Deep Learning Technique for Automatically Classifying Food Images

Dr.K. Uday Kumar Reddy[1]
Associate Professor in CSE M.Tech
AITS, Rajampet, AP, India
pranith.dwh@gmail.com

Sangaraju Swathi[2]
Student in CSE
AITS, Rajampet, AP, India
ssslmp16@gmail.com

Dr.M. Subba Rao[3]
Professor in CSE
AITS, Rajampet,AP, India
msraoswap@gmail.com

Article Info
Page Number: 2362 - 2370
Publication Issue:
Vol 71 No. 4 (2022)

Article History
Article Received: *25 March 2022*
Revised: *30 April 2022*
Accepted: *15 June 2022*
Publication: *19 August 2022*

Abstract

Because of its growing advantages in the health and medical domains, food image categorization is becoming a more popular study topic.Future diet monitoring systems, calorie estimates, and other projects will undoubtedly benefit from automated food identification techniques. This research presents automated systems for classifying foods using deep learning techniques.The classification of food images using Squeeze Net and VGG-16. These networks are suitable for usage in real-world scenarios in the medical and healthcare industries since it has been shown that employing data augmentation and fine-tuning the hyperparameters significantly improved their performance.Because Squeeze Net is a lightweight network, it is simpler to set up and frequently more appealing. VGG-16 can accomplish quite a decent accuracy even with less parameters. Extracting intricate elements from food photographs allows for even higher categorization accuracy. The suggested VGG-16 network considerably enhances the effectiveness of automated food image categorization. Squeeze Net was suggested as having significantly improved accuracy because of increased network depth.
Squeeze Net performs better in the categorization of food images than VGG-16, according to the results. The name of the food item is categorised with pictures that help you identify it.
With deep learning, larger datasets, and more readily available computer resources, image categorization has become less challenging. The most common and widely applied method for classifying images in the present is the convolution neural network. Various transfer learning algorithms are used to classify images from a broad variety of food datasets. Food is important to life since it gives us various nutrients, thus it's important for everyone to keep an eye on their eating patterns. To live a healthier lifestyle, categorising food is so vital. In this project, pre-trained models are employed rather than the more conventional approach of creating a model from scratch, which reduces computing time and costs while also producing superior outcomes.For training and validation purposes, the food dataset is utilised. It consists of several classes, each with many photos. These pre-trained models will be used to identify the provided food, and they will make predictions about its nutritional value based on the colour of the image.

Key words: Food 101 dataset, Food classification, Deep Learning, Transfer Learning, image processing, CNN, VGG-16, Squeeze Net.

INTRODUCTION

Not just for the social network domain element, automatic food identification is a developing study area. Because of the growing advantages it offers from a medical standpoint, researchers are in fact concentrating on this field. The development of diet monitoring systems to fight obesity, as well as the assessment of calories and food quality, will be made easier with the use of automatic food identification technologies [1].

Conversely, food has a great degree of natural deformation and visual variation. Because food photos have low interclassed variance and high intraclass variance, complex characteristics are not recognised by traditional techniques. Since complex aspects cannot be detected by conventional methods, food recognition is a challenging issue. These traits may be quickly and easily recognised by CNNs automatically, improving classification accuracy [2].In this study, we thus attempt to use CNNs to categorise food photos.Even though the photos in Fig. 1 appear to be fairly like one another, they are from two separate dietary categories.Food items interclass variation is minimal, which is the cause of this.

(a) Chocolate Cake **(b) Donuts**

Fig. 1. Food courses with varying types

Convolution, pooling, and fully linked layers make up the core of a convolution neural network.To the input picture, the convolution layer applies biases and weights that may be learned. By summing the features in the input data, the pooling layer reduces the trainable parameters and down samples the data. Having complete connections to all neurons, the completely connected layer is present at the end. The chance of the picture being a member of a certain class was estimated using the SoftMax activation function.

CNNs can readily recognise some of the sophisticated aspects that are not identified by machine learning methods because of the large intraclass variation and low interclass variance present in food photos. By automatically identifying extremely high-level characteristics, these deep learning-based network models have shown tremendous success in improving classification accuracy.As a result, CNNs will be used in the planned study to categorise food images. By applying convolutional operations to specific input data layers, these systems continuously feature extraction to produce feature maps.

