A Convolution Neural Network Based Classification Approach for Recognizing Traditional Foods of Bangladesh from Food Images

Chapter *in* Advances in Intelligent Systems and Computing · January 2020

DOI: 10.1007/978-3-030-16657-1_79

3 authors:

Nishat Tasnim
Daffodil International University
4 PUBLICATIONS **30** CITATIONS
SEE PROFILE

Md Romyull Islam
Kennesaw State University
12 PUBLICATIONS **53** CITATIONS
SEE PROFILE

Shaon Bhatta Shuvo
Daffodil International University
18 PUBLICATIONS **203** CITATIONS
SEE PROFILE

A Convolution Neural Network Based Classification Approach for Recognizing Traditional Foods of Bangladesh from Food Images

Nishat Tasnim(✉), Md. Romyull Islam, and Shaon Bhatta Shuvo

Daffodil International University, Dhanmondi-32, Dhaka, Bangladesh
{tasniml5-5709, islaml5-5833, shaon.cse}@diu.edu.bd

Abstract. The process of identifying food items from an image is one of the promising applications of visual object recognition in computer vision. However, analysis of food items is a particularly challenging task due to the nature of their has achieved by traditional approaches in the field. Deep neural networks have exceeded such solutions. With a goal to successfully applying computer images, which is why a low classification accuracy vision techniques to classify food images based on Inception-v3 model of TensorFlow platform, we use the transfer learning technology to retrain the food category datasets. Our approach shows auspicious results with an average accuracy of 95.2% approximately in correctly recognizing among 7 traditional Bangladeshi foods.

Keywords: Object recognition · Computer vision · Deep neural networks · Inception-v3

1 Introduction

Traditional foods play a significant role in a country's culture, which reflects its unique history, identity, heritage, lifestyle, values, and beliefs, as well as helps to easily understand people's humor. Some traditional foods have geographical indications. Most of our traditional foods have evolved over centuries, and it is a gift from a generation to the next generation. Knowledge about the processing and preservation techniques of traditional foods has been known for many generations in a country. Traditional foods have historical precedence in a countries national dish. Regional cuisine or local cuisine refers to specific ingredients, and the location of the production. These traditional foods vary significantly from each other and use locally available spices, herbs, fruits, and vegetables. Almost every country has unique recipes to represent their respective traditional foods. Like other countries, Bangladesh also has its own tradition of foods. People of Bangladesh tend to identify themselves with their food.

The process of identifying food items and extraction of information from food image is quite an interesting as well as a challenging issue. With the development of computer and digital image processing technology, people began to search the method of automatic food items recognition by computer. The ultimate goal of our research is

A. Abraham et al. (Eds.): ISDA 2018, AISC 940, pp. 844–852, 2020.
https://doi.org/10.1007/978-3-030-16657-1_79

to develop a new strategy to recognize food items automatically by utilizing new techniques of Computer Vision and Machine Learning and improving the accuracy. During this paper, the key technological innovation is the deep learning-based food image recognition algorithms.

In this paper, we use the transfer learning technique to retrain the Inception-v3 [1] model of Tensor Flow [2] on the dataset of 7 food items, traditionally known as Kalavuna, Khichuri, Murgir-Roast, Panta-Ilish, Sorse-Ilish, Gorur-Rezala, and Tehari. We implemented an effective food items recognition model using a short training time and achieved a higher accuracy.

The rest of the paper arrangement is as follows: in Sect. 2, details of TensorFlow, transfer learning [3], Convolutional Neural Network (CNN) [4] and Inception-v3 model are presented. Section 3 deals with the Literature Review. Data collection and training are explained in Sect. 4. Result analysis has done in Sect. 5; and finally, some future work scopes and conclusion are discussed in Sect. 6.

2 Background Study

We have deployed an Inception-v3 [1] model of TensorFlow [2] platform and used CNN [4]. TensorFlow is an open source library, developed by Google Brain Team within Google's Machine Learning Intelligence analysis association, for numerical computation that makes machine learning simpler and faster. TensorFlow joins the computational algebra of compilation optimization techniques, making simple the calculation of many mathematical expressions where the problem is the time needed to accomplish the computation.

