Indian Journal of Food Engineering (IJFE)
ISSN: 2583-1275 (Online), Volume-1 Issue-1, December, 2021

Indian Food Image Recognition using a Deep Learning Approach

E. Emerson Nithiyaraj, S. Rajaseela

Abstract: ***In today's scenario, food image recognition is an interesting and useful application of visual object recognition. Convolutional Neural Network (CNN) is the best deep learning architecture for image classification tasks since it automatically learns the representations from the input images. Due to unlikeness and varieties of food available across the country, food recognition becomes very challenging. In this paper, a deep learning algorithm is proposed to recognize and classify 21 Indian food image categories. A transfer learning approach using Alex Net is developed for this task. For the experimentation, the dataset India-Food-21-Categories-Small is used from Kaggle and the Alex Net architecture is fine tuned for this application. Since the dataset has only limited amount of images, the available dataset is augmented to enhance the system's performance. The proposed CNN architecture results in an accuracy of 96.6% while trained for just 5 epochs.***

Keywords: Deep Learning, Convolutional Neural Network (CNN), Indian Food Image Recognition, Data Augmentation

I. INTRODUCTION

Computer Vision is a field of science that visualizes and understands the visual world using the concepts of image processing and machine learning. A computer vision technique tries to match the level of human perception in visual object recognition. However the system should be robust against noise, unwanted backgrounds, occlusion, lighting conditions, etc. Numerous machine learning techniques and algorithms are proposed for the task of image recognition. Since low-cost imaging devices are wide popular in today's camera market and due to the availability of free datasets in the internet, computer vision applications are being developed in great numbers among which food image recognition has recently gained wide attention. Food image recognition is an interesting and promising application of object recognition, since it will help the healthcare sector by monitoring the eating habits of people and to estimate the food calories consumed by a person every day. Food image recognition is the first step in a food consumption monitoring system. The complex and tedious part is that different food images have a much higher inter class similarity and intra class variation than usual ImageNet[1] objects like cars, animals etc.

Manuscript received on 12 November 2021 | Revised Manuscript received on 25 November 2021 | Manuscript Accepted on 15 December 2021 | Manuscript published on 30 December 2021.
***** Correspondence Author
E. Emerson Nithiyaraj*, Research Scholar, Department of ECE, Mepco Schlenk Engineering College, Sivakasi (Tamil Nadu), India.
S. Rajaseela, PG Student, Department of ECE, Pandian **Saraswathi** Yadav Engineering College, Sivaganga, (Tamil Nadu), India.

Different cars have the same outer shape but a single food dish doesn't look alike when it is prepared by different persons, which has high intra class variation. Deep Learning is a subdivision of Machine Learning which is rising quickly over the recent times. Deep learning techniques facilitates automatic feature extraction (which is a statistical representation of images) from the images where the user is unworried about manual feature extraction and the deep learning architectures like CNN extracts the best representative features from the images on its own as shown in Fig 1. Convolutional Neural Network (CNN) is best known for their ability to recognize patterns present in images. Machine learning approaches needs to be fed with hand crafted features where different applications works well on different features and finding the appropriate feature for a particular application is a time consuming process. In this work, classification between 21 Indian food images is done by the transfer learning approach using AlexNet and data augmentation techniques. The paper is divided into the following sections: Section II explains the existing works related to this application. Section III briefs the working of CNN architecture for the image classification task. In section IV, the experimentation is explained. Section V gives the results and discussion and in Section VI, conclusion and the direction for the future work are given.