Since these systems contain millions of characteristics, training them takes a lot of data and tremendous processing power. Thus, the researchers made the decision to use pre-trained networks that had been improved using data from a certain topic. Utilizing the which was before models' knowledge of relevant data is made possible by the transfer learning technique [3]. Using CNN's Squeeze Net and VGG-16 models, this study examined the classification of food-related pictures. Over a million images from the 1000-class Labelled data were used to train these pre-trained networks. Using the learnt values and characteristics from the pre-trained deep CNN model to Squeeze Net and VGG-16, transfer learning has been done on the Food-101 dataset.

Both conventional methods and deep learning methods are applied to categorise images. Traditional methods can only identify fundamental aspects of a picture, such as colour, form, texture, etc. Traditional machine learning techniques like SVM [5], random forest algorithm [6], and artificial neural networks (ANN) [7] can be used to classify images, although they are less accurate than deep learning techniques.Deep learning algorithms enable the quick identification of deep and complicated characteristics, improving the efficiency of recognition tasks. Among the deep learning techniques that are often used are CNN, learning models, data augmentation, and deep feature fusion networks [8].

The remainder of the document is structured as follows: The specifics of our technique are provided in section 2. The categorization models for foods utilised in this work are discussed in Section 3. Section 4 provides examples of the experiment's findings. The outcomes of the experiment are summarised in Section 5.

METHODOLOGY

The suggested framework to identify food items from the food picture collection is described in this section. The suggested framework for classifying food images is shown in Fig. 2.

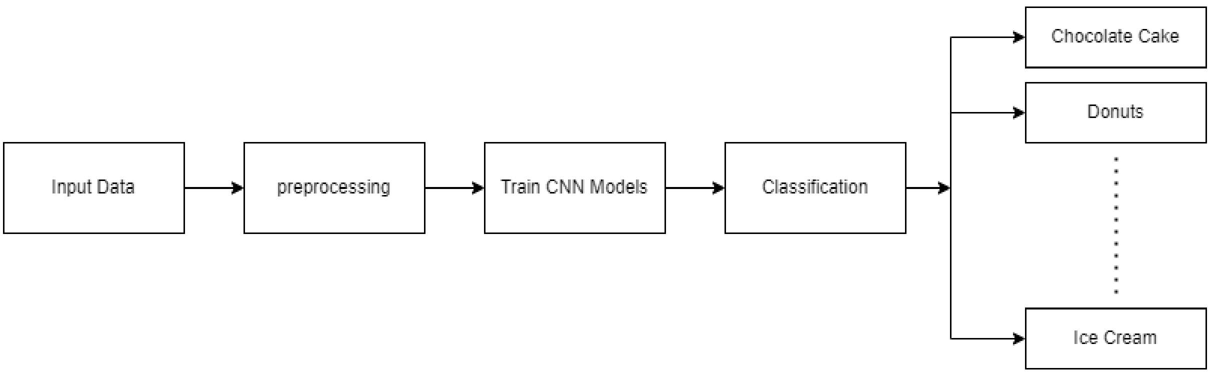


Fig. 2. There is a suggested framework for identifying food from a series of images.

A. Dataset

There are 1000 photographs of each sort of food in the 101,000-photo food-101 collection, which is divided into 101 distinct categories. 5000 images from ten distinct food groups were taken into consideration for categorisation. Cupcakes, French fries, Fried rice, Tee-cream, Omelette, Veggie Pizza, Chocolate Cake, Samosa, and Vegetable rolls are among the ten food categories from the Food-101 dataset. These 10 categories were developed with Indian foodstuffs in mind. To avoid overfitting, this dataset's training data is validated using 30% of the whole dataset.

B. Image Pre-processing

By eliminating undesired distortions and boosting a few key visual elements, image pre-processing enhances the features of an image. Additionally, some of the data augmentation approaches listed below can increase the dataset's efficacy.

Data Augmentation: Data augmentation acts as a regularize, reducing imbalance class problems and preventing overfitting in neural networks [9]. The most common techniques used to alter original pictures are trimming, enlarging, turning, translating, and flipping. Rotation range and random reflection are the methods used in this instance, and they are simply detailed here.

