Journal of Artificial Intelligence and Computing (JAIC)
ISSN (e): 3005-3358
Volume: 2, Number: 1, Pages: 9- 18, Year: 2024

A Cloud Edge Collaboration of Food Recognition Using Deep Neural Networks

Muhammad Talha Khan[1], Muhammad Hassan Khan[2]

[1]FAST National University of Computer and Emerging Sciences, Lahore, Pakistan (Email: talhakhan_official@outlook.com)
[2]University of Western Sydney, Sydney, Australia (Email: hassan.khan@westernsydney.edu.au)

***Abstract*—Deep neural network-based learning methods are commonly used for classifying images or object detection with excellent performances. In this paper, we looked at how effective a deep convolution neural network (DCNN) is to identify food photography. Food identification is a sort of visual fine-grain recognition that is more difficult than traditional image recognition. Mobile apps in many countries have been omnipresent in many aspects of people's lives in the last few years. Fitting the healthcare potential of this pattern has become a focal point in the industry and researchers' basic applications for architecture that patients should use in their well-being, prevention or treatment method. Mobile cloud computing has been introduced as a possible mobile well-being paradigm interoperability problems management service in various information formats. In this paper, I am integrating deep neural networks with cloud architecture to avoid substantial memory loss in mobile devices or web platforms where I can upload images, and it will predict the actual images and the names of food categories. The datasets I will be using are UECFOOD101; I will be running my deployment on the system as well as on the cloud for retrieving the data easily on mobile phones and web pages. The cloud architecture helps me offload the data, which is not required using computational offloading and profiling.**

***Index Terms*—Artificial intelligence, machine learning, neural network, image processing.**

I. Introduction

Food identity based on photographs raises new challenges and mainstream algorithms for computer vision. The latest functions in the area are based on, or learning from, manufactured representations by using deep neural networks (DNN). Though DNN-based works are successful, they use profound and non-distributive architectures. We think that if the architecture is described well, better results can be achieved with reference to dietary composition research. Quick, precise and automatic food attribute determination is a realistic necessity in everyday life. Modern techniques, including computer vision, spectroscopy, and spectrum imaging, are most commonly used for food attribute identification. They can acquire information related to the composition of food. Data review is of the utmost significance since the vast volumes of data have a lot of redundancy and Knowledge that is meaningless. How can you cope with so much? Data and the retrieval of valuable data are highly required, and it is also a challenge to implement these methods and essential problems to apply to the real world (APP).

Many methods of data processing were developed for the management of a huge number of data, such as partial squares for simulation, The artificial neural network (ANN) support vector Machine (SVM), random forest in 2012, K- nearest neighbour (KNN)), etc. For the first time. extraction of functions such as principal component analysis (PCA), wavelet transform (WT), independent correlation amongst components scale-invariant transform function, speedup Robust Characteristics, histogram of oriented gradient, and so forth. These techniques in dealing with these data have shown great value.

Convolutional neural networks and their derivative algorithms have been mostly surveyed in dietary or food-related articles, which can learn digital input information for subsequent regression or classification tasks. CNN successfully collected a vast amount of data from the tools for evaluating food quality and protection. CNN can successfully accommodate optical cameras and so on. Food image recognition is one of the promising uses for identifying visual objects, as it helps to measure food calories and evaluate food patterns treatment with well-being. Many works were then written [1, 2, 3]. Increasing the amount of recognized foods is important to make food identification more successful. We built an EUCFOOD 100 food data collection and performed experiments with 100 food classes. The consistency of classification recorded to date is 72.26%, which is still required to improve for practical changes. In addition, we recently suggested an extension structure automatically an internal dataset, and we expanded it with a data collection of 100 class Foods in 256 class Food Dataset FOOD101. We need a more advanced classification of food images to boost the practicality of food classification.

Problem & Motivation:

There are several documentation and works related to food recognition. Still, the problem arises if the dataset is too large

and there is a big amount of memory consumption in loading the single datum. And it is not possible to load this into a mobile phone.

Proposed Solution:
I will use the deep neural network technique along with Cloud computing to perform the analysis of food recognition on mobile devices having less memory consumption and perform the computational offloading for deploying the recognition of food images on mobile or Web platforms.

II. LITERATURE REVIEW

A. A Dataset for Large-Scale Food Recognition via Stacked Global-Local Attention Network. [1]:

Food computing is arising as another field to enhance the issues from food-important fields, for example, farming, nutrition, and medication. Existing works center on using more modest datasets for food recognition. These informational collections are absent of diversity and inclusion in food classes excludes a wide scope of food pictures. The article includes a broad assessment for the proposed dataset and the other two mainstream food benchmark datasets to confirm the method's viability. Another enormous scope and exceptionally different food image dataset with five hundred classes and around four lac pictures, which will be made freely accessible to the advancement of versatile food recognition, is presented. After collecting and annotating the image, there are as yet many food classifications with few pictures. Work is likewise extremely applicable to fine-grained image recognition in Fig. 1, which expects to arrange subordinate classes.

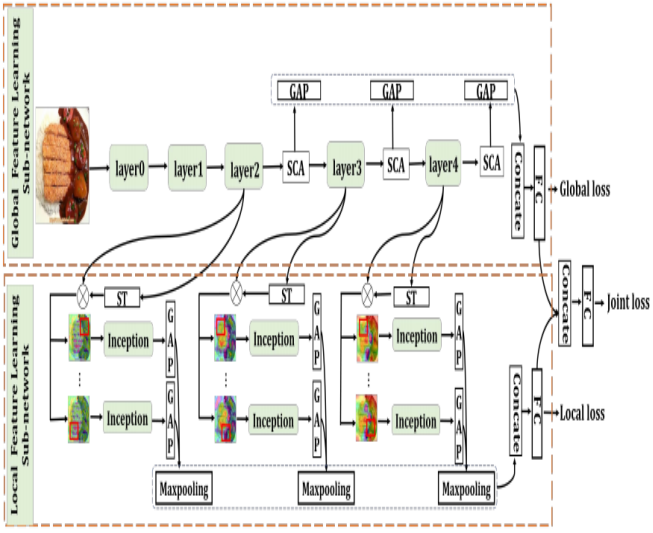

Figure 1: Fine-grained Image recognition.