Inception-v3 is the 2015 iteration of Google's Inception architecture and a widely used image recognition network model that has been presented to obtain greater than 78.1% accuracy on the ImageNet dataset. Inception is a remarkable architecture and it is the result of multiple cycles of trial and error. The model is the climax of many ideas developed by multiple researchers over the years. The original paper on which Inception-v3 is based on is: "Rethinking the Inception Architecture for Computer Vision" by Szegedy et al. [5] (Fig. 1).

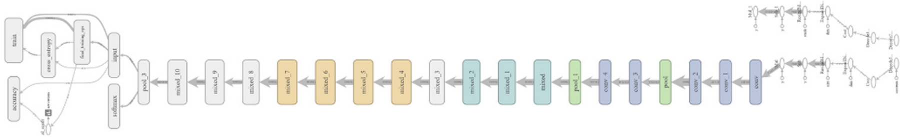

Fig. 1. Main graph of Inception-v3 model from tensor board.

Transfer learning [3] is the application of knowledge obtained from completing one task to help in solving a different but related problem. The development of algorithms that assist transfer learning processes has become a goal of machine learning technicians as they attempt to make machine learning as human-like as possible [3]. For example, the knowledge obtained by a machine learning algorithm to recognize trucks could later be transferred for use in other machine learning model being developed to recognize other types of vehicles, such as buses. Compared with the traditional neural network, it only needs to use a small amount of data to train the model, and attain high accuracy with a short training time.

Convolutional neural networks (CNN) [4] are deep artificial neural network architectures, which are utilized basically to classify images, cluster them by similarity, and to accomplish object recognition within scenes. A CNN consists of one or more convolutional layers and then followed by one or more completely associated layers as in a standard multilayer neural network [6]. It learns especially from images. A CNN can be prepared to do image analysis tasks including classification, object detection, segmentation, and image processing.

Many types of layers are used to build ConvNet architectures like Convolutional Layer, Non-Linearity Layer, Rectification Layer, Rectified Linear Units (ReLU), Pooling Layer, Fully Connected Layer, Dropout Layer (Fig. 2).

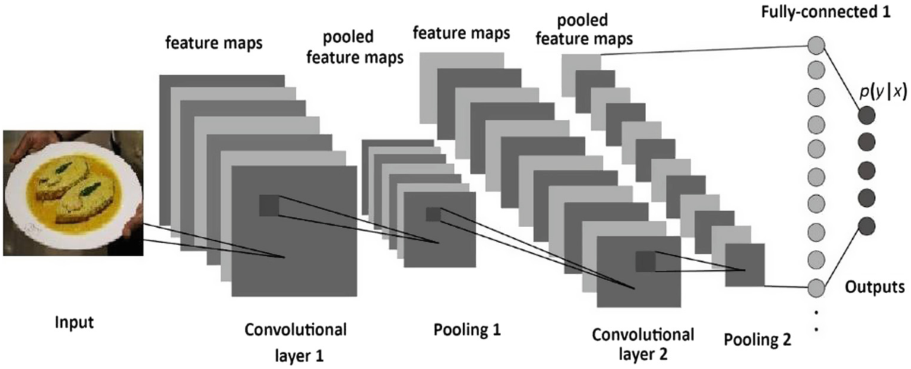

Fig. 2. The Architecture of Convolutional Neural Network (CNN) [7].

The main task of the convolutional layer is to detect local conjunctions of features from the previous layer and mapping their appearance to a feature map. As a result of convolution in neuronal networks, the image is split into perceptrons, creating local receptive fields and finally compressing the perceptrons in feature maps of size $m_2 \times m_3$. Thus, this map stores the information where the feature occurs in the image and how well it corresponds to the filter. Hence, each filter is trained spatial in regard to the position in the volume it is applied to [8].

