Recognizing and Classifying Indian Foods by State Using CNN and MobileNetV2

Dr Shashikala Parameshwarappa[#1], Dr B.V.Dhandra [*2]

Computer Science and Engineering Department
Government Engineering College ,Raichur-584135
State: Karnataka, Country:India
Dr B.V. Dhanrda
Department of Master of Computer Application
Gulbarga University Gulbarga- 58102

Abstract— **A method for automatically recognizing food from different food images was made available. Our project's main objective is to propose a novel method for automatic food recognition. Food_classification, which comprises various photos of food dishes in 12 distinct classes, was used in this dataset. There are classes on Indian cuisine. In this research, we present a transfer learning based MobileNetV2 model with a custom-built CNN model for food recognition and classification.To estimate the number of calories in the identified food image, a method for calorie estimation is presented, which takes into account both nutritional data and image attributes. There are 12 classes in the collection, each having 100 photos of Indian cuisine. Following testing, it was discovered that the MobileNetV2 model performed 77% more accurately than the bespoke CNN model. Indian Thali and particular meals unique to each state are also categorized. Additionally, calories are being tracked.**

Keywords—**Food, CNN model, MobileNetV2, Indian cuisine, calorie**

I. Introduction

With open-ended continuous learning, new and updated visuals of current lessons arise frequently. There are two major forms of incremental learning that are relevant to open-ended continuous learning: 1) With regard to progressive data learning 2) Instruction in small groups. By employing newly accessible photographs, data incremental learning enhances current class recognition performance and adjusts to domain changes. On the other hand, class incremental learning continuously picks up knowledge from new classes.

Similar to the two elements of human learning, these two forms of incremental learning are necessary for acquiring new ideas and enhancing present class categorization abilities.

This illustrates the usefulness of open-ended continuous learning in a variety of real-world recognition tasks, including food recognition, since the dataset is dynamic and new concepts of interest appear over time.

Users regularly add food photos from their smartphones to both new and current classes for food recognition.

It is crucial to expect that various users will view the same class from different perspectives since users' views of existing classes vary. Photos show notable intra-class variance as well as inter-class similarities, according to the food datasets used to assess the suggested framework.

.

II. Related Work

S. Rousseau [1], Food "porn" refers to still or moving images of food and/or eating that appear in a variety of media, such as cookbooks, magazines, television, blogs, websites, and social media platforms such as Twitter, Facebook, Instagram, and Pinterest.

T. C. Chen[2], Intelligence technology is an important area of study. Hand-held object recognition, as a very special but important case of object recognition, plays an important role in intelligence technology for its many applications such as visual question-answering and reasoning. In real-world scenarios, datasets are open-ended and dynamic, with new object samples and classes added on a regular basis. To efficiently learn new information, intelligence technology must enable hybrid incremental learning, which supports both data-incremental and class-incremental learning. D. Maltoni and V. Lomonaco, [4],A new approach, denoted as AR1, combining architectural and regularization strategies is then specifically proposed. AR1 overhead (in term of memory and computation) is very small thus making it suitable for online learning. When tested on CORe50 and iCIFAR-100, AR1 outperformed existing regularization strategies by a good margin. G. Ditzler, M. Roveri,

U. C. Alippi, and R. Polikar,[5]The nonstationarity can be due, for example, to seasonality or periodicity effects, changes in the users' habits or preferences, hardware or software faults affecting a cyber-physical system, thermal drifts or aging effects in sensors. In such nonstationary environments, where the probabilistic properties of the data change over time, a non-adaptive model trained under the false stationarity assumption is bound to become obsolete in time, and perform sub-optimally at best, or fail catastrophically at worst.J. D. Power and B. L. Schlaggar[6]This essay surveys some of the key concepts related to neural plasticity, beginning with how current patterns of neural activity (e.g., as you read this essay) come to impact future patterns of activity (e.g., your memory of this essay), and then extending this framework backward into more development-specific mechanisms of plasticity.

A. Proposed System

Utilizing Transfer Learning and a Convolutional Neural Network (CNN) model, the suggested work is implemented in Python. The models were used to predict food class after they had been trained on 1000 photos over 12 classes. Before a fresh input is sent to the trained model, it passes through various picture processing phases, such as scaling and color space conversion. The output is anticipated by comparing each learned class's features with those of the input.

.

III. Methodology

CNN, or Convolutional Neural Network One type of deep learning neural network is the convolutional neural network (CNN). The model is composed of.MaxPool, Convolution, Dropout, Fully Connected Layer, Conv2D Layer, and ReLu Function. An input layer, an output layer, and a few hidden layers in between make up a Deep Neural Network (DNN).