Food recognition has a place with fine-grained image recognition. GloFLS first embraces the SCA to gain more discriminative highlights from each layer of the organization, and at that point totals a group of highlights from these layers to catch various types of worldwide level highlights, for example, shape and surface prompts about food. SGLANet is prepared with various types of losses, including global loss, joint loss, and local loss in a start to finish style to augment their corresponding advantage in terms of the discriminative force. The exploratory outcomes exhibit that it gains the best recognition execution when various losses are used.

Table I: Various methods with Top-1 and Top-5.

Method	Top-1 acc.	Top-5 acc.
VGG-16 [47]	55.22	82.77
GoogLeNet [36]	56.03	83.42
ResNet-152 [17]	57.03	83.80
WRN-50 [46]	60.08	85.98
DenseNet-161 [21]	60.05	86.09
NAS-NET [60]	60.66	86.38
SE-ResNeXt101_32x4d [19]	61.95	87.54
NTS-NET [55]	63.66	88.48
WS-DAN [20]	60.67	86.48
DCL [11]	64.10	88.77
SENet-154 [19]	63.83	88.61
SGLANet	**64.74**	**89.12**

The existing strategies are far from handling huge scope recognition tasks with high exactness like ImageNet, highlighting energizing future bearings. This is because these techniques, for example, IG-CMAN, presented extra ingredient data. Broad assessment of ISIA Food-500 datasets has confirmed its adequacy and, in this manner, can be considered as a number gauge.

B. Mobile multi-food recognition using deep learning [2]:

In this article, we propose a mobile device for food identification using the food image user mobile application, for example, Steak and potatoes, to recognize several food items at the same food tray and to predict calorie and meal nutrition. To increase the speed and precision of the operation, The consumer will be promptly asked to describe the general food region by drawing a food binding circle on tapping the screen photo. The device then uses numerical and image recognition to Recognize food objects. Rather than remembering the entire meal, the benefit of the method is that it is trained only with pictures of single food objects.
The author proposes a portable system of food recognition. At the stage of testing, they initially created a group of different regions of the candidate. In every locale, it registers a grouping score dependent on its removed CNN, which includes and expects food names of the chosen districts. The proposed strategy uses modern deep learning strategies to highlight extraction and characterization. The model has accomplished a precision of seventy-five percent in characterizing forty-three unique types of organic product. Trial results show 83.6 percent exactness for the food class recognition. They recommend an algorithm that mutually distinguishes food category recognition. The framework manages an enormous arrangement of images and cycles them to infer brings about different configurations. Specific Search has been extensively utilized as a recognition technique by many item indicators.

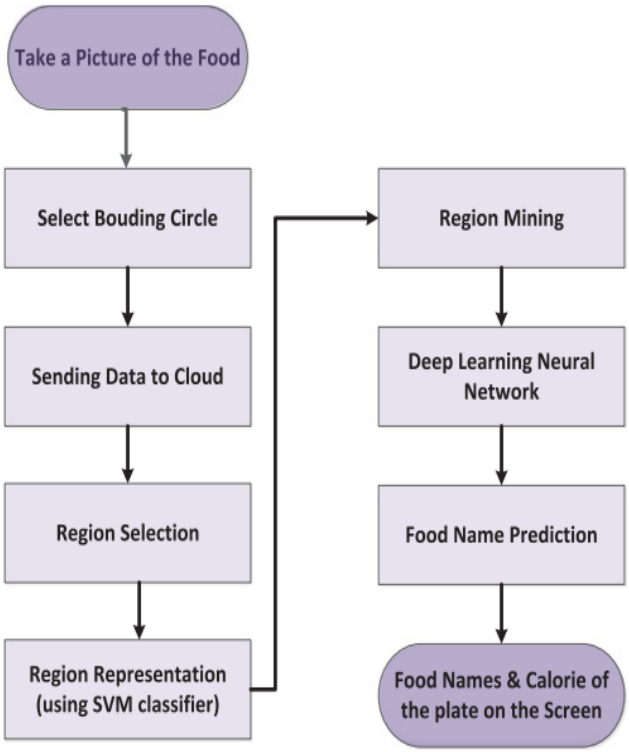

Figure 2: Flow chart of system.

For positive areas, since various proposed locales have various types of segregation toward the aim item class, an area mining method assesses the segregation to find the positive districts that would best separate the aim item class from the foundations. The plan of the profound neural organization depends on a two-venture measure in which the initial step. Two hidden layers merge to frame the RBM, which further structure the pile of RBMs. The proposed framework will pick the food thing with the most noteworthy likelihood. Giving a few food things comes from the overall idea driving profound learning.

Table II: Recognition System.

	Recognition Rate (%)			
N	Food items	Recall	Precision	Accuracy
1	Red Apple	93.64	96	96
2	Orange	95.59	97.5	98
3	Corn	84.85	80	85
4	Tomato	89.56	97	96
5	Carrot	93.25	98	98
6	Bread	98.39	89	90
7	Pasta	94.75	98	98
8	Sauce	88.78	92	93
9	Chicken	86.55	89	88
10	Egg	81.22	87	90
11	Cheese	95.12	97	97
12	Meat	95.73	96	96.5
13	Onion	89.99	93	95
14	Beans	98.68	95	97
15	Fish	77.7	85	88
16	Banana	97.65	97	97

The classifiers prepared without area mining gives a low accuracy rate. The local mining method in the proposed pipeline depends on the presumption that object areas of images having a place with a similar class will group intently together in element space. In the proposed framework, the mix of those strategies gives a ground-breaking instrument for the precision of food recognition.