In each layer, there is a bank of m_1 filters. The number of how many filters are applied in one stage is equivalent to the depth of the volume of output feature maps. Each filter detects a particular feature at every location on the input [6]. The output $Y_i^{(l)}$ of layer l consists of $m_1^{(l)}$ feature maps of size $m_2^{(l)} \times m_3^{(l)}$. The i^{th} feature map, denoted $Y_i^{(l)}$, is computed as

$$Y_i^{(l)} = B_i^{(l)} + \sum_{j=1}^{m_1^{(l-1)}} K * Y_j^{(l-1)} \tag{1}$$

where $B_i^{(l)}$ is a bias matrix and $K_{i,j}^{(l)}$ is the filter of size $2h_1^{(l)} + 1 \times 2h_2^{(l)} + 1$ connecting the j^{th} feature map in layer $(l - 1)$ with i^{th} feature map in layer.

The pooling layer is responsible for reducing the spatial size of the activation maps [8].

The pooling layer l has two hyperparameters, the spatial extent of the filter $F^{(l)}$ and the stride $S^{(l)}$. It takes an input volume of size $m_1^{(l-1)} \times m_2^{(l-1)} \times m_3^{(l-1)}$ and provides an output volume of size $m_1^{(l)} \times m_2^{(l)} \times m_3^{(l)}$ where;

$$m_1^{(l)} = m_1^{(l-1)} \tag{2}$$

$$m_2^{(l)} = m_2^{(l-1)} - F^{(l)})/S^{(l)} + 1 \tag{3}$$

$$m_3^{(l)} = (m_3^{(l-1)} - F^{(l)})/S^{(l)} + 1 \tag{4}$$

The goal of the complete fully connected structure is to tune the weight parameters to create a stochastic likelihood representation of each class based on the activation maps generated by the concatenation of convolutional, non-linearity, rectification and pooling layers [8].

If $l - 1$ is a fully connected layer;

$$y_i^{(l)} = f\left(z_i^{(l)}\right) \text{ with } z_i^{(l)} = \sum_{j=1}^{\mathrm{m}_1^{(l-1)}} w_{i,j}^{(l)} y_i^{(l-1)} \tag{5}$$

Otherwise;

$$y_i^{(l)} = f(z_i^{(l)} \text{ with } z_i^{(l)} = \sum_{j=1}^{\mathrm{m}_1^{(l-1)}} \sum_{r=1}^{\mathrm{m}_2^{(l-1)}} \sum_{s=1}^{m_3^{(l-1)}} w_{i,j,r,s}^{(l)} \left(Y_i^{(l-1)}\right) r, s \tag{6}$$

3 Literature Review

A plethora of research and development efforts have been made in the field of computer vision over the last few years to ease the effort of automatic object recognition, among which food image recognition has recently gained much importance. However, none of the work has found recognizing traditional foods.

Inception-v3 [1] model is used in many researches of different categories. One of the work by Xiaoling Xia and Cui Xu from College of Computer Science, Donghua

University used the transfer learning technique to retrain the Inception-v3 model of TensorFlow on the flower category datasets [9] of Oxford-I7 and Oxford-102 for Flower Classification in 2017. The classification precision of the model was 95% on Oxford-I7 flower dataset and 94% on Oxford-102 flower dataset [10].

Mathew et al. from bVuelogix Technologies Pvt Ltd utilized Google's TensorFlow deep learning a framework to train, validate and test the network for Intrusion Detection in 2017 [11]. The precision was 95.3%. However, the proposed network is found to be harder to train due to vanishing gradient and degradation issues. Kieffer et al. used CNN and Inception-v3 model for Histopathology Image Classification in 2017 [12]. All experiments were done on Kimia Path24 dataset. The precision was 56.98%. Xia et al. worked for Facial Expression Recognition based on the Inception-v3 model of TensorFlow platform in 2017. They used CK+ dataset [13] and selected 1004 images of facial expression. The precision was 97% but it was not based on dynamic sequences. Batsukh and Tsend worked on Effective Computer Model for Recognizing Nationality from the Frontal Image in 2016 [14]. They used SVM [15], AAM [16], ASM [17]. The precision was 86.4%. The analysis was worked manually and images must be the frontal face image that has smooth lighting and does not have any rotation angle.