These networks are capable of handling non-linearity in addition to unstructured and unlabeled data. The 'DNN' model wrapper offered by TFLearn is capable of automatically carrying out neural network classifier operations like training, prediction, and save/restore.

The proposed CNN Structure consists of:

Conv2D - The first Convolutional 2D layer consists of 32 kernels of 3x3. Takes an input of size 50x50x3 where 50x50 is the rescaled size of images from the dataset. RGB, the color aspect of the image is denoted by

Convolutional Layer - A (50,50,3) input size is used by convolution function and this layer generates the feature maps through convolving the input data.

$$W_{out} = \frac{W - F + 2P}{S} + 1$$

Formula for Convolution Layer

Pooling Layer - The second layer with a pool size of 2x2 is the max-pooling layer. For better feature extraction, these layers are repeated once again. Then, to get more filtered images for the fully connected layers, the kernel’s size is increased from 32 to 64.

$$W_{out} = \frac{W - F}{S} + 1$$

Fully Connected Layer - This is used to connect all neurons to a one layer as well as to another layer, which works on the basis of traditional multi-layerpreceptor(MLP) neural networks. Two fully connected layers are used next with 128 and 90 neurons respectively.

$$z_l = W_l * h_{l-1}$$

Dropout layer - To prevent overfitting, dropouts have been added in between the dense layers.

ReLu - Activation functions are mathematical equations that determine the output of a neural network model. All the convolutional 2D layers and the fully connected layers have an activation function of Rectified Linear Unit (ReLu).

$$\boxed{\mathrm{ReLU}(z_i) = \max(0, z_i)}$$

The network is configured to output 12 values, one for each class in the classification task, and the softmax function is used to normalize the outputs. Adam optimization algorithm was used and the learning rate was configured as 0.0001.

MobileNetV2: is a convolutional neural network architecture that seeks to perform well on mobile devices. It is based on an inverted residual structure where the residual connections are between the bottleneck layers.MobileNet has been pre-trained on the ImageNet database which contains more than a million images. It captures the edges, color and pattern in initial layers and complex patterns related to our task in the final layers. The proposed work uses a mobilenet model for the purpose of classification. Fig - 3:Visualization of proposed MobileNetV2 Architecture The architecture of MobileNetV2 contains the initial fully convolution layer with 32 filters, followed by 19 residual bottleneck layers. The primary network (width multiplier 1, 224 × 224), has a computational cost of 300 million multiply-adds and uses 3.4 million parameters. Instead of adding fully connected layers on top of the feature maps, an average of each feature map is taken, and the resulting vector is fed directly into the softmax layer

Calorie Estimation :

By default, the image sequences are recorded in BGR color space. In BGR(Blue,Green,Red), red takes up the least amount of space, followed by green and blue. Hue Saturation Value, or HSV, space is created by converting the BGR color space.

This is achieved by a function of the OpenCV module called cvtColor(using COLOR_BGR2HSV.The idea of indexed images—a direct mapping of pixel values to colormap values—is used by the algorithm. Nutrition websites [11] provide the calorie range for each food class, which is then utilized in conjunction with the HSV approach to estimate the calories of the food image.

IV. System Architecture

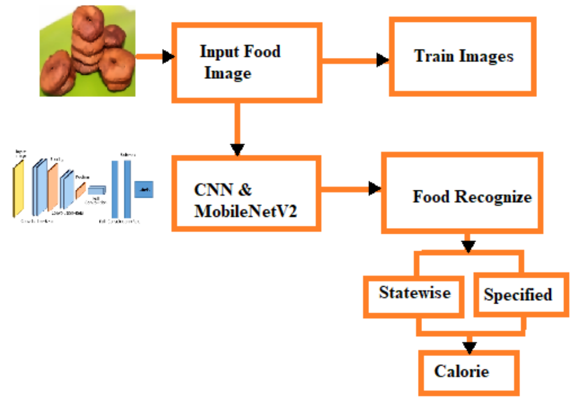


Figure 1: System Architecture

In the above architecture, the system accepts test food image, applies train images, using CNN & mobileNetV2 model the system recognizes the food by state wise as well as specified food. Then finally displays the calorie of that food.