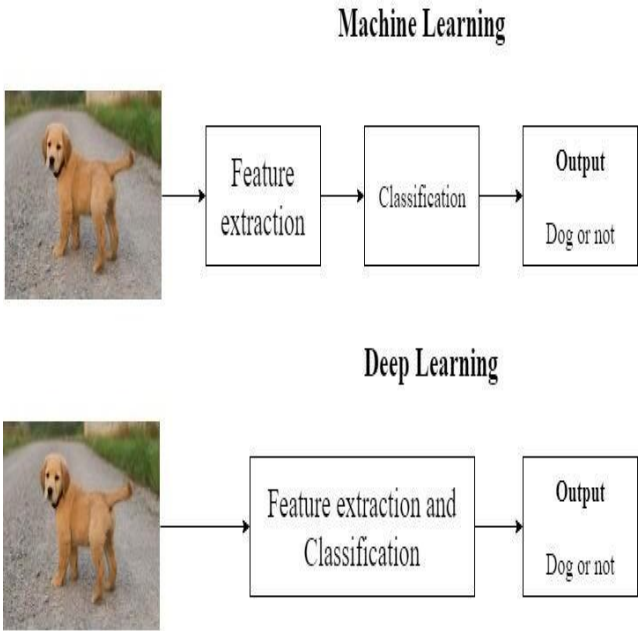

Fig 1: Machine Learning vs. Deep Learning

II. RELATED WORKS

Food image recognition applications are recently more popular using deep learning techniques [2] [3]. In [4] the authors have proposed Deep Convolutional Neural Network (DCNN) for Indian food image recognition and classification. The dataset consists of 60,000 grayscale images of size 280*280 pixels belonging to ten classes of Indian food. The algorithm has achieved a classification accuracy of 96.95%. In [5] a small dataset of ten food categories with consisting of 5822 images was used and a CNN with five layers was constructed to recognize the images where data augmentation techniques based on geometric transformation has resulted an accuracy of 94%. In [6] the authors have developed a food/non-food classification of images where they have achieved an accuracy of 99% using a CNN named 'Network in Network model' [7]. [8] has developed a multilayered deep CNN for automatic food classification on real-world food recognition database ETH Food-101which has provided a result of 95% using fine tuning approach. Since food image recognition is a new and trending area of research, very few existing works are present in the literature as discussed.

III. CONVOLUTIONAL NEURAL NETWORKS

CNN is a multilayer neural network in which input to each layer is fed with results of the previous layer. CNN architecture consists of convolutional layer next to the input layer and followed by various other layers such as pooling layer, Relu layer, normalization layer, loss layer (drop out), fully connected layer etc. Convolutional layer is the first layer in CNN that performs automatic feature extraction using filters or kernels. Each filter performs convolution operation with the input image where each filter may learn some feature and the matrix obtained after convolution operation all through the image is called the feature map'. Each convolution layer is followed by a RELU (Rectified Linear Unit) layer which adds non-linearity to the model and this function will output the same input if it is positive, otherwise, it will output zero. It has become the default activation function for many types of neural networks because a model that uses it is easier to train and often achieves better performance. To increase the performance and stability of a neural network further, normalization techniques are used that normalizes the output values of a previous feature map. Cross channel normalization, batch normalization, layer normalization, group normalization etc. are the different normalization techniques that has its own advantages and drawbacks. Pooling layer generally reduces the spatial size of the feature map which reduces the number of parameters and the complexity of computation in the network. Dropout layer prevents the drawback called overfitting while training deep neural networks. The fully connected layer is supplied with values from the final

pooling or final convolutional layer, where the values are vectorized and then fed as a vector into the fully connected layer. The final layer uses the softmax classifier which is used to get probabilities of each input belonging to a particular class. Implementing a CNN model requires significantly huge data to train the parameters of the network. CNN architecture may consist of various numbers of layers that depends upon the application and there are three ways to build a CNN: 1) Using a pre-trained model, 2) Transfer Learning, 3) Building a CNN from the scratch. Building a network from the scratch is a tedious task where it requires huge background knowledge about the CNN. Transfer learning [9] is a simple approach in building deep learning applications where a pre- trained network including its architecture and its parameters is used to learn a new task. Transfer learning is usually much faster and easier than building the network from scratch and training the network with randomly initialized weights. We can easily transfer the learned features from a network to a new task using a smaller number of training images. AlexNet, GoogleNet, VGG-16 are some the available pre-trained models and we have used AlexNet [10] for this work.

IV. EXPERIMENTATION

A. Dataset

The publically available "India-Food-21-Categories-Small" dataset from Kaggle is used for experimentation that consists of 21 food categories that Indian people loves [11]. The dataset consists of 460 RGB images of 21 different classes of Indian food images. From the available 460 images, 80% images are used to train the model, 10% for validation and 10% for testing. Some example images are shown in fig
3. All the images are resized to (256 x 256 x 3) pixels which is the input size of the AlexNet architecture.

B. Proposed Methodology

The proposed work uses the AlexNet CNN architecture and the transfer learning approach. AlexNet is a popular CNN architecture which participated in the ImageNet challenge in 2012 and achieved a top-5 error of only 15.3. AlexNet was trained with millions of images and can classify images into 1000 object categories (such as keyboard, coffee mug, pencil, and many animals). AlexNet contains five convolutional layers (CL) and three fully-connected layers (FCL), totally eight layers. Each convolution layer in AlexNet is followed by RELU, layers such as max pooling, cross channel normalization, dropout were also present in AlexNet which has been briefed in section III. Since our application has only 21 classes, the final fully connected layer of AlexNet is modified to 21 from 1000. The architecture of the proposed fine tuned AlexNet is presented below in Fig .2.