The aim of our research is to provide a suitable methodology for accurate automation of traditional food images classification as the first work of its kind.

4 Methodology

The following diagrammatic representation explains the methodology of our proposed model (Fig. 3):

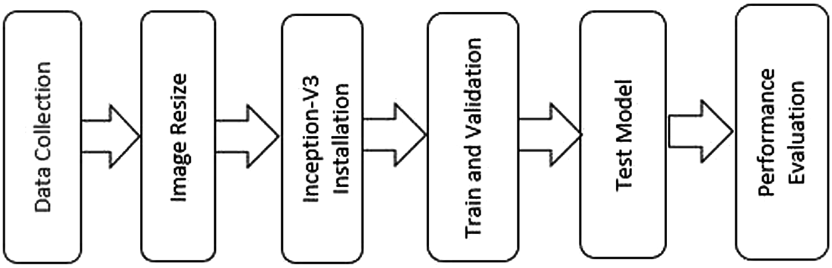

Fig. 3. Flowchart of the proposed model.

4.1 Dataset

Bangladesh has plenty of traditional foods. For recognizing different traditional food, we collected 700 images of seven traditional foods. They are locally known as, Kalavuna, Khichuri, Murgir-Roast, Panta-Ilish, Sorse-Ilish, Gorur-Rezala, and Tehari. We collected 100 images of each of the item. The following images represent a portion of our dataset (Fig. 4).