C. Voting combinations-based ensemble of fine-tuned convolutional neural networks for food image recognition [3]:

CNN is a profound neural organization roused by the natural cycle called the animal visual cortex. ResNet, VGGNet, InceptionV3, and GoogleNet with fine-tuning are utilized in the article and are dependent on profound learning for PC-assisted food recognition tasks. From an innovative perspective, they can utilize the proposed strategy to improve the execution of different Image recognition issues, for example, determining clinical illness, plant or animal species recognition, security, etc. The authors intend to merge the benefits of the group and profound learning methods for food recognition.

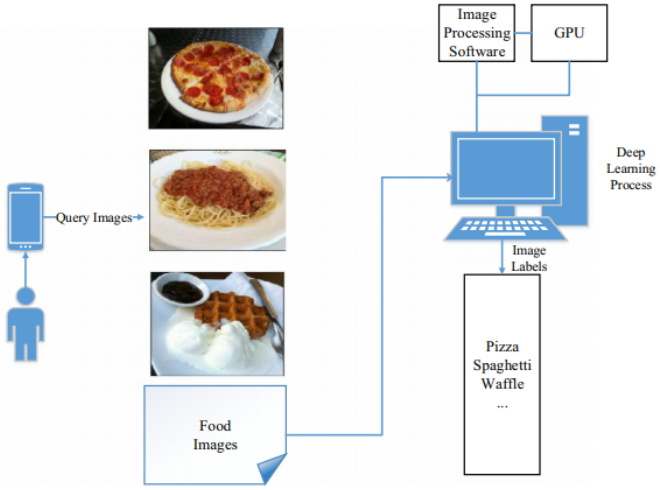

Figure 3: AI Platform.

The Bayesian optimization algorithm is broadly utilized for AI algorithms to upgrade significant expense discovery functions around the world, and it advances by iteratively building up a worldwide measurable model of the goal function. In addition, weighted democratic activity gives more viable and adaptable outcomes for the democratically based grouping measure.

Table II: CNN Architecture.

CNN Architecture	Layers Deep	Fine-tuning Time		
		Food-101	UEC-FOOD100	UEC-FOOD256
VGG16	16	13h 50m	1h 57m	4h 21m
VGG19	19	47h 45m	7h 35m	16h 48m
GoogleNet	22	5h 30m	39m	1h 32m
ResNet101	101	46h 6m	6h 33m	14h 18m
InceptionV3	48	36h 56m	5h 10m	11h 27m

The framework requires both of them to have a profound learning cycle, such as pre-handling and fine-tuning. There is a compromise between the exactness and speed contrasted with the utilization of CNN designs. In future work, they might

assess the created framework with client-defined datasets of food images for noticing and looking at performing the framework.

D. Multi-Tasks Guided Multi-View Attention Network for Chinese Food Recognition [4]:

Food recognition assumes a basic function in different medical services applications. This paper recommends a novel Multi-View Attention Network within various tasks learning system that fuses different semantic highlights into food recognition from both formula displaying and constituent recognition. It is fundamental to present more semantic data from different points of view, such as ingredients and plans. The proposed Multi-View Attention Fusion instrument uses a module to change loads of the semantic highlights. Recently, a quality-based perform various tasks adaptation loss is projected to enhance the exactness in the transformation case, while they propose the incompletely shared structures to undertake explicit portrayal.

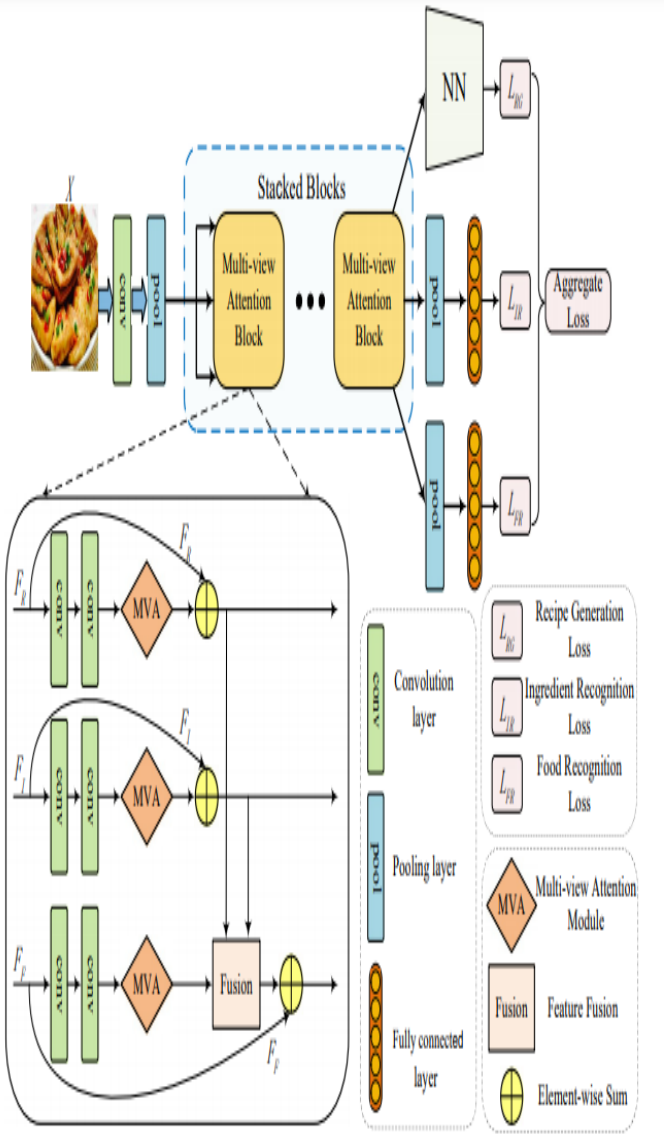

Figure 4: Combination of models

Various combinations of multi-see consideration module in various assignments likewise accomplish further advancement, showing that the multi-view consideration combination system empowers the various undertakings to share task-explicit highlights viably with one another. The exhibition of MVANet50-bilinear is near that of MVANet50- link. The multi-scale data can undoubtedly influence the exhibition of food recognition. MVANet264 accomplishes better, showing that movement can incorporate more extensive and strong semantic highlights from various tasks. MVANets not just can accomplish a better exhibition, yet in addition, needs fewer boundaries. A modest number of boundaries lessens the danger of over-fitting and spare computational assets. For a more profound comprehension of how the proposed multi-view consideration system changes load of various semantic highlights in the displaying cycle, the authors examine the dynamic changes. MVANet has a more grounded actuation for target object districts.

