Food Image Classification Using Convolutional Neural Network

Sagidur Rahman
ID: 2014–2–60–067

B.M. Nafiz Karim Siddique
ID: 2014–3–60–078

Md Tohidul Islam
ID: 2015–1–60–32

A thesis submitted in partial fulfillment of the requirements for the degree of Bachelor of Science in Computer Science and Engineering

Department of Computer Science and Engineering
East West University
Dhaka-1212, Bangladesh

September, 2018

Declaration

We, Sagidur Rahman, B.M. Nafiz Karim Siddique and Md Tohidul Islam hereby, declare that the work presented in this thesis titled 'Food Image Classification Using Convolutional Neural Network' is outcome of the investigation performed by me under the supervision of Dr. Taskeed Jabid , Assistant Professor, Department of Computer Science and engineering, East West University. I also declare that no part of this thesis has been or is being submitted elsewhere for the award of any degree or diploma.

....
Dr Taskeed Jabid
Supervisor

....
Sagidur Rahman
ID: 2014-2-60-067

....
B.M. Nafiz Karim Siddique
ID: 2014-3-60-078

....
Md Tohidul Islam
ID: 2015-1-60-032

Letter of Acceptance

I hereby declare that, this thesis is the students own work and best effort of mine. All other sources of information used have been acknowledged. This thesis has been submitted with my approval.

....

Dr. Taskeed Jabid
Assistant Professor
Department of Computer Science and Engineering
East West University
Aftabnagar, Dhaka-1212, Bangladesh

Supervisor

....

Dr. Ahmed Wasif Reza
Associate Professor
Department of Computer Science and Engineering
East West University
Aftabnagar, Dhaka-1212, Bangladesh

Chairperson

Abstract

In our thesis we tried to classify food images using convolutional neural network. Convolutional neural network extracts spatial features from images so it is very efficient to use convolutional neural network for image clasification problem. Recently people are sharing food images in social media and writing review on food. So there is a lot of food image but some image may not be labeled. It will be very helpful for restaurants if they can advertise their food to those people who is looking similar kind of foods they offer. Food classification system can help social media platform to identify food. Food classification system can enable an opportunity for social media platform to offer advertisement service for restaurants and beverage companies to their targeted users. It will be financially beneficial for both social media platform and beverage companies. Food classification is very difficult task because there is high variance in same category of food images. We developed a convolutional neural network model to classify food images in food-11 dataset. We also used transfer learning technique using Inception V3.

Acknowledgments

First of all, We would like to express my deepest gratitude to the almighty Allah for His blessings on us. Next, our special thanks go to our supervisor, “Dr. Taskeed Jabid”, who gave us this opportunity, initiated us into the field of “Computer Vision and Deep Learning”, and without whom this work would not have been possible. His encouragements, visionaries and thoughtful comments and suggestions, unforgettable support at every stage of our BS.c study were simply appreciating and essential. His ability to muddle us enough to finally answer our own question correctly is something valuable what We have learned and We would try to emulate, if ever We get the opportunity.

There are numerous other people too who have shown me their constant support and friendship in various ways, directly or indirectly related to our academic life. We will remember them in our heart and hope to find a more appropriate place to acknowledge them in the future.

Sagidur Rahman
August, 2018

B.M. Nafiz Karim Siddique
August, 2018

Md Tohidul Islam
August, 2018

Table of Contents

Declaration of Authorship **i**

Letter of Acceptance **i**

Abstract **i**

Acknowledgment **ii**

Table of Contents **iii**

List of Figures **iv**

List of Tables **vi**

1 Introduction **1**

2 Background Study **4**

2.1 Convolutional Neural Network 4

2.2 Transfer Learning 7

3 Related Work **10**

3.1 Featuee Based Model 10

3.2 Deep Learning Based Model 12

4 Methodology **14**

4.1 Dataset 15

4.2 Pre-Processing 15

4.3 Training Convolution Neural Network Model from Scratch 17

4.4 Transfer Learning 17

5 Experimental Results **19**

5.1 Evaluation of Model 19

5.2 Obtained Results 20

6 Concluson and Future Work **23**

Bibliography **23**

List of Figures

2.1 Typical CNN Architecture 5
2.2 Max Pooling 6
2.3 Depiction of Inductive Transfer Taken from Transfer Learning. 8

4.1 Methodology 14
4.2 Sample Dataset 16
4.3 Architecture of our Convoluional Neural Network. 17
4.4 Architecture of Inceptin V3. 18

