DEEP LEARNING BASED FOOD IMAGE RECOGNITION: A SURVEY

[1]Tejaskumar Patel, [2]Dr. Chirag Patel, [3]Prof. Ankita Gandhi

[1]MTech Student of CSE, [2]H.O.D of Computer Department of Parul Institute of Technology, [3]Professor of Computer Department of Parul Institute of Engineering &Technology

[1]Parul University, Limda, Waghodia, Vadodara, India

Abstract : Deep learning is considered as one the state-of-the-art approach in the field of food image recognition. It is an active and wider approach in the research area. In deep learning for food image detection, CNN is considered as one of the most important method to acquire accurate results. However, along with the time, the limitations of CNN were developed in R-CNN and furthermore developed in Fast R-CNN and in Faster R-CNN to overcome former's limitations. Even YOLO is also one of the quick approach used in the recent research area for food image recognition. Taking this technique to even furthermore, it has been quite successful in applying these techniques to the modern areas of application such as nutritional facts of food from food image , food calorie estimation of food intake, Diet intake etc.

***keywords* – Image-based food object recognition, Deep Learning, CNN, Food Image Detection**

I. INTRODUCTION

It is considered to be one of the major tasks in visionary world of computer system : To detect and identify objects in an image, since it may contain variable objects. For the human vision system to be known, it has become habitual for the humans to implement this task easily as it is naturally shaped. But as we are not fully aware of human vision system, similar procedures cannot be implemented directly to machine. Image is a group of pixels or vectors in any given box of computer. It is very difficult to make an explanation to machines that a small change/changes of pixels will change the image. It takes huge, inventiveness and calculus skills for a machine to garner a simple task a detection and identification.

One of the most developing application of computer world is food detection. Food is most essential intake of humans. Even though it shows as to provide necessary nutrients or supplements to the body. Food is an important element in human's life. It is imbibed in once natural life cycle. That's the reason it gives a way to research more on food image recognition.

With the widespread use of low cost imaging devices like smart phone cameras, more applications are being developed to facilitate automatic object recognition has made easier to calculate and record food intake and energy consume. There are various applications available to record food intake but less in number to track nutritional habits.

With the invent of CNN, various approaches has been proposed to ease the task of nutrition books. One of the most use proposal is food identification. These avenue gives direct access to identify food in an image and even allows to have knowledge about nutritional values. Moreover, no applications or websites contains maximum records of worldwide food dishes (more than 8000 according to Wikipedia).

II. EXISTING SYSTEM

A pretrained model has a fix numbers of classes but the most challenging part is detection issue of the new images generated which are not included in the existing model. Hence accuracy of model and performance is compromised. An algorithm like CNN requires more times to train the model and has less accuracy compare to other algorithms like R-CNN, Fast R-CNN, Faster R-CNN. The latest approach called YOLO is fast and accurate among the other algorithms but it also compromises on accuracy if the objects are very small.

III. LITERATURE REVIEW

[1], In this paper author apply CNN for the task of food object detection and recognition with the help of SPM(Spatial Pyramid Matching) using color histogram and SVM(Support Vector Machine). Author uses Food Logging (FL) as dataset for their experiment and it results better accuracy for the recognition of food images. After experiments author gets success to achieve 93.8% accuracy.

In this paper[2], author uses Faster R-CNN algorithm as food detector to detect each dish in food image. Basically they apply this method to detect food object in a multi-dish food image. And author also uses two kind of food photo datasets [school lunch images annotated with bounding boxes and UEC FOOD-100] for the experiment, they also applied food detector for food calorie estimation for food photos of multiple dishes[2].

[3], In this paper author has applied CNN, as contemporary approach, which is optimize to the task of food detection and recognition with the help of dataset acquired from publicly available internet sources(like Instagram) for the most consume local Malaysian food items. Food-101, UEC-FOOD-100 and UEC-FOOD-256 is used as dataset for optimizing deep conventional network for food image classification.

In this paper[4], A novel image representation that comprise of the covariances of conventional layer feature map is proposed by the author for food image recognition. The author has used FMCD(Feature Map Covariance Descriptor) method along with ETHZ FOOD-101 and ImageNet-2012 as a dataset to classify images using Linear SVM learned with fully connected layer activation of trained CNN.

The author in[5], uses method to classify food images by using the model of Bayesian Network. In this paper, the author has experiment to update correction suggested by users and acquires better results(up to 92%) after analysis. Naive Bayes method used by author is more ideal for updating model based on users feedback.

[6], In this paper, author has applied CNN along with data augmentation techniques based on geometric transformation to increase the size of training images, achieving accuracy of more than 90%. The objective was to expand the data technique and as per comparison, and

conventional BoF(Bag of Features) model along with linear SVM approach was implemented for food image recognition which resulted in improving performance of CNN and the effectiveness of data expansion technique.