Rotation Range:a 45° angle. Images are randomly rotated within this range. By considering all food picture appearances for this range, this aids in improved performance.

RandomReflection: It is true that the pixel range [-30 30] has random X- and Y-translation. All the changes are made to the original image using an affine transformation, which hasthe next structure:

$$I\,(x, y) \xrightarrow{T} J\,(x', y') \qquad (1)$$

The coordinates of a picture's points are changed using a geometric operation to create a new image. It's a continuous coordinate transform that serves as a mapping function [10].

C. NetworkParameters

The food dataset was trained using the VGG-16 CNN architecture and the Squeeze Net architecture. The table below displays a few of the training factors that were considered by both algorithms to improve performance.

Table 1: Squeeze Net with VGG-16 training parameters

Input Variables	Frequency
Minibatch Size	64
Initial Learn Rate	0.001
Solver	SGDM
Dropout	0.5

Squeeze Net: Complex visual characteristics cannot be accurately identified by conventional machine learning classification techniques. Due to its numerous parameters and greater

depth, CNNs can automatically extract complicated characteristics required in computer vision applications. Squeeze Net is the name of a deep neural network that was first created in 2016 with the intention of having a few parameters and a small design. With a size of 227 MB and 61.0 million parameters, the ILSVRC 2012 champion, Alex Net, is bigger than Squeeze Net. Squeeze Net, on the other hand, is significantly smaller (4.60 MB), has fewer parameters (1.24 million), and has less parameters [11] but still achieves almost the same accuracy on the ImageNet dataset.
With 68 layers and 75 connections, Squeeze Net has 18 layers. The pre-trained network uses fire modules, which compress parameters using [1 x 1] convolutions, as part of its design strategy to minimise the number of parameters. Furthermore, this model can accept input in the following sizes: [227 x 227 x 3] and [3 x 3] convolution with stride [1 x 1]. Utilizing maximum pooling with stride [2 x 2] allows for the reduction of [3 x 3] in size. To reduce overfitting, fire9 module is followed by layers with a 0.5 probability dropout. This network has gaps in each layer's connectivity. The Classification layer outputs the classes, and the SoftMax layer computes the probability after that.
With 64 batches, a 1-validation-per-batch validation frequency, and a learning rate of 0.001, Squeeze Net is trained. A 50% dropout rate occurs during the network's 810 training cycles. Squeeze Net was created as a more transportable replacement that operates around three times quicker and has 50 times less parameters than Alex Net [12].

VGG-16: Due to its superior performance in the ILSVRC-2014, the second model, VGG-16, has been chosen. It is a variant of convolution neural networks that is more advanced but simpler. There are 138 million parameters in VGG-16, which has a depth of 16, and it is 515 MB in size. Images having a size of [224 224 x 3] are used as input to the 41-layer model in question.
Convolutional layer filters have [1 x 1] size and [3 x 3] stride dimensions. The stride [2 x 2] technique makes use of Max Pooling across a [2 x 2] pixel window. The last three levels have been improved to meet the issue with the new categorization [13]. To lessen the possibility of vanishing gradient problems, ReLU activation function is used in all hidden layers. 64 batches were used to train the VGG-16 at a learning rate of 0.001. 50 was chosen as the validation frequency. Following that, the network is trained for 408 cycles with a 50% dropout rate.

D. Transfer Learning Models

Transfer learning will be used in place of building the network from scratch, using a pre-trained model. It was created using the learnt weights and trained on the ImageNet dataset [14]. The pre-trained network is utilised for transfer learning. Transferring information from prior, similar tasks helps students learn more in the current activity. Only the last layers, which make predictions based on specific visual components, are trained.

E. Food Classification

Following is a description of the CNN network setup used to classify food images:

Convolution Layer: To extract a feature map, the convolution process is used in this layer. Filter size and stride are some of its properties. ReLU is an activation function that turns all negative values into zero and is utilised in hidden layers. It increases nonlinearity. Because ReLU is easier to train and frequently produces better performance, it is used [15].