5.1 Accuracy curve of proposed model. 21
5.2 Loss curve of proposed model. 21
5.3 Accuracy curve of Inception V3 (Transfer Learning). 22
5.4 Loss curve of Inception V3 (Transfer Learning). 22

List of Tables

5.1 Test results of models 20

Chapter 1

Introduction

Due to emergence of social media thousands of pictures are being uploaded in social media. The rise of mobile devices put a camera in everyones pocket. Suddenly, consumers no longer needed a desktop device to log into their social media accounts, and very quickly they began to engage on social media platforms on the go and on the same phones or tablets they used to snap photos. Image-based social networks such as Instagram and Snapchat emerged to cater to this new reality. We have also seen text-based social platforms, such as Twitter, respond to user behavior by offering more opportunities to put images front and center. No doubt, marketers know the importance of social media marketing.

The amount of information flooding the Internet, namely social media platforms, is huge. For brands, this data represents both a challenge and an opportunity as they look to effectively market themselves, protect their image, and excel in the era of information overload. Image recognition tools are the key to unlocking the potential hidden within this every-growing pool of images online.

For brands, this means exposure to more data than ever before, especially image-based data. Social media users have fully-embraced the concept of sharing photos in place of or accompanying text.

Evidence of this real-life explosion of, a picture is worth a thousand words can be seen in the growing popularity photo-based social media platforms such as Snapchat, with 187M daily users at the end of 2017 - up from 46 million at the start of 2014. And, Instagram, with 1B monthly users, up from 800M in September 2017.

The ability to identify, analyze, and exploit this growing trend became essential. With the future of digital marketing being dominated by visual data - image recognition technology had to exist. Without it, brands were missing out on a whole heap of valuable data.

The photos people share on social media represent the consumer behaviors, wants and needs that often are undetected by marketers. If a person posts a photo of a new product, but doesn't include any text saying the products name, social media monitoring probably won't capture it. As a result, companies miss out on big opportunities to learn about and communicate with their customers. Now, artificial intelligence and image recognition are making it easier for marketers to find visuals within social media, even when they're not accompanied by an explicit text mention.

In our thesis we developed a CNN model to classify food from images. Millions of food images are being uploaded in social media. With food recognition system social media can cluster people based on their food choices. It would be helpful for restaurants, beverage manufacturer and social media platform to advertise targeted audience. Food recognition model will be helpful to recognize food even if the food is not labeled in actual post.

Image processing and computer vision techniques are now being applied in many domains. One of the domains is food recognition from image. Food recognition is a challenging task since there is less intra-class difference in food images. Besides any image recognition also needs more computation power than most of the text based data classification. But to be benefited from food recognition model people should be able use it in a less expensive device. In present day there are affordable smartphones with high computation power that can process high quality image data. So the model described in this thesis can be implemented in smartphones.

The rest of the chapters of this thesis are arranged in following ways:

- The chapter 2 reviews the background and describe the CNN and transfer learning.
- The chapter 3 summarizes the related works involving food image classification.
- The chapter 4 describe our approaches to classify food images using CNN in details.
- The chapter 5 analyze the experimental results found in our thesis.
- The chapter 6 discuss about some future works about food image classification that can be implemented in future.

Chapter 2

Background Study

In our research we used convolutional neural network to classify food images. We also used transfer learning method in our work. This chapter describes the convolutional neural networks and transfer learning method in details.