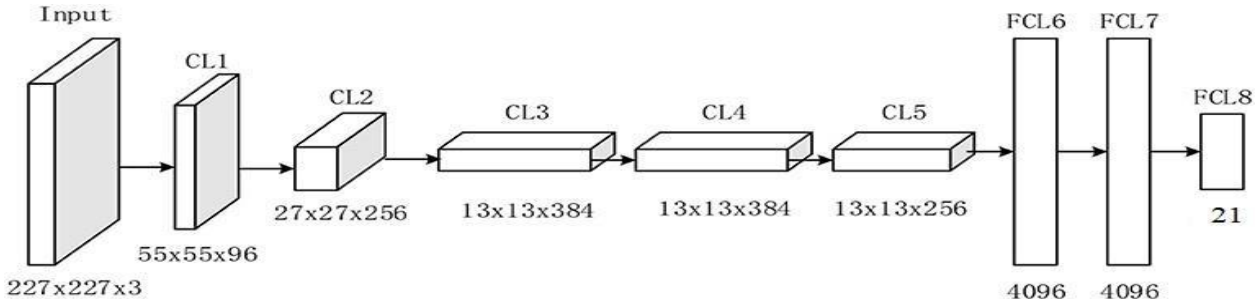

Fig 2: Proposed CNN Architecture using AlexNet

Fig 3: Sample Images from India-Food-21-Categories-Small dataset

V. RESULTS and DISCUSSION

The AlexNet is trained on CPU, Intel I5 processor and on 8 GB RAM and the working platform is Matlab. The optimal network parameters are: number of epochs – 5, learning rate – 0.0001, mini batch size – 10, optimizer – SGDM (Stochastic Gradient Descent). Using the above parameters and the food dataset, the network is re-trained using AlexNet's architecture and has achieved a validation accuracy of 75%. In-order to improve the performance of the system, data augmentation is proposed.

A. Data Augmentation

The performance of the CNN can be enhanced using data augmentation [12] technique. This is done because the deep neural network model has many parameters and the model should be trained with a huge amount of data so that the model is trained with the optimal parameters. For each original image, the images are horizontally flipped, vertically flipped and horizontally vertically flipped. Each image results in 4 images and totally 1840 images are generated using the data augmentation approach. The same 80%, 10%, 10% convention is used for training validation and testing the proposed model. Data augmentation improved the accuracy from 75% to 96.6% using the same AlexNet architecture.

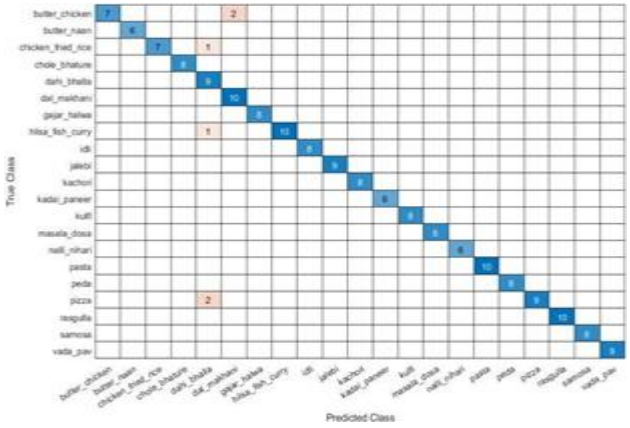

Fig 4: Confusion Matrix for the testing set

A **confusion matrix** is a matrix or a performance metric that is used to evaluate the performance of a classification model (or classifier) on the validation or test data for which the actual labels are known. The confusion chart for the proposed model is shown in fig 4.

Table 1: Comparison of our results with existing works

	No of imas	Epoch	Learning Rate	Accuracy
[4]	60,000	10	0.001	96.9%
[5]	5822	100	0.001	94%
Proposed work without Data Augmentation	460	5	0.0001	75%
Proposed work with Data Augmentation	**1840**	**5**	**0.0001**	**96.6%**

While comparing the proposed work to the other existing works in the literature related to food image recognition, our work has achieved a far good result. [2] has used around 60,000 images to train for 10 epochs and [3] has used 5822 images to train for 100 epochs which had been trained for a longtime using CPUs and the proposed model has achieved a best accuracy of 96.6% with just 1840 images trained for 5 epochs where the performance of our system has been improved using data augmentation. The training time is also very less even while using a CPU. The testing results from the dataset are shown below in fig 5 and testing results using Google images are shown in fig 6.

