Food Image Classification Using Various CNN Models

Rudraja Vanutre
B.Tech (CSE)/Shri Shankaracharya Institute of Professional Management and Technology

***Abstract*— Food picture classification is a burgeoning area of research since it is becoming more and more significant in the health and medical fields. Deep learning models are now often employed for picture recognition and categorization. Image categorization is a common research issue in the areas of machine learning, computer vision, and image processing. There are numerous unique approaches to classifying and identifying foods that have been established in recent research papers.**

***Index Terms*—CNN, Deep Learning, Food-101, image classification, pre-trained CNN.**

I. INTRODUCTION

Pictures are essential in today's world because they carry a lot of information. Finding meaningful image information within the effective time is critical for big photos. As a result, the image classification algorithm's high performance has an impact on the picture classification results. [1].

By using data to train the computer, image classification was developed to close the gap between computer vision and human vision. [2]. Due to the growth in data in a number of industries, including e-commerce, automotive, healthcare, and gaming [3], it has recently become more popular among technology developers. Numerous application gateways in numerous fields have been made possible by access to this data and various image categorization methods and algorithms.

Among these domains is the visual perception of foods. Its uses include tracking a person's eating patterns and automatically detecting food waste. It can also anticipate how many calories will be ingested. Sorting foods into categories It is challenging to capture food images because they depend on a variety of technical, environmental, and technical environmental elements such as illumination, image quality, noise level, etc. The same food can have varied shapes, sizes, and colours. When the image consists of a group of one or more foods, the task is made considerably more challenging. Food is therefore highly deformable, with a high intraclass variance and a low interclass variance. [4]. Machine learning is used to classify foods in the conventional way. Edges, textures, and other features are retrieved using a feature attraction module in machine learning, and then they are categorised using a classification module. The main disadvantage of utilizing ML is that it can only extract a limited selection of characteristics from the data set, not the distinguishing traits. Deep Learning is employed to get around this. [2]

In order to analyse data having a grid-like design, such as an image, convolutional neural networks (CNNs) are used. Fully connected, pooling, and convolution layers make up the three layers of a convolution neural network. The input picture is subjected to learning biases and weights in the convolution layer.

The pooling layer minimizes trainable parameters by down sampling the input data and averaging the features. There is a fully linked layer with complete connections to every neuron at the very end. The Softmax activation function was used to calculate the likelihood that a given image belongs to a particular class. [5]

II. RELATED WORK

Popular deep learning research areas include pattern recognition and picture categorization.
Since it may be used to calculate food calories and investigate people's eating patterns for health-care purposes, food picture identification and classification is one of the most potential applications in the field of image and object recognition, drawing the attention of several researchers. There has been a great deal of study in this area, and many papers have been presented. The bulk of current work on the categorization of food images has employed deep learning.

The Pittsburgh Fast-Food Image Dataset (PFID), which consists of 4545 still photos, 606 stereo image pairs, 303 360 food clips, and 27 eating films of 101 different items, was created in 2009 to promote further research in the area. The color histogram methodology had a classification accuracy of 11% and the bag-of-SIFT

(Scale-Invariant Feature Transform)-features method had a classification accuracy of 24% when a Support Vector Machine (SVM) classifier was used on this dataset. [7].

On the basis of the multiple segmentation hypotheses, a segmentation and classification framework was used in [8]. To do this, it chooses the most effective segmentations based on the confidence ratings each sector receives. The approach combines efficient techniques to create several divisions of segmentation and gives an iterative stability metric to assign the best class label to each picture pixel is 44% accurate on average.

The components of food picture analysis are also examined, along with their combinations, using a classification technique based on word trees and k-nearest neighbors in [9]. 42 different categories and 1453 photos are used in the testing. According to the study, the performance of food classification was increased by 22% for Top 1 classification accuracy and 10% for Top 4 classification accuracy by adding three features, DCD, MDSIFT, and SCD.