2.1 Convolutional Neural Network

In machine learning, a convolutional neural network (CNN, or ConvNet) is a class of deep, feed-forward artificial neural networks, most commonly applied to analyzing visual imagery. CNNs use a variation of multilayer perceptrons designed to require minimal preprocessing [1]. They are also known as shift invariant or space invariant artificial neural networks (SIANN), based on their shared-weights architecture and translation invariance characteristics [2, 3].

Convolutional networks were inspired by biological processes[4] in that the connectivity pattern between neurons resembles the organization of the animal visual cortex. Individual cortical neurons respond to stimuli only in a restricted region of the visual field known as the receptive field. The receptive fields of different neurons partially overlap such that they cover the entire visual field.

CNNs need relatively little pre-processing of datasets compared to other image classification algorithms. This means that the network learns the filters that in traditional algorithms were hand-engineered. This independence from prior knowledge and human effort in feature design is a major advantage. They have applications in image and video

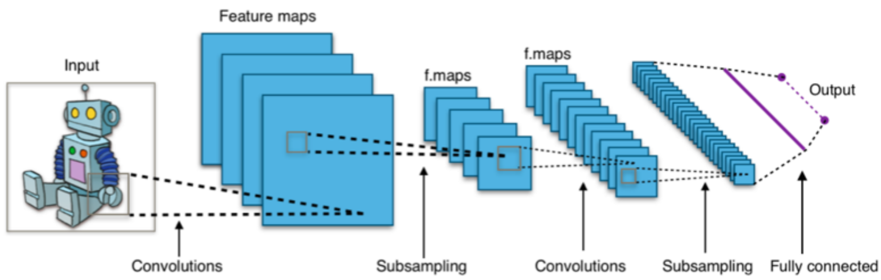

Figure 2.1: Typical CNN Architecture

recognition, recommender systems and natural language processing.

Convolutional Neural Network architectures make the explicit assumption that the inputs are images, which allows us to encode certain properties into the architecture. These then make the forward function more efficient to implement and vastly reduce the amount of parameters in the network. The description of layers in convolutional neural networks described below.

- Convolution Layer: The Convolution layer is the core building block of a Convolutional Network that does most of the computational heavy lifting. Three hyper-parameters control the size of the output volume: the depth, stride and padding. The Output Size of a convolutional layer is:

$$\left[\frac{n+2p-f}{s}+1*\frac{n+2p-f}{s}+1\right]$$

  where p = padding, s = stride, f = number of filters, n = image width = image height [1].

- Max Pooling Layer: Max pooling is a sample-based discretization process. Max pooling is done by applying a max filter to (usually) non-overlapping sub regions

[1] http://cs231n.github.io/convolutional-networks/

of the initial representation. The Output Size of a Max-Pooling layer is:

$$\left[\frac{n_h - f}{s} + 1 * \frac{n_w - f}{s} + 1\right]$$

where, s = stride, f = number of filters, n_h = input height n_w= image width [4].

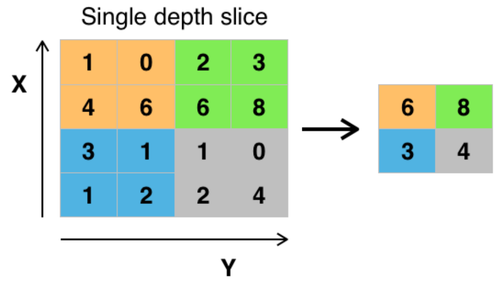

Figure 2.2: Max pooling with a 2x2 filter and stride = 2.

- Average Pooling Layer: Average pooling layer reduces the variance and complexity in the data. It also performs down-sampling by dividing the input into rectangular pooling regions and computing the average values of each region. The Output Size of a Average-Pooling layer is:

$$\left[\frac{n_h - f}{s} + 1 * \frac{n_w - f}{s} + 1\right]$$

where, s = stride, f = number of filters, n_h = input height n_w= input width [4].