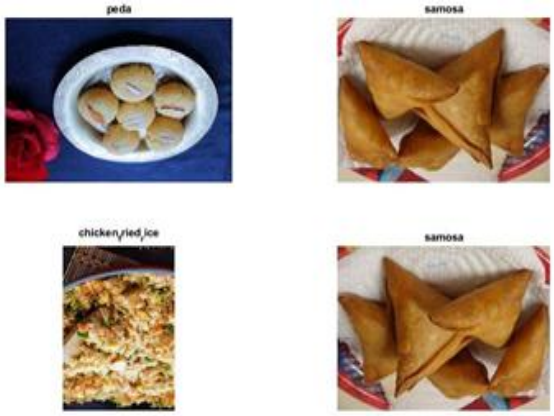

Fig. 5: Testing Results using Dataset Images

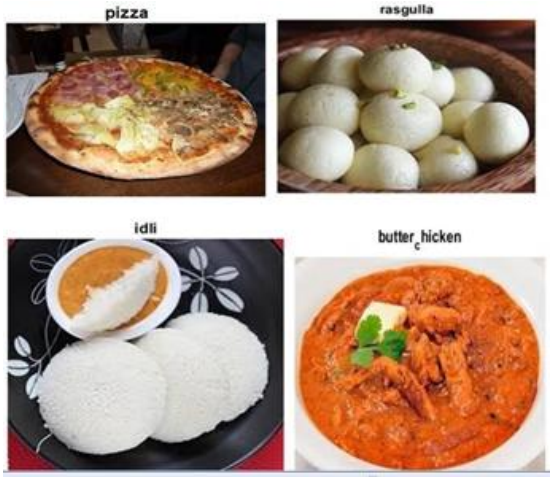

Fig 6: Testing Results using Google Images

VI. CONCLUSION

In this paper, the proposed Indian food image recognition using transfer learning approach and data augmentation technique is discussed. AlexNet architecture has been fine tuned to classify between 21 Indian food categories which has resulted in a remarkable accuracy of 96.6% trained for only 5 epochs. Once we feed the deep learning architectures with huge amount of data they perform better and so AlexNet has learnt the food image features very well and has provided a very good result. Hence, this work resulted in a good accuracy with limited images and in very less time as compared to other works in literature as shown in table 1. These kinds of works can be used in healthcare monitoring system for monitoring the patient's diet habits, apps can be developed for recognizing the unknown food items and many more. Food image recognition helps to monitor the amount of calories consumed by a person per day and to monitor the eating habits of people by tracking their food consumption using cameras. In future, more no of classes can be added and also CNN architecture can be built from scratch for this application.

REFERENCES

1. Deng J, Dong W, Socher R, Li LJ. ImageNet: a large-scale hierarchical image database. In: *2009 IEEE Conference on Computer Vision and Pattern Recognition.* Miami, FL (2009). p. 248–55.
2. Chen, X., Zhou, H., & Diao, L. (2017). ChineseFoodNet: A large-scale Image Dataset for Chinese Food Recognition. *ArXiv, abs/1705.02743.*
3. Aguilar E., Bolaños M., Radeva P. (2018) Exploring Food Detection Using CNNs. In: Moreno-Díaz R., Pichler F., Quesada-Arencibia A. (eds) Computer Aided Systems Theory – EUROCAST 2017.
4. Jahan, Shamay. "Deep Indian Delicacy: Classification of Indian Food Images using Convolutional Neural Networks.", IJRASET, Vol 6 Issue III, March 2018
5. Yuzhen Lu , "Food Image Recognition by Using Convolutional Neural Networks (CNNs) arXiv:1612.00983.
6. Kagaya H., Aizawa K. (2015) Highly Accurate Food/Non-Food Image Classification Based on a Deep Convolutional Neural Network. In: Murino V., Puppo E., Sona D., Cristani M., Sansone C. (eds) New Trends in Image Analysis and Processing -- ICIAP 2015 Workshops. ICIAP 2015. Lecture Notes in Computer Science, vol 9281. Springer, Cham
7. Lin, M., Chen, Q., Yan, S.: Network in network. In: Proceedings of International Conference on Learning Representations (2014)
8. Pandey, P., Deepthi, A., Mandal, B., & Puhan, N.B. (2017). FoodNet: Recognizing Foods Using Ensemble of Deep Networks. *IEEE Signal Processing Letters, 24*, 1758-1762.
9. Hussain M, Bird JJ, Faria DR. A study on CNN transfer learning for image classification. In: *Advances in Computational Intelligence Systems*. Lotfi A, Bouchachia H, Gegov A, Langensiepen C, McGinnity M, editors. Cham: Springer International Publishing Ag (2019) p. 191– 202.
10. Alex krizhevsky, "ImageNet classification with deep convolutional neural networks" *NIPS'12: Proceedings of the 25th International Conference on Neural Information Processing Systems* - Volume 1, 2012, pp. 1097–1105
11. https://www.kaggle.com/sandy1112/indiafood21categoriessmall
12. Ouchi T, Tabuse M. Effectiveness of data augmentation in automatic summarization system. In: Sugisaka M, Jia Y, Ito T, Lee JJ editors. *International Conference on Artificial Life and Robotics (ICAROB).* Oita: Alife Robotics Co., Ltd. (2019). p. 177–80.
13. EUROCAST 2017. Lecture Notes in Computer Science, vol 10672. Springer, Cham.