V. Implementation

Figure 2: Main

It's a main page of the program

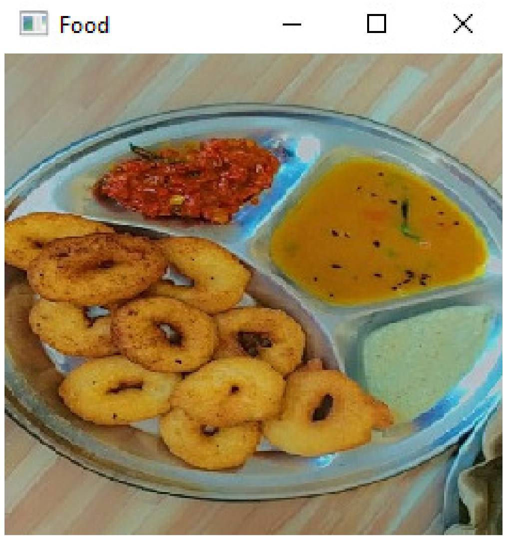


Figure 3 :Read Image

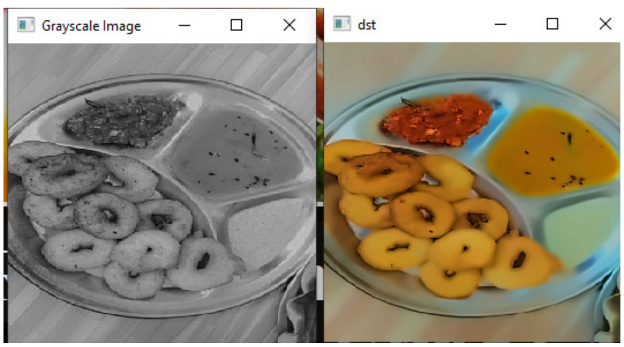


Figure 4: Preprocessing
It converts the image into gray scale

Food distribution: -> Pizza: 33.33%; Softdrinks: 33.33%; burgers: 33.33%; Vada:90%
Training: 4860 pcitures. Evaluation: 540 pictures.

Figure 5:Food Recognition
It recognise te food category and displays the food name with recognised accuracy %

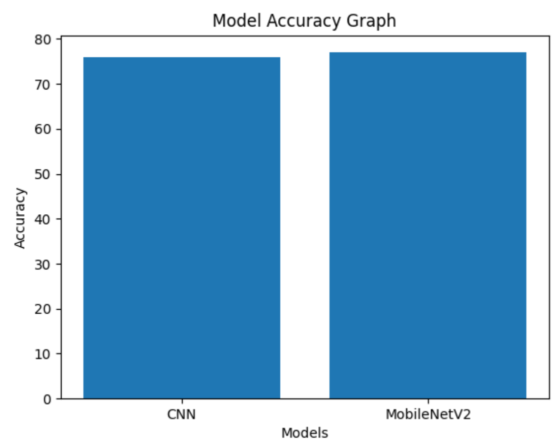
Classification Report:
----

	precision	recall	f1-score	support
burger	0.97	0.95	0.96	38
butter_naan	0.71	0.84	0.77	38
chai	0.92	0.89	0.91	38
chapati	0.56	0.61	0.58	38
chole_bhature	0.65	0.77	0.71	31
dal_makhani	0.71	0.88	0.79	40
dhokla	0.78	0.79	0.78	39
fried_rice	0.83	0.90	0.86	39
idli	0.69	0.72	0.71	40
jalebi	0.94	0.88	0.91	33
kaathi_rolls	0.73	0.77	0.75	39
kadai_paneer	0.76	0.83	0.79	35
kulfi	0.79	0.70	0.75	44
masala_dosa	0.86	0.68	0.76	37
momos	0.72	0.85	0.78	34
paani_puri	0.85	0.72	0.78	39
pakode	0.52	0.74	0.61	39
pav_bhaji	0.76	0.54	0.63	52
pizza	0.87	0.77	0.82	35
samosa	0.94	0.64	0.76	53
accuracy			0.77	781
macro avg	0.78	0.77	0.77	781
weighted avg	0.78	0.77	0.77	781

Figure 6: Classification Report
This gives the CNN classification Report with Accuracy 77%

Model Name	Accuracy
CNN	76%
MobileNetV2	77%

Table 1: Model Accuracy Table



Graph 1: Model Accuracy Graph
Model Vs Accuracy

VI. CONCLUSION

This study compares the accuracy of the MobileNetV2 model versus the conventional CNN model for multi-class classification of an Indian cuisine image dataset. It is observed that in terms of accuracy, the MobileNetV2 model performs better than the CNN model. In conclusion, transfer learning is preferable than conventional CNN when a sizable dataset is unavailable.

In subsequent research, it will be possible to identify the ingredients in a given food class and enhance the prediction of calories by utilizing volume estimation and additional extracted attributes. Using more than 12 classifications, a more advanced image classification tool can be created.