Table IV: Various Models.

Model	Params	Top-1 Acc (%)	Top-5 Acc (%)
ResNet50 [3]	23.97M	60.53±0.28	86.42±0.08
MVANet50-addition (ours)	18.05M	63.50±0.25	88.41±0.15
MVANet50-bilinear (ours)	18.05M	63.70±0.31	88.50±0.18
MVANet50-concatenation (ours)	18.05M	**63.72±0.25**	**88.51±0.14**

The research study proposes a novel strategy, MVANet, for recognizing food, depending on the recently proposed performance of various tasks learning systems and multi-see consideration instruments. The proposed multi-view consideration can adaptively combine many semantic highlights from various undertakings to get a better element portrayal. MVANet not only gains the huge development of the execution for food recognition, but it additionally has fewer boundaries with the goal that it mostly consumes a more modest space and costs less time.

E. An approach for food recognition on mobile devices using convolutional neural networks and depth maps. [5]:

The technique used to assess the volume is known as depth map fusion and includes capturing various pictures from different points and then figuring a 3D model of the item. The paper presents a method for identifying sorts of food and calculating its mass on cell phones by utilizing just the phone's camera. The proposed method comprises two fundamental segments: one is food type detection and the other is volume assessment. The strategy presented in the article straightforwardly and uses a point cloud merging method known as ICP to get the intertwined model. With the headway of innovation accessible to the normal buyer, it is conceivable to use PC vision algorithms to appraise the mass and the certain volume straightforwardly from the telephone.

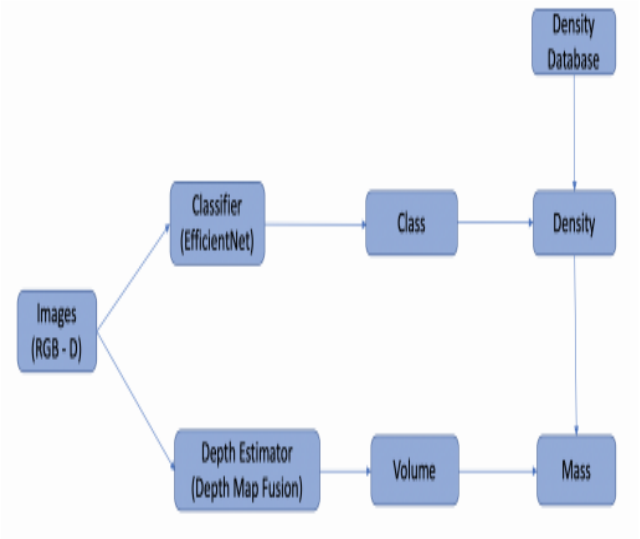

Figure 5: Algorithmic flowchart.

The unwavering quality and exactness of the technique proposed for food volume assessment are observationally demonstrated by the trial results. The article elaborates various recreations with standard sizes of vowels and makes a function that elaborates the ideal Voxel size: figuring the volume of the item. They planned SLAM systems to use sparse point mists to decide the continuous position and direction of the camera. This paper filled in as a beginning stage for the model. The authors presented the idea of a Depth Map, it elaborates the connection between the camera and the object.

Table V: Depth map of objects.

Object	Banana	Pepper	Mushroom	Pear
Estimated volume (cm^3)	140	337	83	149
Actual volume (cm^3)	130	325	80	135

Object	Carrot	Apple	Lemon	Tomato
Estimated volume (cm^3)	125	149	148	86
Actual volume (cm^3)	120	165	135	85

A discretionary third tracking should be possible again with the general purpose. This impact is far less exceptional and takes more time to have an effect. Since the dataset was defective, it was a repetitive change between the dataset and model. It used the information to prepare the model, yet in addition, it used the model to channel the information by eliminating top losses that were produced because of mistakes. This cycle incredibly improves model execution.

F. Wide-slice residual networks for food recognition. [6]:

Image-based food acknowledgment presents additional difficulties for mainstream PC vision calculations. Towards the ultimate target, the authors first introduced a slice convolution square to catch such explicit information. Then, they influenced the ongoing accomplishment of profound leftover squares and combined those with the slice convolution to deliver the order score. The proposed WISeR arrangement expands upon such a thought by proposing a profound learning design that intends to catch the food structure. Setting information regarding the area where the food image was taken, together with extra information about the eatery. By combining the highlights recognized through such a layer with a heap of leftover learning blocks, they got a decent portrayal of food dishes that do not show a particular structure.

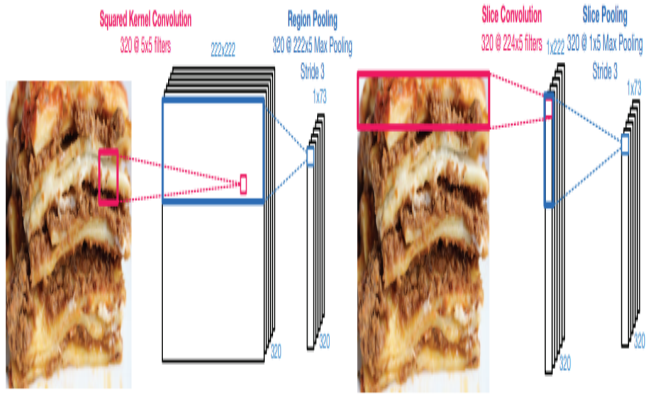

Figure 6: Image processing

They will probably make a single-effort of a food dish and yield the corresponding food classification. The proposed model points to accomplish quite a goal by discovering the underlying idiosyncrasies of the image by combining a slice convolution layer with remaining learning. For a particular vertically organized food classification, they do not ensure that the vertical layers show up similarly situated. To handle this issue, they performed max pooling on vertically extended windows. Results got for the three datasets exhibit that the answer shows better execution regarding existing methodologies either dependent close by highlights or on profound learning plans. They do not intend this shows that our method can address the nontrivial challenges in food acknowledgment and to tackle the particular issues gained by a single dataset.