- Concat Layer: The Concat layer concatenates its multiple input blobs to one single output blob.

- Dropout Layer: A dropout layer randomly sets input elements to zero with a given probability. Dropout is a technique used to improve over-fit on neural networks [5].

- Fully Connected Layer: The fully connected (FC) layer in the CNN represents the feature vector for the input. This feature vector holds information that is vital to the input.

- Softmax Layer: Softmax assigns decimal probabilities to each class in a multi-class problem. Those decimal probabilities must add up to 1.0. This additional constraint helps training converge more quickly than it otherwise would [6].

2.2 Transfer Learning

Transfer learning is a research problem in machine learning that focuses on storing knowledge gained while solving one problem and applying it to a different but related problem. In transfer learning a model developed for a task is reused as the starting point for a model on a second task.

It is a popular approach in deep learning where pre-trained models are used as the starting point on computer vision and natural language processing tasks given the vast compute and time resources required to develop neural network models on these problems and from the huge jumps in skill that they provide on related problems.

Transfer learning is related to problems such as multi-task learning and concept drift and is not exclusively an area of study for deep learning.

Nevertheless, transfer learning is popular in deep learning given the enormous resources required to train deep learning models or the large and challenging datasets on which deep learning models are trained.

Transfer learning only works in deep learning if the model features learned from the first task are general.

This form of transfer learning used in deep learning is called inductive transfer. This is where the scope of possible models (model bias) is narrowed in a beneficial way by using a model fit on a different but related task.

Two common approaches of transfer learning are as follows:

- Develop Model Approach
- Pre-trained Model Approach

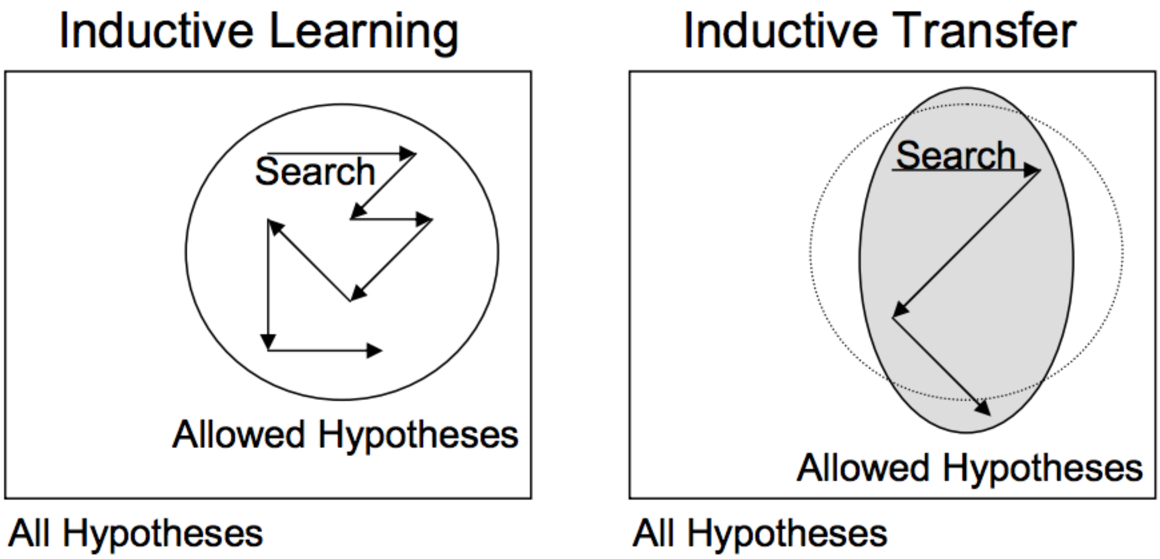

Figure 2.3: Depiction of Inductive Transfer Taken from Transfer Learning.