The author shows in[7], perspective view of multi-label learning. With the help of two mix dish datasets: mixed economic rice and economic beehoon, the experiment shows the effectiveness of dish recognition on region level with multiple granularities irrespective of the dish size. It also reduces the tedious work of manual labelling which results in increasing all indicators as. Compare to plain multi-label classification.

In paper[8], A significant method to detect food object can effectively evaluate food nutrition. The author uses FOOD-101 dataset along with the architecture base on the Inception-ResNet and Inception-V3 model for getting impressive recognition results. Finetuning on all layers gain better recognition accuracy with updated weights.

The author has addressed in[9], personalization problem of food image detection and has proposed personalization framework which combines a fixed class NCM (Nearest Class Mean) classifier that performs effectively based on deep features. To get realistic results the author generated a new dataset of daily food images collected from daily food recording application called FoodLog and effectively achieve great results.

[10], The author presents an approach to identify and classify food images for further diet monitoring application by using CNN. When the number of classes are more, the author has observed that the CNN is more ideal to classify the food images and hence has achieved 86.97% accuracy from the classes of FOOD-101 dataset.

IV. Overview of Different Methods

4.1. Data Augmentation

Different techniques such as cropping, padding and flipping are used to train large neural networks in a strategy called data augmentation. It helps to increase the range of data without actually collecting new data.

4.2 Scale-invariant Feature Transform (SIFT)

Scale Invariant Feature Transform (SIFT) : It is a technique for detecting salient, stable feature points in an image. for each such point, it also provides a group of “features” that “characterize/describe” a little image region round the point. These features are invariant to scale and rotate.

4.3 Speeded Up Robust Features (SURF)

This is a quick and robust algorithm towards local feature detector for comparison of images. The main feature of this approach is its fast computation of tasks such as tracking and object detection, image classification etc.

4.4 Principal Component Analysis (PCA)

PCA may be a process consists of orthogonal conversion to rework a group of observations of correlated variables into linearly uncorrelated variables which are called principal components. it's considered together of the simplest statistical procedure and it also reduces 2 dimension data to 1 dimension data.

4.5 Linear Discriminant Analysis (LDA)

It is a dimensionality reduction algorithm to find a linear combination of features that classifies or divides multiple classes of objects and the results used as a linear classifier.

4.6 Support Vector Machine (SVM)

Support Vector Machine (SVM) : A SVM may be a discriminative classifier generally referred to as supervised machine learning algorithm which may be wont to create decision boundary called hyperplane. It are often utilized in both classification and regression problems.

4.7 Convolutional Neural Network (CNN)

A CNN consists of an input and an output layer and multiple hidden layers. The hidden layers in CNN contains a series of convolutional layers that convolve with a multiplication or other dot product. RELU layer is activation function, followed by pooling layers and fully connected layers, generally called as hidden layers because their inputs and outputs are combined by the activation function and final convolution.

4.8 R-CNN

In this algorithm and image is divided into 2000 regions and they are selected with the help of selective search process. Henceforth CNN (Convolutional Neural Network) extract features over 2000 regions which may ascertain an object in an image. Moving further it is passed through SVM(Support vector machine) to classify an object over the proposed region.

4.9 Fast R-CNN

The image in fast R-CNN is passed through CNN to produce a convolutional feature map. With the help of this feature map, it detects a region proposals and framed in to square using RoI polling layer. It is reframed into a fixed sized so that it can be feed into the fully connected layer form RoI layer, a class of proposed region is forecasted using SoftMax layer. Henceforth, it creates a set values for the bounding box. With the support of offset values it helps to locate an object in an image.

4.10 Faster R-CNN

In order to surmount the problem of fast R-CNN, faster R-CNN approach was developed, same as in fast R-CNN, a conventional feature map is generated out of an image using CNN. Here the difference in the working of faster R-CNN is : it uses different network to speculate region proposal and then follows the same procedure of fast R-CNN of reshaping to a definite size to be put up into fully connected layer. Henceforth, using SoftMax layer a class of proposed region is predicted which creates offset values of the bounding boxes hereby detecting an object.

4.11 YOLO

A part from above three different types of CNN, YOLO is quite different and quick approach. While the other types uses CNN to divide an image, YOLO uses a single convolutional network for anticipating bounding boxes and class probability of bounding boxes. And images divided into SxS grid. Amidst each grid, N number of bounding boxes are taken into account and it yields to network output that is class probability and offset values for bounding boxes. Here the bounding boxes having class probability above pre-decided value is chosen and gradually use to find the object within an image.