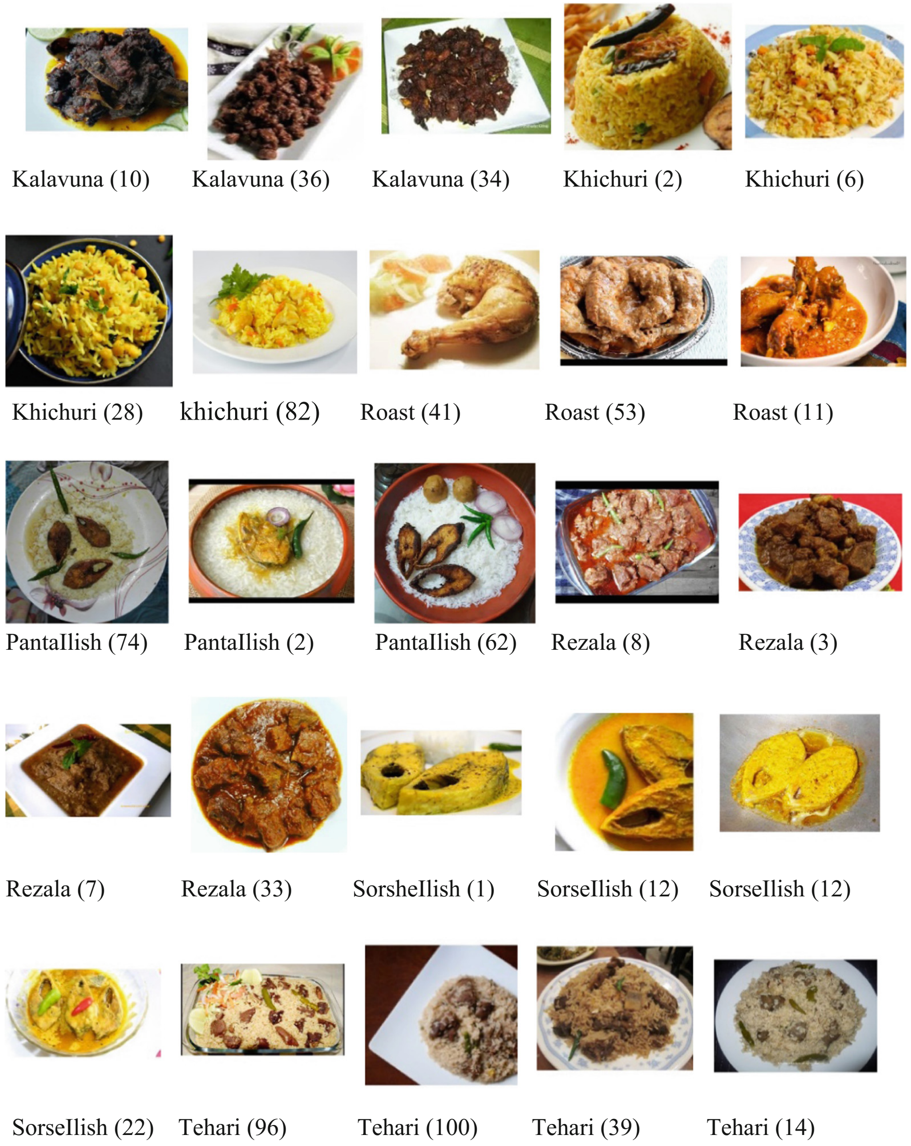

Fig. 4. A portion of our dataset of traditional foods of Bangladesh.

4.2 Data Pre-processing

Image pre-processing is an important stage to promote the effect of image classification. The learning method of CNN directs the execution of our activity in machine learning, so in the image pre-processing step we have labeled and resized the images for training and testing.

4.3 Model Installation

First, we have downloaded TensorFlow. Then, we have downloaded Inception-v3 model. We have also used the transfer learning method that keeps the parameters of the earlier layer and have expelled the final layer of the Inception-v3 model, then retrain a final layer.

4.4 Train Model

In this progression, we should keep the parameters of the past layer, then expel the final layer and input our dataset to retrain the new last layer. The last layer of the model is trained by backpropagation algorithm [18], and the cross-entropy cost function [19] is utilized to integrate the weight parameter by calculating the error between the output of the softmax layer, and the label vector of the given test category [10, 13].

5 Result Analysis

The variations in accuracy based on cross-entropy in our training dataset are shown in Figs. 5 and 6 respectively. The training set is represented by the orange line, and the validation set is represented by the blue line.

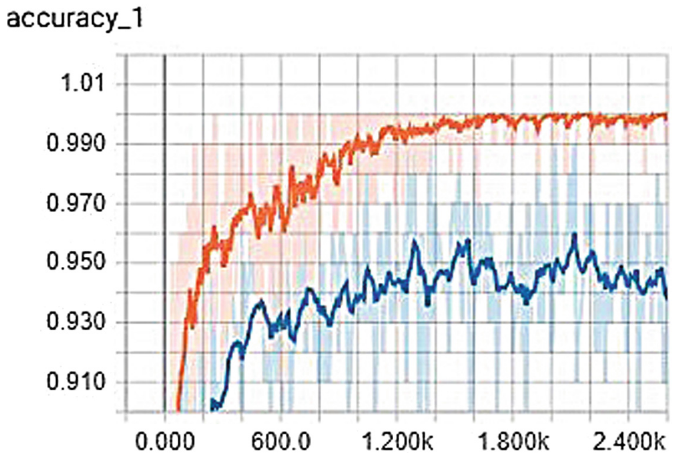

Fig. 5. The variation of accuracy on training and validation set.

The description of the two figures is shown in the following table (Table 1).

The training accuracy reached 99%, and the validation accuracy maintained between 91% to 95% for our dataset.

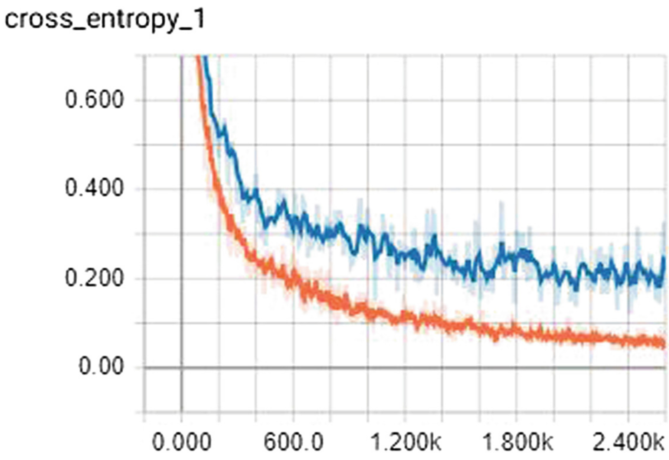

Fig. 6. The variation of cross-entropy

Table 1. Description of the two figures.