Table VI: Various methods of publication

Method	Top-1	Top-5	Publication
MLDS	42.63	-	ECCV2014 [2]
RFDC	50.76	-	ECCV2014 [2]
SELC	55.89	-	CVIU2016 [26]
AlexNet-CNN	56.40	-	ECCV2014 [2]
DCNN-FOOD	70.41	-	ICME2015 [42]
DeepFood	77.4	93.7	COST2016 [24]
Inception V3	88.28	96.88	ECCVW2016 [9]
ResNet-200	88.38	97.85	CVPR2016 [12]
WRN	88.72	97.92	BMVC2016 [44]
WISeR	**90.27**	**98.71**	Proposed

The WISeR architecture combines highlights separated from two organization categories. The learning category gives a profound pecking order which can catch the food attributes of most of the existing food classifications. Correlations with existing techniques have shown that by exploiting the two branches preferred execution over state-of-the-craftsmanship approaches is accomplished.

G. *Bilinear CNN models for food recognition. [7]:*

Because of the variety of types of food and the contrasts among various dishes, the class of images is another challenge in PC vision. The authors have abused quite a comparable structure in which it uses two profound networks as highlight extractors and the yields of them melded to get fine-grained highlights. It treats the food acknowledgment as an acknowledgment task, so the authors investigate the utilization of CNN Models for food distinguishing proof. It supplants the element extractor in the model with a further developed profundity learning to engineer to upgrade the capacity to highlight the concept of extraction.

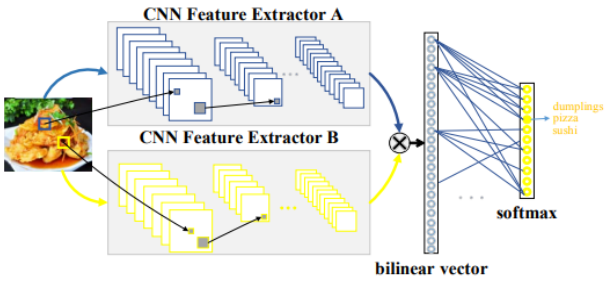

Figure 7: CNN Feature extraction

An undeniable method to expand the exhibition of the CNN network is to raise profundity and width. To approve the method, the authors have noticed results on three standard datasets for food acknowledgment. In the method, they reflect two CNNs for extracting highlights in the bilinear design–the Inception-v3 also, the Inception-v4. They assess the datasets based on three benchmarks to illustrate the prevalence of the approach proposed in the article. The research results elaborate that the performance is in the present status of approaches.

Table VII: Comparison of Methods

Method	UECFOOD-100		UECFOOD-256	
	TOP-1	***TOP-5***	***TOP-1***	***TOP-5***
FC7[26]	58.03	83.71	-	-
Patch-FV+Color Patch-FV(flip)[27]	59.6	82.9	-	-
Color-FV[28]	-	-	41.60	64.00
Color-FV+HOG-FV[28]	-	-	52.85	75.51
Fisher Vector[16]	65.32	86.70	52.85	75.51
DeepFood[27]	77.2	94.8	63.8	87.2
DCNN-FOOD[16]	78.77	95.15	65.57	88.97
Inception-v3[17]	81.45	97.27	76.17	92.58
WISeR[18]	89.58	99.23	83.15	95.45
B-CNN[In-v3, In-v3]	88.64	99.21	83.06	95.44
B-CNN[In-v4, In-v4]	91.26	99.27	84.58	95.49
B-CNN[In-v3, In-v4]	**91.87**	**99.28**	**84.92**	**95.49**

The outcome shows that the grouping of two-element extractors of various structures can accomplish enhanced execution. Presumably, because the yields of highlight extractors with various structures have a low relationship. There are a few outwardly comparable disarrays classes that prompt a few errors in the dataset.

H. *FoNet-Local Food Recognition Using Deep Residual Neural Networks. [8]:*

Recognition capacity for the local food mirrors the social strength. The Computer Vision people group has concentrated on visual food investigation, for example, food location, food acknowledgment, food confinement, and bit assessment. This is the reason they propose a novel method for local food acknowledgment, where the authors have made and used a Deep Residual Neural Network. In the paper, the authors display another model of CNN. Above all, they gather the information, by then, the information is pre-handled, then the information is resized since the authors need the information for the test. Starting there forward, they increased the information by growth technique.

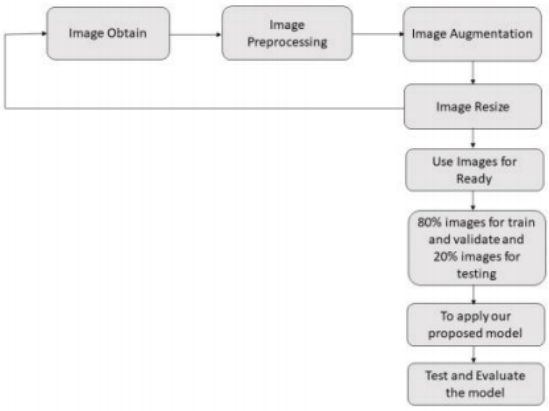

Figure 8: Optimizer

In the proposed strategy, it uses Adam Optimizer. The model has been considered in the food dataset and discovered a great outcome on the train, test, and confirmation set. The authors used the proposed model with CNN models. In the wake of using this model. The exactness level is the model when assessed is high that is 98.16%. We have used an actuation capacity to make the yield non-linear.

Table VIII: Models with accuracy

Model Name	Training Accuracy	Testing Accuracy	Parameters (millions)
Inception-V3	94.70%	95.80%	23.8
MobileNet	93.80%	94.50%	4.2
FoNet (Proposed)	98.26%	98.16%	25.6

In deep learning, it orders explicit highlights into each network by essentially passing information through various layers of the network as input. The loss steadily decreases in each age of approval loss and training loss. Then again, the exactness is increased in each age of training exactness and approval precision. The real utilization of deep learning neural frameworks is significantly greater. Since the concept of machine learning calculations needs marked information, they

are definitely not reasonable for solving complex questions that comprise huge measures of information.