Deep convolutional neural networks (DCNN) effectiveness in identifying foods from photographs was examined in [10]. It makes use of a variety of DCNN-related techniques, such as activation characteristics gleaned from the pre-trained DCNN and pre-training on large-scale ImageNet data. With classification accuracy rates of 78.77 percent and 67.57 percent for the UEC-FOOD100/256 dataset, it had the greatest classification success rate.

In [11], the efficiency of the deep-learning technique Inception for classifying food images was evaluated. Inception is based on the requirements of Google's image recognition framework. To categorize photos from the ETH Food-101, UEC FOOD100, and UEC FOOD 256 food image datasets, it uses a Deep Convoluted Neural Network (DCNN) with 54 layers that has been fine-tuned. It had accuracy scores of 88.28 percent, 81.45 percent, and 76.17 percent on the top-1 and 96.88 percent, 97.27 percent, and 92.58 percent on the top-5.
In [12] improved the Inception V3 model, which on the 3,960 picture TFF dataset had a classification accuracy of 88.33% and was trained on the ImageNet dataset.

In order to quantify calorie intake, a method is presented in [13] for identifying and classifying high-calorie fast food snacks from the test photographs. It trains an image category classifier using a feature extractor that has already been trained on a convolutional neural network (CNN). The components of food picture analysis are also examined, along with their combinations, using a classification technique based on word trees and k-nearest neighbors in [9]. The tests employ 1453 images across 42 categories. The study found that adding our three features, DCD, MDSIFT, and SCD, increased food classification performance by 22% for Top 1 classification accuracy and 10% for Top 4 classification accuracy.

In [14], a pre-trained CNN model is utilized to train the CNN classifier. It makes use of a data set on Indian food that consists of 20 distinct classes with 500 photos each. The models used are ResNet, IncceptionV3, VGG16, and VGG19. With a data loss of 0.5893, it had an accuracy of 87.9%.

The FCN, ENet, SegNet, DeepLabV3+, and Mask RCNN algorithms used in our methodology are compared in [15]. It makes use of a dataset comprising 1250 pictures and 9 classes, including the most well-liked Brazilian food subcategories.

In [16], a food genre classification model is presented and developed using CNN, the Python Tensor Flow library, Keras, and a data collection of 17 genres and 170 photos. A 92.9% accuracy rate was achieved.

In [17], smartphone software was developed to recognize food items from live images. The smartphone software uses a trained deep CNN model to recognize food items from live photos. In addition, we combine two CNN architectures to create a fresh deep convolutional neural network (CNN) model for food recognition. Through ensemble learning, the new deep CNN model was created. A unique dataset for food recognition is used to train the deep CNN food identification model. 29 different food and fruit types make up the tailored food identification dataset. This model was able to achieve an accuracy rating of 95.55%.

In [4], the most popular deep learning techniques for classifying foods are examined. Public food databases are also shown, along with the results of a food

classification experiment that had an accuracy of 90.02% and was conducted across five trials. It makes use of the ResNeXt-101, DenseNet-161, and UEC Food 100 databases.

For the segmentation, recognition, and calorie calculation of food items, [18] proposes a hybrid network built on GAN and CNN. Food ROIs were calculated with Pix2Pix GAN and the extracted food RoIs were then categorized using ResNet50. It has a 95.21% accuracy rate.

III. CONVOLUTION NEURAL NETWORK

The A neural network's input layer, hidden layers, and output layer are its fundamental building components. The structure of CNNs is patterned after that of the human brain. Artificial neurons, or nodes in CNNs, receive inputs, analyse them, and offer the result as output, much like a neuron in the brain does when it distributes information throughout the body. The input is a picture. The input layer is capable of accepting image pixel arrays as input. There might be a number of hidden layers in CNNs that employ mathematics to extract data from the image. This may be seen in many different ways, such as convolution, pooling, rectified linear units, and completely linked layers. In order to extract features from an input picture, convolution is the first layer. In the output layer, the fully connected layer classifies and recognizes the item.

