (ACM, 2014), pp. 589–593
91. G. Ciocca, P. Napoletano, R. Schettini, Food recognition: a new dataset, experiments and results., IEEE J. Biomed. Health Inform. **99** (2016)
92. L. Yuzhen, Food image recognition by using convolutional neural networks, in *Vision and Pattern Recognition* (2019)
93. A. Eduardo, A., B. Marc, R. Petia, exploring food detection using CNNs, in *Computer Aided Systems Theory (EUROCAST 2017)* (Springer, Berlin, 2017), pp. 339–347
94. A. Eduardo, B. Marc, R. Petia, Food recognition using fusion of classifiers based on CNNs, in *Image Analysis and Processing (ICIAP 2017)* (Springer, Berlin, 2017), pp.213–224
95. H. Luis, M. Weiqing, J. Shuqiang, Food recognition and recipe analysis: integrating visual content, context and external knowledge, in *Computer Vision and Pattern Recognition* (2018)
96. I. Sanasam, S. Ayan, K. Pritee, Classification of food images through interactive image segmentation, in *Asian Conference on Intelligent Information and Database Systems (ACIIDS)* (Springer, Berlin, 2018), pp. 519–528