Develop Model Approach

1. Select Source Task: You must select a related predictive modeling problem with an abundance of data where there is some relationship in the input data, output data, and/or concepts learned during the mapping from input to output data.

2. Develop Source Model: Next, you must develop a skillful model for this first task. The model must be better than a naive model to ensure that some feature learning has been performed.

3. Reuse Model: The model fit on the source task can then be used as the starting point for a model on the second task of interest. This may involve using all or parts of the model, depending on the modeling technique used.

4. Tune Model: Optionally, the model may need to be adapted or refined on the

input-output pair data available for the task of interest.

Pre-trained Model Approach

1. Select Source Model: A pre-trained source model is chosen from available models. Many research institutions release models on large and challenging datasets that may be included in the pool of candidate models from which to choose from.

2. Reuse Model: The model pre-trained model can then be used as the starting point for a model on the second task of interest. This may involve using all or parts of the model, depending on the modeling technique used.

3. Tune Model: Optionally, the model may need to be adapted or refined on the input-output pair data available for the task of interest.

Chapter 3

Related Work

Several approach has been done to classify food from images. In previous years many feature based model is being used to classify food images. SCD [7], EFD [7], GFD [7] and LBP [8] are the common features that has been used to classify food images. In modern literature there are nerual networks specially convolutional neural networks have been used to classify food images.

3.1 Featuee Based Model

Feature based recognition is quite different from a variety of object recognition algorithms [9, 10, 11]. These algorithms rely on a single type of feature: an edge. Because the world is full of edges that look roughly the same, the set of edges extracted from an image cannot be directly compared to the set of edges extracted from a model object. Before a comparison can be made one must first discover the mapping, called a correspondence, from image edges to model edges. There are many approaches to finding a correspondence, but in every case significant computation is required. FBR attempts to finesse the problem of finding a correspondence. Instead it constructs representations of the image and the model that are directly comparable. As a result recognition can be viewed as a classification problem. FBR proceeds by computing a number of properties of the input image and combining them into a feature vector. An object model is a set of feature vectors associated with a set of representative images of the object. A novel image is classified by computing the feature vector of the image and comparing it directly to the model vectors. An image is identified as an instance of an object when that object model contains the feature vector that is closest to the image feature vector.

The scale-invariant feature transform (SIFT) is an algorithm in computer vision to detect and describe local features in images. In [12] they implemented k-nearest neighbor and vocabulary tree 42 food categories with 1453 images. For distance measure the chose L1 norm for the SCD, EFD and GFD features and the Euclidean distance (L2 norm) for the DCD feature. They got 64.5 % top 1 accuracy and 84.2 % top4 accuracy with combination of DCD, MDSIFT, SCD, SIFT features.

In [13] they proposed a method that applies SIFT and LBP features in SVM classifier with PFI dataset. SIFT feature is used detect and describe local features in images and LBP is a type of visual descriptor and was used because it has many advantages. LBP is simple to compute, immune to illumination changes. Support vector machines are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis.

In [14] they proposed a method that classifies food images with sphere shaped support vector machine. The support vector machine can efficiently perform non-linear classification with kernel trick. They applied this method to FoodLog dataset which consisting of 6512 images. They used FCM algorithm to segment food images. FCM is similar to k-means clustering algorithm. In FCM coefficients are assigned randomly to each data point for being in the clusters then centroid for each cluster is computed and for each data point coefficients is computed of being in the clusters and these steps are repeated until convergence. After applying FCM to segment food images they used sphere shaped support vector machine to classify segmented images. They got an accuracy of 95

The authors of [15] used random forests classifier to classify food-101 dataset. Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks, that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. Food-101 is a large dataset with 101 categories. The researchers got 50.76% of accuracy with the RFDC approach.