A. Pooling Layer

Shift occurs at every activation following the application of each layer's filter, which has a specified size and step. Max Pooling is the most often used pooling technique. If the maximum pooling is 4*4, then there are four phases. Each filter will have a maximum pooling value, and average pooling will choose the average value. If the structure or dimensions of any data are large, we can employ pooling layers with the convolutional layer to convert the high-dimensional feature map created by the convolutional layer to the low-dimensional form, and the remaining computational work will only take little effort.

By pooling layers, the featured map is summarized, negating the need for the model to be trained on precisely positioned features. As a result, a model becomes more reliable and sturdy.

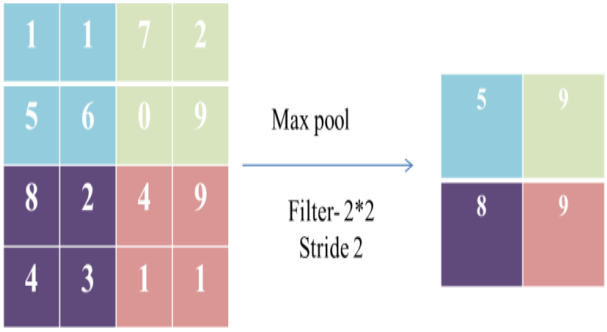

Figure 1 Pooling Layer

B. Dropout Layer

The regularization method used to stop model over fitting is called dropouts. The network's neurons are randomly switched along with dropouts in a specified proportion. Neurons are also shut off when their incoming and outgoing connections are. To enhance the model's learning capabilities, this is done. A neural network is compelled by dropout to acquire more dependable characteristics that function well with several different random subsets of the other neurons. The number of convergence iterations required is more than doubled when dropout occurs. The training period is reduced with each epoch, though.

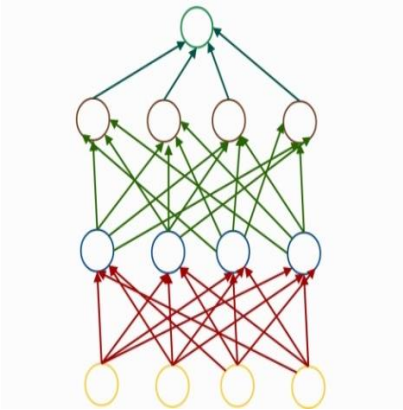

Figure 2 Normal Neural Network

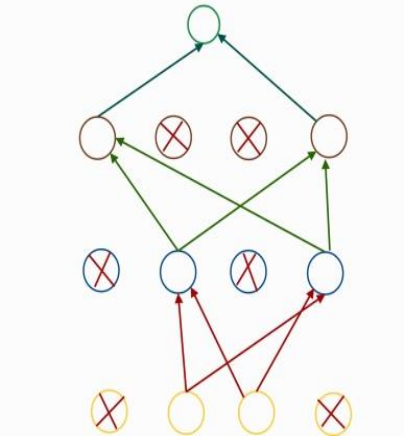

Figure 3 After Dropout

C. ReLu Activation Function

The (ReLU) enactment work is the initiating work that is typically used for the CNN model. Rectified Linear Unit is a CNN model implementation that uses the following capabilities: According to the formula f(x) = max (0, x) (2), this capability applies thresholding with a value of 0 to each pixel's value in the information picture. This action causes any pixel values in a picture that are less than 0 to be changed to 0.