$F(x) = \max(0, x)$ (2)

Pooling Layer: The depth and many parameters of CNNs make computations using them costly. To reduce the dimensionality between the layers, it is necessary. By using down sampling, the pooling layer does this.

Fully Connected Layer:

Typically, this layer appears at the conclusion of the CNN architecture. Each input is coupled to every neuron in a single-dimensional vector created by this layer, which transforms input.

RESULTS

The experiment employed an Intel Core i7 CPU running at 2.60 GHz, an NVIDIA GeForce GTX graphics card, and 8.00 GB of RAM, together with the programming language MATLAB and a Windows 10 64-bit machine. 64 batches and a 0.001 beginning learning rate were used during the 810 rounds of training the Squeeze Net model on a GPU. Classification precision was used to assess the efficacy of the fine-tuned Squeeze Net model. With a 93.47 percent training accuracy and a 77.20 percent validation accuracy, the model outperformed conventional machine learning models significantly. The Squeeze Net classifier's evolution during training is seen in Figure 3.

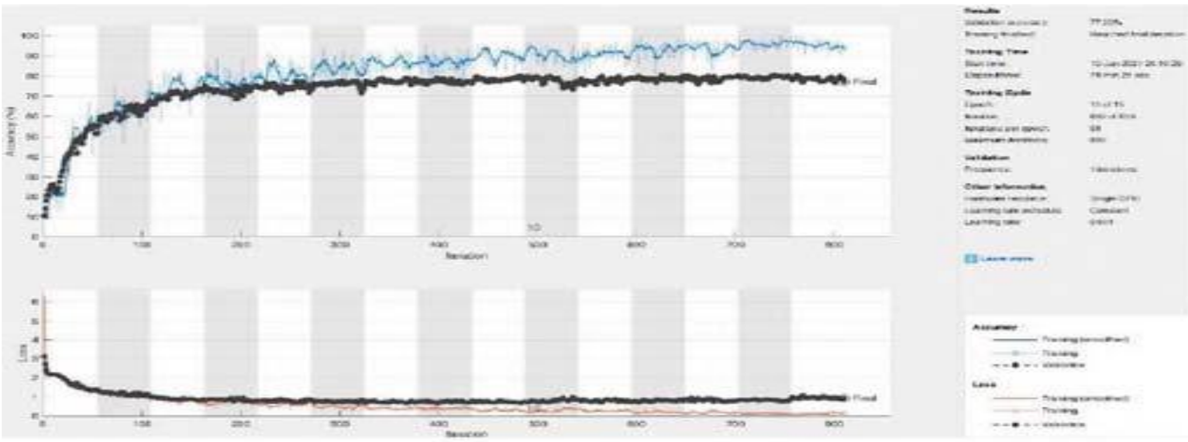


Fig. 3. Training Progress of SqueezeNet Classifier

The Food-101 dataset, which includes the top 10 classes of Indian food, uses VGG-16 as the second model for classifying food images. A better level of categorization accuracy benefits the Squeeze Net model. With the same learning rate and batch size as Squeeze Net, the proposed VGG-16 model was created (0.001 and 64, respectively). With 823 minutes and 408 CPU training cycles with 50 validation intervals, the model was able to classify objects with an accuracy of 85.07 percent. Because of the deeper network and more parameters, the suggested VGG-16's accuracy showed a significant improvement.