3.2 Deep Learning Based Model

In present day deep learning is very popular in computer vision. Deep learning (also known as deep structured learning or hierarchical learning) is part of a broader family of machine learning methods based on learning data representations, as opposed to task-specific algorithms. Learning can be supervised, semi-supervised or unsupervised. Deep learning architectures such as deep neural networks, deep belief networks [16] and recurrent neural networks [16] have been applied to fields including computer vision, speech recognition, natural language processing, audio recognition, social network filtering, machine translation, bioinformatics, drug design and board game programs, where they have produced results comparable to and in some cases superior to human experts.

In recent literatures there are also several approaches that used deep convolution neural network to classify food images. Since convolutional neural networks are scalable for large datasets it is more suitable to use convolutional neural network for food image classification. Deep learning was used in [17] to classify UEC-256 food image for computer aided dietary assessment system. They obtained a top-1 accuracy of 54.7 % and 81.5% accuracy of top-5 In [18] CNN was used to classify food images with food-11 dataset in order to build a dietary management system.

A Pre-trained deep neural network was applied in [19]. The deep convolutional neural network was pre-trained on ImageNet with 1000 food related categories than fine tuned to classify UEC-FOOD100 dataset. They achieved 78.77% top-1 accuracy.

In [20] they used GoogLeNet to classify Thai fast food images in TFF food dataset. They achieved 88.33% accuracy for 11 classes. In [21] they implemented and compared several convolutional neural network models with food-11 dataset. They got 70.12% with their proposed approach, 80.51% with Caffenet and 82.07% accuracy with Alexnet.

We see that deep learning approaches have better result than traditional feature based models with larger datasets. In our thesis we used convolutional neural network to classify food-11 dataset. We tried with a convolutional neural network built from scratch as well as with transfer learning that uses Inception V3 [22] model.

Chapter 4
Methodology

In our thesis we used food 11 dataset for our research. We developed a CNN from scratch to classify food images. We also used transfer learning from Inception v3 model which was pre-trained with Imagenet [23] method in our work. This chapter describes the model of our convolutional neural networks and transfer learning method in details. The methodology is depicted in figure 4.1.

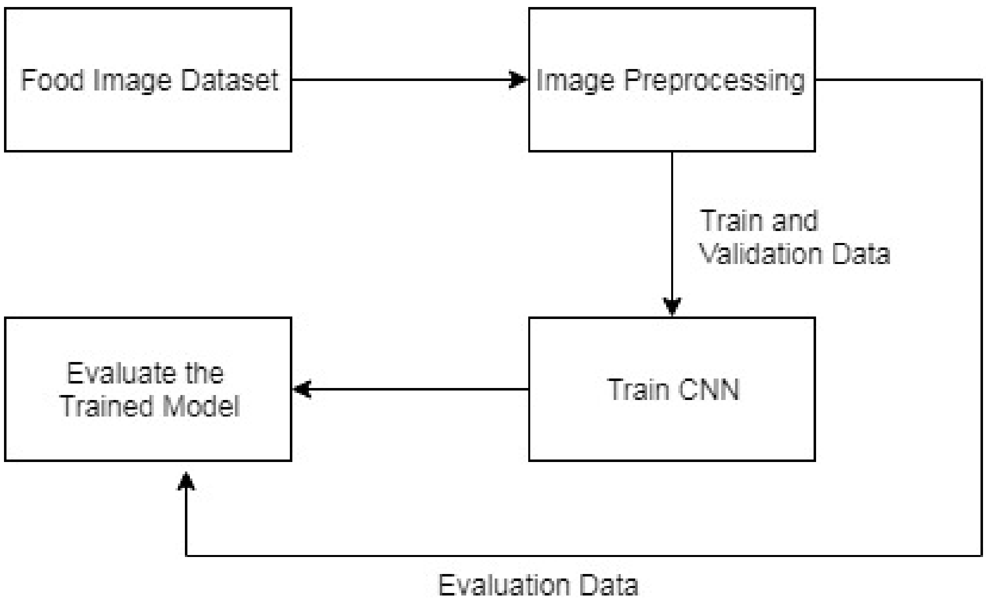

Figure 4.1: Methodology of food image classification.