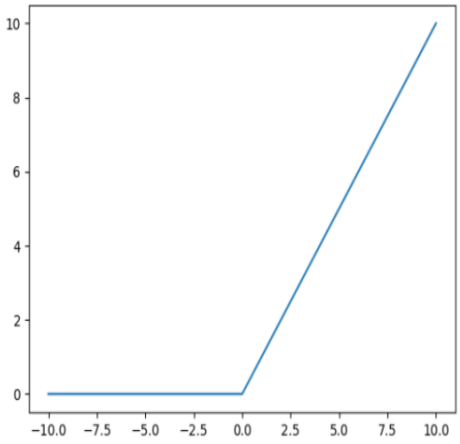

Figure 4 ReLu Activation Function

C. Cross Entropy Loss Function

Another term for the cross entropy function is the logistic, logarithmic, or logarithmic loss. A score/loss is calculated that penalizes the probability in proportion to how much it deviates from the actual expected value for each class after comparing its projected probability to the real class' intended output, which can be either 0 or 1. The penalty is logarithmically calculated, thus significant discrepancies close to 1 result in a large score while minor differences close to 0 result in a small score. Cross-entropy loss is employed to modify the model weights during training. Since loss reduction is the aim, a better model will result in a lesser loss. A perfect model has zero cross-entropy loss.

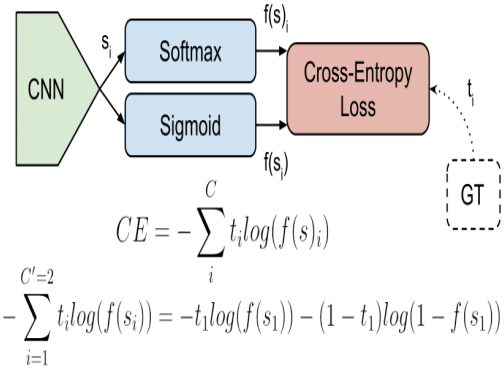

Figure 5 Cross Entropy Loss Function

IV. DATASET

The benchmark dataset Food-101 has 101 food categories and 101,000 photos. 250 exam photos and 750 practice images are offered for each class. Beef tartare, chicken curry, chocolate mousse, French breakfast, fried rice, hot dogs, ice cream, lasagna, oysters, pizza, and takoyaki are among the 11 classes we'll be training. 2750 photographs are in the test folder and 8250 are in the train folder.

Figure 6 Random images from each class

V. RESULT

This paper examines several pre-trained cnn models on traditional Food-101 benchmark dataset. Results from various cnn models are stated in the table below.

TABLE 1 : ACCURACY OF EACH CNN MODEL

CNN Model	Classes	Accuracy
MobileNetV2	11	92.50%
InceptionV3	11	93.89%
Efficient Net B2	11	93.25%
Resnet152	11	93.79%
Resnet50	11	92.46%

VI. ACKNOWLEDGEMENT

I would like to express my gratitude to Dr. Pradeep Singh (Asst. Professor, CSE, NIT Raipur) for providing his expertise and guidance throughout this project and providing me with this opportunity to learn and work on this project during the course of my internship at NIT Raipur.

VII. REFERENCES

[1] P. Wang, E. Fan, and P. Wang, "Comparative analysis of image classification algorithms based on traditional machine learning and deep learning," Pattern Recognit. Lett., vol. 141, pp. 61–67, 2021, doi: 10.1016/j.patrec.2020.07.042.

[2] M. M. Krishna, M. Neelima, M. Harshali, and M. V. G. Rao, "Image classification using Deep learning," Int. J. Eng. Technol., vol. 7, no. March, pp. 614–617, 2018, doi: 10.14419/ijet.v7i2.7.10892.

[3] M. A. Abu, N. H. Indra, A. H. A. Rahman, N. A. Sapiee, and I. Ahmad, "A study on image classification based on deep learning and tensorflow," Int. J. Eng. Res. Technol., vol. 12, no. 4, pp. 563–569, 2019.

[4] B. Arslan, S. Memis, E. B. Sonmez, and O. Z. Batur, "Fine-Grained Food Classification Methods on the UEC FOOD-100 Database," IEEE Trans. Artif. Intell. vol. 3, no. 2, pp. 238–243, Aug. 2021, doi: 10.1109/tai.2021.3108126.

[5] S. Yadav, Alpana, and S. Chand, "Automated Food image Classification using Deep Learning approach," in 2021 7th International Conference on Advanced Computing and Communication Systems, ICACCS 2021, Mar. 2021, pp. 542–545. doi: 10.1109/ICACCS51430.2021.9441889.