I. Food image recognition using deep convolutional network with pre-training and fine-tuning. [9]:

Food recognition includes visual acknowledgment, which is a moderately more troublesome issue than ordinary image acknowledgment. To handle this issue, the authors looked for the best relation strategies related to DCNN, for example, pre-training the enormous scope of ImageNet information, fine-tuning also, and enactment highlights extricated from the pre-trained DCNN. In this technique, information of DCNN yield is a class-name likelihood. DCNN includes all the item acknowledgment steps, such as nearby element extraction, highlight coding, and learning. If a huge scope of training information is consistently required, it restricts relevant issues of a DCNN. To avoid such circumstances and to make a DCNN successful for some scope information, they have proposed two significant strategies until this point.

The authors applied the food classifier of the DCNN strategies to Twitter photograph information. The extraordinary enhancements for food photograph mining include both the number of food photographs and exactness. Also, as of late, a system is proposed to expand an existing dataset and a 100-class food dataset is broadened into a 256-class food dataset called UEC-FOOD256.

It removes highlights from the locales within the following given bounding boxes. It assesses the exactness of the arrangement within the top N up-and-comers, employing 5-overlap cross-approval. Then, in this paper, DCNN highlights are applied fine-tuning with broadened training information for 100/256-class food datasets and examining the viability of DCNN highlights. In this paper, the adequacy of pre-training and fine-tuning of DCNN is examined with a little scope food dataset which has around 100 training images for every one of the food classifications.

J. Ontology alignment using Named-Entity Recognition methods in the domain of food. [10]:

In recent years, they have done a lot of research in prescient modelling in the domain of medical care. UMLS links biomedical vocabularies to empower interoperability. In the food domain, such resources are scant. To address this issue, the paper investigates a strategy for metaphysics arrangement in the food domain by NER strategies dependent on various semantic resources. To play out the arrangement the FoodBase corpus is utilized, which comprises plans commented on with food elements and includes a ground truth variant which is furthermore utilized for assessment. This paper uses the curated version to perform the cosmology arrangement just as assess strategy.

Topic specialists physically checked this form, so it took the false-positive food substances out. Every annotation contains the area of the separated element, for example, wherein the crude content of the surface structure representing the idea happens and its corresponding semantic labels from the Hansard corpus. Web scraping pages for the URIs gives helpful information that can distinguish food from non-food ideas, for example, the more extensive class to which idea of interest has a place. Authors can find an ongoing correlation of existing food NER strategies, where the creators contrast the exhibition of FoodIE and NER strategies using other food ontologies accessible in the BioPortal.

It uses the recently proposed FoodIE NER technique and the Wikifier text annotation instrument. The test results show that FoodIE gives even more promising outcomes than Wikifier, achieving an F1 score of 0.9605, analyzed to 0.5611. This is normal since FoodIE is explicitly intended for the food domain, while Wikifier uses general jargon and explains text with Wikipedia ideas. It can include more food semantic resources to give mapping between many ontologies. Doing this is subject to a NER strategy that works with ideas from the ideal food semantic asset.

K. DeepFood: Deep Learning-based Food Image Recognition for Computer-aided Dietary Assessment. [12]:

The aim of this paper is to enhance food evaluation accuracy through study of food pictures taken by mobile devices (e.g., smartphone). Deep-learning food-recognition algorithms are the main computational breakthrough in this article. Important evidence has shown that digital imaging estimates nutrition in certain conditions correctly and that there are many benefits relative to other approaches. How to easily and reliably draw nutritional knowledge from the food picture .We recommend a new one Food image recognition algorithm based on the Convolutional Neural Network (CNN), to solve this problem.

A precise estimation is necessary to determine the dietary caloric intake Efficacy in treatments for weight loss. Present dietary appraisal techniques rely on auto-reports and manually recorded equipment (e.g. 24-hour meal recall and food frequency questionnaires). Although the 24-hour dietary recall is a gold standard, the individual also exhibits a bias in measuring their intake (short and long term). Evaluation of a participant's dietary intake can contribute to undervaluation and underestimation of the consumption of food. Our method was specifically influenced by LeNet-5, GoogleNet and AlexNet. The CNN concept was originally inspired by Primate visual cortex neuroscience paradigm. The paper's key insights are how to render machine learning like the human mind with many neurons.

The brain of humans is known to regulate multiple neurons, which is how the machine should comprehend and think in a human way. Many artificial intelligence researchers have long been on the issue. This network demonstrates the essential components of neural convolution networks (CNN). There are three layers marked as C1, C3 and C5, subsampling layers marked as S2, S4 and fully connected layers as F6 and output layers. A receptive field (we call it a fixed-size patch or kernel) is for convolutional layer. Picked to measure the convolution in the input for the same patch size. A stride has been taken to make sure that all pixels are protected and created in the original image or function map and generate the corresponding output feature map. The pre-trained model has been trained with 1.2 million training photos.

Moreover, 100,000 test images. Using the UEC-256 data set with a category number of 256 productions, we can further

refine the model based on the pre-trained model. The models were precisely tuned (ft) at 0,01 base learning rate, 0,9 momentum and 100,000 iterations at their base-learning frequencies. The results are shown:

Table IX Comparison of accuracy on UEC-256 at different iterations using UEC-256

# of Iterations	Top-1 accuracy	Top-5 accuracy
4,000	45.0%	76.9%
16,000	50.4%	78.7%
32,000	51.2%	79.3%
48,000	53.1%	80.3%
64,000	52.5%	80.3%
72,000	**54.7%**	**81.5%**
80,000	53.6%	80.1%
92,000	54.0%	81.0%
100,000	53.7%	80.7%

III. Proposed Methodology

In this section of the paper, we will discuss the proposed methodology that is being followed to build a new system. As we have discussed a little bit about our proposed system in Chapter 1, we will apply deep learning techniques such as Deep Residual Networks and Wide Slice Neural Networks. Previously, it was highly focused on the vertical layer of the food dishes, e.g. structural data. The slice convolution helps detect the specific food layer at a specific location. We'll try to extend this research to a cloud-based system. The research is basically on moving datasets to the cloud and computing costs. The approach utilizes mobile edge computing to offload application computations and communications to the edge, thus increasing the processing capacity and improving user comfort. The system leverages the cloud to train convolutional neural networks for food recognition and classify photos segmented by mobile devices or web platforms.