.

REFERENCES

1. S. Rousseau, Food ‘Porn’ in Media. Dordrecht, The Netherlands: Springer, Nov. 2014, doi: 10.1007/978-94-007-0929-4_395.
2. C. Chen, W. Min, X. Li, and S. Jiang, ‘‘Hybrid incremental learning of new data and new classes for hand-held object recognition,’’ J. Vis. Commun. Image Represent., vol. 58, pp. 138–148, Jan. 2019, doi: 10. 1016/j.jvcir.2018.11.009.
3. T. Mensink, J. Verbeek, F. Perronnin, and G. Csurka, ‘‘Distance-based image classification: Generalizing to new classes at near-zero cost,’’ IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 11, pp. 2624–2637, Nov. 2013, doi: 10.1109/TPAMI.2013.83.
4. D. Maltoni and V. Lomonaco, ‘‘Continuous learning in single-incrementaltask scenarios,’’ CoRR, vol. abs/1806.08568, pp. 56–73, Apr. 2018.
5. G. Ditzler, M. Roveri, C. Alippi, and R. Polikar, ‘‘Learning in nonstationary environments: A survey,’’ IEEE Comput. Intell. Mag., vol. 10, no. 4, pp. 12–25, Nov. 2015, doi: 10.1109/MCI.2015.2471196.
6. J. D. Power and B. L. Schlaggar, ‘‘Neural plasticity across the lifespan,’’ Wiley Interdiscipl. Rev., Develop. Biol., vol. 6, no. 1, p. e216, Jan. 2017, doi: 10.1002/wdev.216.
7. G.-B. Huang, L. Chen, and C.-K. Siew, ‘‘Universal approximation using incremental constructive feedforward networks with random hidden nodes,’’ IEEE Trans. Neural Netw., vol. 17, no. 4, pp. 879–892, Jul. 2006, doi: 10.1109/TNN.2006.875977.
8. X. Liu, C. Gao, and P. Li, ‘‘A comparative analysis of support vector machines and extreme learning machines,’’ Neural Netw., vol. 33, no. 9, pp. 58–66, Sep. 2012, doi: 10.1016/j.neunet.2012.04.002
9. P. Jackman, D.-W. Sun, P. Allen, N. A. Valous, F. Mendoza, and P. Ward, ‘‘Identification of important image features for pork and turkey ham classification using colour and wavelet texture features and genetic selection,’’ Meat Sci., vol. 84, no. 4, pp. 711–717, Apr. 2010, doi: 10. 1016/j.meatsci.2009.10.030.
10. G. Ciocca, P. Napoletano, and R. Schettini, ‘‘Food recognition and leftover estimation for daily diet monitoring,’’ in New Trends in Image Analysis and Processing—ICIAP, V. Murino, E. Puppo, D. Sona, M. Cristani, and C. Sansone, Eds. Cham, Switzerland: Springer, 2015, pp. 334–341.
11. P. Duan, W. Wang, W. Zhang, F. Gong, P. Zhang, and Y. Rao, ‘‘Food image recognition using pervasive cloud computing,’’ in Proc. IEEE Int. Conf. Green Comput. Commun. IEEE Internet Things IEEE Cyber, Phys. Social Comput., Aug. 2013, pp. 1631–1637, doi: 10.1109/GreenCom-iThingsCPSCom.2013.296.
12. W. Phetphoung, N. Kittimeteeworakul, and R. Waranusast, ‘‘Automatic sushi classification from images using color histograms and shape properties,’’ in Proc. 3rd ICT Int. Student Project Conf. (ICT-ISPC), Mar. 2014, pp. 83–86, doi: 10.1109/ICTISPC.2014.692322.
13. F. Kong and J. Tan, ‘‘DietCam: Automatic dietary assessment with mobile camera phones,’’ Pervas. Mob. Comput., vol. 8, no. 1, pp. 147–163, Feb. 2012, doi: 10.1109/ICT-ISPC.2014.6923223.
14. L. Bossard, M. Guillaumin, and L. Van Gool, ‘‘Food-101—Mining discriminative components with random forests,’’ in Proc. Eur. Conf. Comput. Vis., 2014, pp. 446–461.
15. H. Wang, W. Min, X. Li, and S. Jiang, ‘‘Where and what to eat: Simultaneous restaurant and dish recognition from food image,’’ in Advances in Multimedia Information Processing—PCM, E. Chen, Y. Gong, and Y. Tie, Eds. Cham, Switzerland: Springer, 2016, pp. 520–528.