[6] M. Chen, K. Dhingra, W. Wu, L. Yang, R. Sukthankar, and J. Yang, "PFID : PITTSBURGH FAST-FOOD IMAGE DATASET Intel Labs Pittsburgh , 2 Carnegie Mellon University , 3 Columbia University," Baseline, pp. 289–292, 2009.

[7] F. Zhu, M. Bosch, N. Khanna, C. J. Boushey, and E. J. Delp, "Multilevel Segmentation for Food Classification in I)ieta\ssessment," 2011.

[8] Y. He, C. Xu, N. Khanna, C. J. Boushey, and E. J. Delp, "ANALYSIS OF FOOD IMAGES : FEATURES AND CLASSIFICATION School of Electrical and Computer Engineering , Purdue University Department of Electronics and Communication Engineering , Graphic Era University Cancer Epidemiology Program , University of Hawaii Can," pp. 2744–2748, 2014.

[9] K. Yanai and Y. Kawano, "FOOD IMAGE RECOGNITION USING DEEP CONVOLUTIONAL NETWORK WITH PRE-TRAINING AND FINE-TUNING Keiji Yanai Yoshiyuki Kawano Department of Informatics , The University of Electro-Communications , Tokyo , Japan," Int. Conf. Multimed. Expo Work. . IEEE, pp. 1–6, 2014.

[10] H. Hassannejad, G. Matrella, P. Ciampolini, I. De Munari, M. Mordonini, and S. Cagnoni, "Food image recognition using very deep convolutional networks," MADiMa 2016 - Proc. 2nd Int. Work. Multimed. Assist. Diet. Manag. co-located with ACM Multimed. 2016, pp. 41–49, 2016, doi: 10.1145/2986035.2986042.

[11] N. Hnoohom and S. Yuenyong, "Thai fast food image classification using deep learning," in 1st International ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering, ECTI-NCON 2018, Jun. 2018, pp. 116–119. doi: 10.1109/ECTI-NCON.2018.8378293.

[12] A. B. Akhi et al., "Recognition and Classification of Fast Food Images," 2018.

[13] J. R. Rajayogi, G. Manjunath, and G. Shobha, "Indian Food Image Classification with Transfer Learning," CSITSS 2019 - 2019 4th Int. Conf. Comput. Syst. Inf. Technol. Sustain. Solut. Proc., pp. 17–20, 2019, doi: 10.1109/CSITSS47250.2019.9031051.

[14] C. N. C. Freitas, F. R. Cordeiro, and V. MacArio, "MyFood: A Food Segmentation and Classification System to Aid Nutritional Monitoring," Proc. - 2020 33rd SIBGRAPI Conf. Graph. Patterns Images, SIBGRAPI 2020, pp. 234–239,2020,doi: 10.1109/SIBGRAPI51738.2020.00039.

[15] R. R. G. Hquh et al., ")Rrg *Hquh &Odvvlilfdwlrq Iurp)Rrg ,Pdjhv E\ 'Hhs 1Hxudo 1Hwzrun Zlwk 7Hqvruiorz Dqg .Hudv".

[16] A. Fakhrou, J. Kunhoth, and S. Al Maadeed, "Smartphone-based food recognition system using multiple deep CNN models," Multimed. Tools Appl., vol. 80, no. 21–23, pp. 33011–33032, Sep. 2021, doi: 10.1007/s11042-021-11329-6.

[17] R. Jaswanthi, E. Amruthatulasi, C. Bhavyasree, and A. Satapathy, "A Hybrid Network Based on GAN and CNN for Food Segmentation and Calorie Estimation," in 2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), Apr. 2022, pp. 436–441. doi: 10.1109/ ICSCDS 53736. 2022.9760831.

[18] P. Tripathi, "TRANSFER LEARNING ON DEEP NEURAL NETWORK : A CASE STUDY ON FOOD-101 FOOD CLASSIFIER," vol. 5, no. 9, pp. 229–232, 2021.