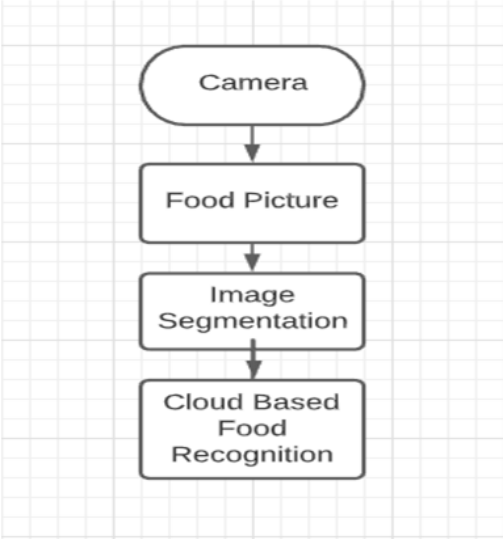

Figure 9 Flowchart for basic understanding

As we can see in Fig. 9. Initially, the mobile phone camera captures the picture, and there are steps for analyzing the information in the picture. We will use the previous methods of extracting information from a food item's picture but will do this in a cloud architecture. The detailed architecture is given below between a mobile device and a cloud server.

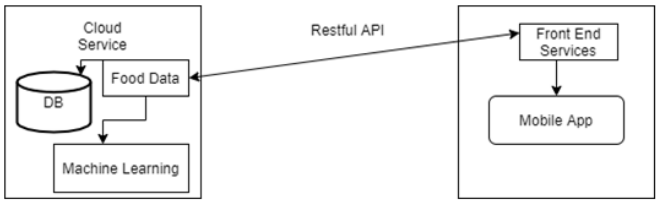

Figure 10: The cloud architecture for food recognition

Figure 10 is the architecture for detecting the food image. The input image will be sent to the cloud to trigger the food recognition and analysis.

A. System Architecture:

Cloud service supports two kinds of operations. The First of all are computer-intensive operations to be carried out on or at the edge of mobile devices. Still, they are not latency-sensitive, such as food detection by machine learning. The second form of job is Data visualization, classification, and storage based. The server for food communicates with the edge operation and stores it in the database. The part of machine learning is the food detection model, where the food picture is used as an input and the relevant regions of the image are detected for the output performing food recognition. It is stored in the nutrition server. Figure 11 will be translated to nutritional information using API calls to the external database.

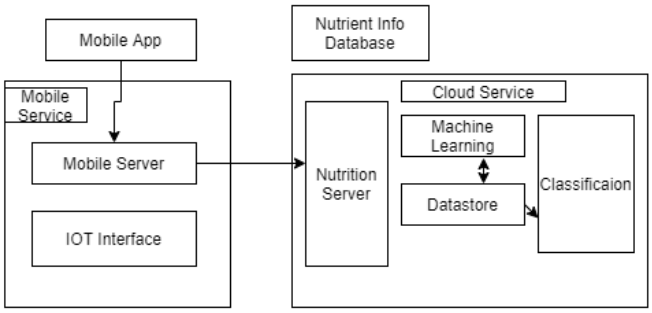

Figure 11 System Architecture for the proposed methodology

B. Implementation for Communication Component:

The communication is done through the API Request; the front-end device, like a web application, will send an HTTP request to send an image to the cloud server. Initially, we establish a connection with the cloud server, and then we create the appropriate HTTP header and fill in the image file with the information; for the transmission, we send the post request to the cloud. We deploy a RESTful Python web server on the Cloud Server GCP using FLASK API, which allows file transfer (image, audio, and video) using HTTP requests [11]. When the server is up and deployed, the port will be listened to and save the requested file to the pre-configured destination. Using fine-trained CNN models, our server will store all the required segments for the deep learning tasks.

C. Backend Implementation:

Our back-end implementation predominantly identifies when we get these pictures from the web platform. We used ImageNet's pre-trained GoogleNet model before testing, fine-

tuned it, and then tuned it to the public food data collection, such as the Food-101 data set. Then, we use the CNN model for the classification of images. When the test food picture is fed into the Convolutional neural network as the data and CNN features are extracted, we first load the model into memory, with layers for max-pooling and slice convolution layer for the vertical detection of the image's vertical positions and acceleration of the convergence of computational working [6].

D. Cloud Implementation:

Google Cloud Platform was used to extract and analyse the data (GCP). As long as a GPU is accessible, any cloud platform (such as Paperspace or AWS) can be used to run the program. The data from the edge service processes enables tracking and generation of user food data that can be used for the classification component. It is also responsible for the food recognition of the data transferred to the machine learning component. The machine learning component will be implemented using FASTAI library using the public dataset of FOOD101. As a layer on top of PyTorch, FASTAI provides extra capabilities for your neural network, such as new techniques for visualizing your data .It has a concept "Callbacks" expand the training utilities by providing additional methods to load and divide data, inferring the number of classes from the dataset you give, and providing more ways to load and split data (which keras also has but PyTorch doesn't). After the classification of the images, the images are sent back to the front-end device, either mobile or web, after food recognition based on input data.