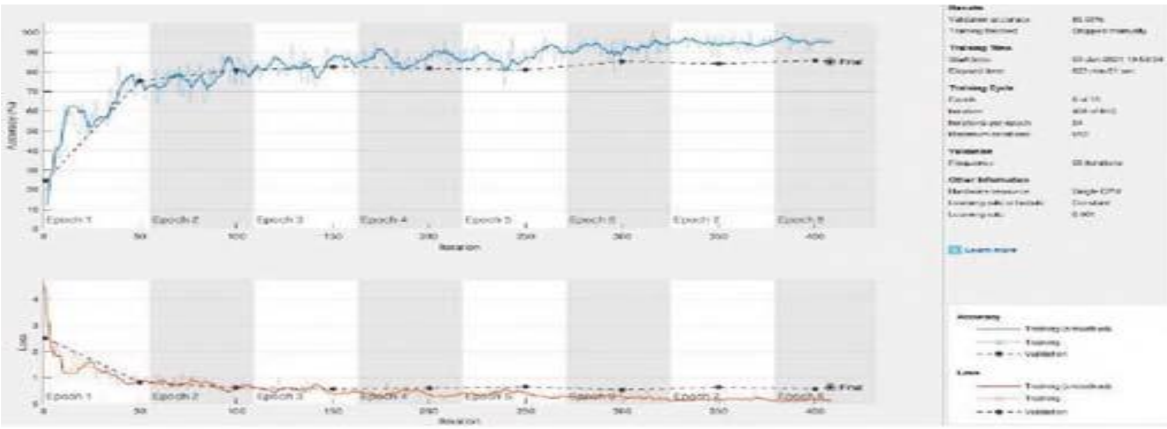
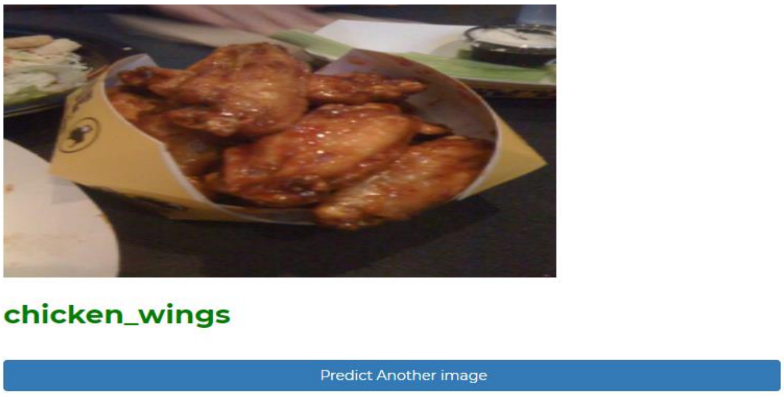


Fig.4. Training Progress of ProposedVGG-16 Classifier

Fig. 5(a) demonstrates the projected class labels for food photos selected at random.



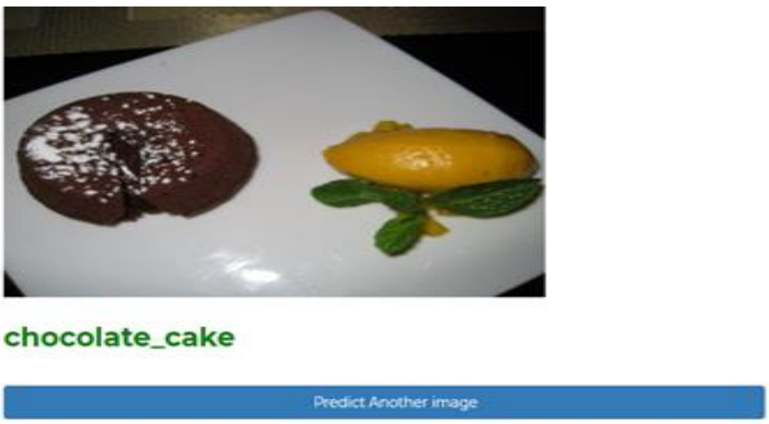


Fig. 5. (b) Randomly selected food photos with predicted class labels

CONCLUSION AND FUTURE ENHANCEMENTS

Using deep learning techniques, automated methods for classifying food images have been provided in this study. Extracting high-level complex characteristics led to improved performance in the categorization of food images. To do this, deep learning models SqueezeNet and VGG-16 have been applied. These networks' hyperparameters were optimised during design to increase network performance. Data augmentation techniques were also applied. With a much smaller model and fewer parameters, SqueezeNet worked admirably, achieving an accuracy of 77.20 percent. Squeeze Net cannot compare to the planned VGG-16's depth and parameterization. Therefore, the recommended VGG-16 has dramatically improved performance and a higher classification accuracy of 85.07 percent for food images.

The food images are categorised into the appropriate classifications in this study using the Convolutional Neural Network, a deep learning approach. The categorization issue could be improved by eliminating noise from the dataset in terms of potential future improvement. It is possible to do the same study with a larger dataset, more classes, and more photos in each class.This is since a larger dataset reduces the loss rate and helps the algorithm learn new characteristics, increasing accuracy. The weights of the model may be used to build online or mobile applications for categorising images and, in addition, for calorie extraction from the food that has been classified.