Dataset	Index	Performance
Dataset	The accuracy of the training set	99%
	The accuracy of the validation set	91% to 95%
	The cross entropy of the training set	0.07
	The cross-entropy of the validation set	22

6 Conclusion and Future Work

We have demonstrated a comprehensive pathway to classify traditional foods of Bangladesh from food images, which is so far the first work of its kind. As a first research work on this domain, the result is quite satisfactory as well as encouraging. We also believe this work will inspire the researchers from various countries to work on their traditional items.

We proposed the classification model based on the Inception-v3 model for seven different food items. Hopefully, in future, we could extend the work with a larger dataset having more varieties of items. We also have the plan to implement some other CNN based models to compare the accuracy on the same dataset.

References

1. Inception-v3. https://cloud.google.com/tpu/docs/inception-v3-advanced. Accessed 1 Sept 2018
2. TensorFlow. https://www.packtpub.com/mapt/book/big_data_and_business_intelligence/9781786468574/1/ch01lvl1sec9/tensorflow–a-general-overview. Accessed 2 Sept 2018

3. Transfer learning. https://searchcio.techtarget.com/definition/transfer-learning. Accessed 3 Sept 2018
4. Islam, M.S., Foysal, F.A., Neehal, N.: InceptB: a CNN Based classification approach for recognizing traditional Bengali games. Procedia Comput. Sci. **143**, 592–602 (2018)
5. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2818–2826 (2016)
6. Convolutional neural network architecture. http://ufldl.stanford.edu/tutorial/supervised/ConvolutionalNeuralNetwork/. Accessed 5 Sept 2018
7. The Architecture of Convolutional Neural Network. https://www.mdpi.com/10994300/19/6/242/htm. Accessed 12 Sept 2018
8. Convolutional Neural Network. https://wiki.tum.de/display/lfdv/Layers+of+a+Convolutional+Neural+Network. Accessed 5 Sept 2018
9. Flower category datasets. https://datascience.stackexchange.com/questions/15989/microaverage-vs-macro-average-performance-in-a-multiclass-classification. Accessed 6 Sept 2018
10. Xia, X., Xu, C., Nan, B.: Inception-v3 for flower classification. In: 2017 2nd International Conference on Image, Vision and Computing (ICIVC), pp. 783–787. IEEE (2017)
11. Mathew, A., Mathew, J., Govind, M., Mooppan, A.: An improved transfer learning approach for intrusion detection. In: 7th International Conference on Advances in Computing and Communications, Cochin, India (2017)
12. Kieffer, B., Babaie, M., Kalra, S., Tizhoosh, H.R.: Convolutional neural networks for histopathology image classification: training vs. using pre-trained networks. In: 2017 Seventh International Conference on Image Processing Theory, Tools and Applications (IPTA), pp. 1–6. IEEE (2017)
13. Xia, X.L., Xu, C., Nan, B.: Facial expression recognition based on tensorflow platform. In: ITM Web of Conferences, vol. 12, p. 01005. EDP Sciences (2017)
14. Batsukh, B.E., Tsend, G.: Effective computer model for recognizing nationality from frontal image (2016)
15. Support vector machine. https://en.wikipedia.org/wiki/Support_vector_machine. Accessed 13 Sept 2018
16. Active appearance model. https://en.wikipedia.org/wiki/Active_appearance_model. Accessed 15 Sept 2018
17. Cootes, T., Baldock, E.R., Graham, J.: An introduction to active shape models, pp. 223–248 (2000)
18. Backpropagation algorithm. https://en.wikipedia.org/wiki/Backpropagation. Accessed 11 Sept 2018
19. Cross-entropy. https://en.wikipedia.org/wiki/Cross_entropy. Accessed 8 Sept 2018