IV. Experimentations and Results

A. Food Datasets:

I have used the dataset Food101. However, it just provides typical photos of fast food in a laboratory. Foodspotting.com photos were used to create a new real-world food dataset. It lets users to snap pictures of what they are eating, annotate the location and type of food, then post these photographs and annotations to the site online. 101 dishes were selected at random from 750 training images. For each class, a further 250 test pictures were taken and manually cleaned. The training pictures were not cleaned on purpose, and as a result, some noise remains. This manifests itself primarily in the form of intense colours and, occasionally, incorrect labels. According to us, if real-world computer vision algorithms are designed for scalability, they should be capable of handling such poorly labelled data. [13]

B. Resnet-34 Architecture:

I have used pre-trained Resnet-34 CNN architecture for classification. Google Cloud Platform (GCP) ran the model in around 1 hour.
ImageDataBunch is used to read the photos into the database. This class in Python does the following functions:

- The picture path is specified, and the training/validation ratio is 80/20.
- Pictures are transformed using the default transformation
- 224 pixels for the image
- 64 pixels for the batch.

I have used pre-trained model resnet34 to train the data, and the learner is used to find the optimal learning rate.
I have used a learning rate of 0.01.

I have used this architecture because the dataset has 101 classes and it takes very long time for the GPU to run the model. As a result, our error rate reduces to 28.9 percent when we fit 8 epochs with our learning rate of 0.01. A new learning rate was then determined by unfreezing parts of the previous layers.

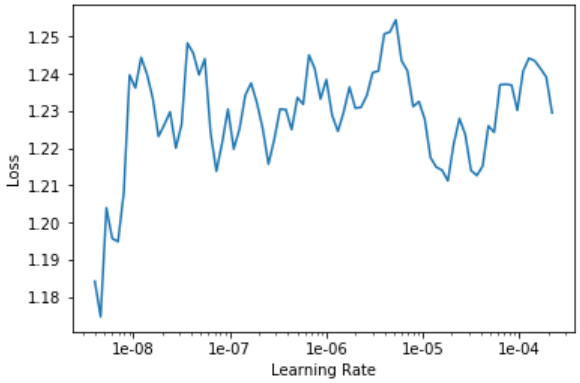

Figure 12: Result of training data.

As a result of training the data for five additional epochs, our error rate dropped from 28.88% to 28.11%.

C. Confusion Matrix:

It was determined by the Confusion Matrix that these foods were the most often misclassified.
The weights file is the model's output. Model.pth is the filename (or final. pth). As in this repo, if you train the model, it is saved in the models folder, which is further used to process the model in the GCP and run a web or mobile application.

D. Heroku App:

I will deploy the model on Heroku to run the web application. The input to the Heroku will be the output of the Model.pth file. I have config,yaml file in Heroku in which I have a path for the code, and I have sample data images. Their names and urls to select these images from the trained model and gives the prediction of the image. The web application will run on Python's Web API Flask, which will take input images from the trained model using Rest API and give the output as a response to the front-end service like a web application.

References

[1] Min, W., Liu, L., Wang, Z., Luo, Z., Wei, X., Wei, X., & Jiang, S. (2020, October). ISIA Food-500: A Dataset for Large-Scale Food Recognition via Stacked Global-Local Attention Network. In Proceedings of the 28th ACM International Conference on Multimedia (pp. 393-401).

[2] *Pouladzadeh, P., & Shirmohammadi, S. (2017). Mobile multi-food recognition using deep learning. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 13(3s), 1-21.*

[3] *Tasci, E. (2020). Voting combinations-based ensemble of fine-tuned convolutional neural networks for food image recognition. Multimedia Tools and Applications, 79(41), 30397-30418.*

[4] *Liang, H., Wen, G., Hu, Y., Luo, M., Yang, P., & Xu, Y. (2020). MVANet: Multi-Tasks Guided Multi-View Attention Network for Chinese Food Recognition. IEEE Transactions on Multimedia.*

[5] *Tomescu, V. I. (2020, May). FoRConvD: An approach for food recognition on mobile devices using convolutional neural networks and depth maps. In 2020 IEEE 14th International Symposium on Applied Computational Intelligence and Informatics (SACI) (pp. 000129-000134). IEEE.*

[6] *Martinel, N., Foresti, G. L., & Micheloni, C. (2018, March). Wide-slice residual networks for food recognition. In 2018 IEEE Winter Conference on Applications of Computer Vision (WACV) (pp. 567-576). IEEE.*

[7] *Chen, H., Wang, J., Qi, Q., Li, Y., & Sun, H. (2017). Bilinear cnn models for food recognition. In 2017 International Conference on Digital Image Computing: Techniques and Applications (DICTA) (pp. 1-6). IEEE.*

[8] *Jeny, A. A., Junayed, M. S., Ahmed, I., Habib, M. T., & Rahman, M. R. (2019, December). FoNet-Local Food Recognition Using Deep Residual Neural Networks. In 2019 International Conference on Information Technology (ICIT)(pp. 184-189). IEEE.*

[9] *Yanai, K., & Kawano, Y. (2015, June). Food image recognition using deep convolutional network with pre-training and fine-tuning. In 2015 IEEE International Conference on Multimedia & Expo Workshops (ICMEW) (pp. 1-6). IEEE.*

[10] *Popovski, G., Eftimov, T., Mladenic, D., & Seljak, B. K. Ontology alignment using Named-Entity Recognition methods in the domain of food.*

[11] *Chang Liu, Yu Cao, Yan Luo, , Guanling Chen, Vinod Vokkarane, Ma Yunsheng, Songqing Chen, and Peng Hou. A New Deep Learning-Based Food Recognition System for Dietary Assessment on An Edge Computing Service Infrastructure. IEEE TRANSACTIONS ON SERVICES COMPUTING, VOL. 11, NO. 2, MARCH/APRIL 2018. IEEE*

[12] *Chang Liu, Yu Cao, Yan Luo, Guanling Chen, Vinod Vokkarane, Yunsheng Ma. DeepFood: Deep Learning-based Food Image Recognition for Computer-aided Dietary Assessment. The University of Massachusetts Lowell, One University Ave, Lowell, MA, 01854, USA*

[13] *Lukas Bossard, Matthieu Guillaumin, Luc Van Gool Food-101 – Mining Discriminative Components with Random Forests. D. Fleet et al. (Eds.): ECCV 2014, Part VI, LNCS 8694, pp. 446–461, 2014.*