4.12 Comparison Table

Table 1: Comparison of Different Methods

Method	Advantages	Limitation
PCA	- Reduce overfitting - Improve visualization.	- Independent variable become less interpretable. - Information loss
ANN	- Storing information on the whole network - It has fault tolerance - Parallel processing ability	- Hardware dependence - The problem of showing the matter to the network - The time period of the network is unknown
CNN	- Fast model training - Automatically extract features without human supervision.	- It doesn't consider Spatial relationship of feature - sometime faced overfitting issue.
R-CNN	- Image divided into 2000 regions so only 2000 regions needs to classify	- Every time 2000 region classification sometimes not required it leads towards more training time
YOLO	- Speed - used for real time object detection	- sometimes face difficulties with small objects.

V. Limitation

1. Food object cannot be identified in some challenging condition like input image is not properly captured, for example blurred image.
2. Lack of training data of food images in dataset leads towards wrong prediction or less accurate predicted result.
3. Because of the wide variety of food objects and its dataset, it takes more time to train the model for better accuracy.

VI. Conclusion and Future work

This survey paper focus on developing Image-based food object recognition system based on Deep Learning approach called CNN by summarizing different literature Review. It also discusses about tradition approach of feature extraction and Deep learning-based CNN model. Deep learning is sub field of machine learning, the way of learning mechanism that uses Image based food object recognition. CNN uses multi-layers and even provides multiple processing layer for linear and nonlinear transformation. This all make food object recognition system accurate and high recognition result. Even in Challenging condition like low light. Through the observation of trained convolution kernels and optimizing the architecture & hyper-parameters of the network, performance is evaluated. It is shown that it is beneficial for classification accuracy to have this depth. In future, we can also develop applications that can invent new kind of food based on the food images and ingredients. That is, it may be possible to develop an AI chef that augments the traditional recipe with new machine-generated ones. Moreover, such application may also be able to predict replacement ingredients in recipes according to the tastes.

Acknowledgement

First of all. Thank you for giving me hope and motivating me, I would like to acknowledge the support that I got from Dr. Vilin Parekh. Principal of Parul Institute of Engineering and Technology. I am greatly thankful to Dr. Kirit Modi, HOD of CSE. I want to express a warm gratitude to all staff members of Parul University Computer Science and Engineering Department and my friends who was always there for me in any situation.

REFERENCES

[1] Kagaya, H., Aizawa, K., & Ogawa, M. (2014). Food Detection and Recognition Using Convolutional Neural Network. *Proceedings of the ACM International Conference on Multimedia - MM 14*. doi: 10.1145/2647868.2654970

[2] Ege, T., & Yanai, K. (2017). Estimating Food Calories for Multiple-Dish Food Photos. *2017 4th IAPR Asian Conference on Pattern Recognition (ACPR)*. doi: 10.1109/acpr.2017.145

[3] 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

[4] Tatsuma, A., & Aono, M. (2016). Food Image Recognition Using Covariance of Convolutional Layer Feature Maps. *IEICE Transactions on Information and Systems*, *E99.D*(6), 1711–1715. doi: 10.1587/transinf.2015edl8212

[5] Maruyama, Yuto, et al. "Personalization of Food Image Analysis." *2010 16th International Conference on Virtual Systems and Multimedia*, 2010, doi:10.1109/vsmm.2010.5665964.

[6] Lu, and Yuzhen. "Food Image Recognition by Using Convolutional Neural Networks (CNNs)." *ArXiv.org*, 25 Feb. 2019, arxiv.org/abs/1612.00983v2.

[7] Yunan Wang, Jing-jing Chen, Chong-Wah Ngo, Tat-Seng Chua, Wanli Zuo, and Zhaoyan Ming. 2019. Mixed Dish Recognition through Multi-Label Learning. In Proceedings of the 11th Workshop on Multimedia for Cooking and Eating Activities (CEA '19). Association for Computing Machinery, New York, NY, USA, 1–8. DOI:https://doi.org/10.1145/3326458.3326929

[8] Yu, Qian and Jingfan Wang. "Deep Learning Based Food Recognition." (2016).

[9] S. Horiguchi, S. Amano, M. Ogawa and K. Aizawa, "Personalized Classifier for Food Image Recognition," in IEEE Transactions on Multimedia, vol. 20, no. 10, pp. 2836-2848, Oct. 2018.

[10] D. J. Attokaren, I. G. Fernandes, A. Sriram, Y. V. S. Murthy and S. G. Koolagudi, "Food classification from images using convolutional neural networks," TENCON 2017 - 2017 IEEE Region 10 Conference, Penang, 2017, pp. 2801-2806.