When calories are calculated for multi-food and complicated food items, the suggested approach will be enhanced, allowing users of our programme to understand the complexity of food more fully. Additionally, the development of the dataset to include a wider variety of food kinds will enhance the system's performance and accuracy. The computational portion of the model that requires more time may be offloaded to the cloud as quick reaction is one of the most important factors now. A model that uses the cloud may be obtained that uses huge calculations to provide results more quickly.

Real-time food photographs were first gathered for the project from a variety of sources. A significant number of food photographs were used in the convolution neural network-based model's training, which improves your model's capacity to acquire the necessary features rapidly. The accuracy of the training dataset of pictures acquired in the outcome analysis is roughly 84%. The necessity for a big dataset to train the CNN and the fact that even a well-trained network cannot achieve 100% segmentation accuracy will cause the recognition error rate to increase in the future. We may thus create a bigger dataset that contains a variety of food photographs to obtain a better outcome.

Due to the lack of knowledge regarding diet and calorie requirements, it is essential to build up a system that records daily food consumption for a balanced diet. It might be difficult to identify meals correctly. So, in order to calculate the amount of calories from various food photographs, we devised a measurement approach that involves analysing and quantifying the features of the image, such as the colour of the food. We have created a reliable method for classifying fruits.

REFERENCES

[1]. Rajayogi, J.R., Manjunath, G. and Shobha, G., 2019, December. Indian Food Image Classification with TransferLearning. In 2019 4th Inter-national Conference on Computational Systems and Information Technology forSustainable Solution (CSITSS) (Vol. 4, pp. 1-4). IEEE.

[2]. Reddy, V.H., Kumari, S., Muralidharan, V., Gigoo, K. and Thakare, B.S., 2019, May. Food Recognition andCalorie Measurement using Image Processing and Convolutional Neural Network. In 2019 4th In-ternationalConference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT) (pp. 109-115).IEEE.

[3]. Subhi, M.A. and Ali, S.M., 2018, December. A deep convolutional neural network for food detection andrecognition. In 2018 IEEE-EMBS conference on biomedical engineering and sciences (IECBES) (pp. 284-287).IEEE.

[4] Burkapalli, V.C. and Patil, P.C., TRANSFER LEARNING: INCEPTION-V3 BASED CUSTOMCLASSIFICATION APPROACH FOR FOOD IMAGES.

[5] Fruit Recognition and its Calorie Measurement: An Image Processing Approach ManpreetkourBasantsinghSardar1, Dr. Sayyad D. Ajij2 (2016)

[6] Raikwar, H., Jain, H. and Baghel, A., 2018. Calorie Estimation from Fast Food Images Using Support Vector
Machine. International Journal on Future Revolution in Computer Science & Communication Engineering, 4(4),pp.98-102.

[7] Subhi, M.A. and Ali, S.M., 2018, December. A deep convolutional neural network for food detection andrecognition. In 2018 IEEE-EMBS conference on biomedical engineering and sciences (IECBES) (pp. 284-287).IEEE.

[8] Zhang, W., Zhao, D., Gong, W., Li, Z., Lu, Q. and Yang, S., 2015, Au-gust. Food image recognition withconvolutional neural networks. In 2015 IEEE 12th Intl Conf on Ubiquitous Intelligence and Computing and 2015 IEEE 12th Intl Conf on Autonomic and Trusted Computing and 2015 IEEE 15th Intl Conf on Scalable Computing and Communications and Its Associated Workshops (UIC-ATC-ScalCom) (pp. 690-693). IEEE.

[9] Pathanjali, C., Salis, V.E., Jalaja, G. and Latha, A., 2018. A Comparative Study of Indian Food ImageClassification Using K-Nearest-Neighbour and Support-Vector-Machines. J. Eng. Technol, 7, pp.521-525.

[10] Christodoulidis, S., Anthimopoulos, M. and Mougiakakou, S., 2015, September. Food recognition for dietary assessment using deep con-volutional neural networks. In International Conference on Image Analysis and Processing (pp. 458-465). Springer, Cham