4.1 Dataset

We used the Food-11 dataset for our research. The dataset was created by the authors who proposed [24]. The dataset consists of 16643 images grouped into 11 major food categories. The 11 food categories are Bread, Dairy product, Dessert, Egg, Fried food, Meat, Noodles/Pasta, Rice, Seafood, Soup, and Vegetable/Fruit. The food images are divided into three parts: training set with 9866 images, validation set with 3430 images and evaluation set with 3347 images.

4.2 Pre-Processing

We applied some image pre-processing technique to increase efficiency to our system. First we re-sized all our images to 224 x 224 x 3 to increase processing time and also to fit in our convolutional neural network model. After that we applied following image pre-processing techniques.

- Feature wise Center = False: Set input mean to 0 over the dataset.
- Sample wise Center = False: Set each sample mean to 0.
- Feature wise standard deviation normalization = False: Divide inputs by standard deviation of the dataset.
- Sample wise standard deviation normalization= False: Divide each input by its standard deviation.
- ZCA whitening = False: Apply ZCA whitening.
- Rotation Range = 0: Randomly rotate images in the range from 0 to 180 degrees.
- Width shift range = 0.2: Randomly shift images horizontally.
- Height shift range = 0.2: Randomly shift images vertically.

Figure 4.2: Images from each class of the dataset.

- Horizontal flip = True: Randomly flip images.

These pre-processing helped us to make our CNN insensitive to the exact position of the object in the image.

4.3 Training Convolution Neural Network Model from Scratch

We developed a convolutional neural network model from scratch to classify food images. We trained our dataset for 100 epoch with Adam optimizer with learning rate 0.01 and 0.01 decay.

The architecture of our convolutional neural network is given below in figure 4.3.

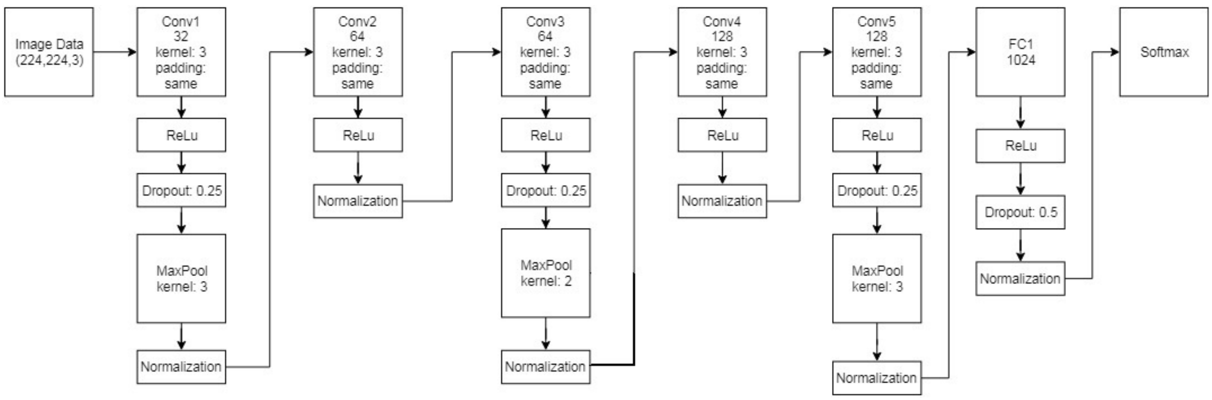

Figure 4.3: Architecture of our Convoluional Neural Network.

4.4 Transfer Learning

We applied also applied transfer learning in our research. We chose google inception v3 which was pre-trained Imagenet dataset. We trained our dataset of SGD optimizer with initial learning rate of 0.01 and 0.9 momentum. We used a learning scheduler to set learning rate to 0.002 after 15 epochs and 0.0004 after 32 epochs. The architecture of Inception V3 is given in figure 4.4.

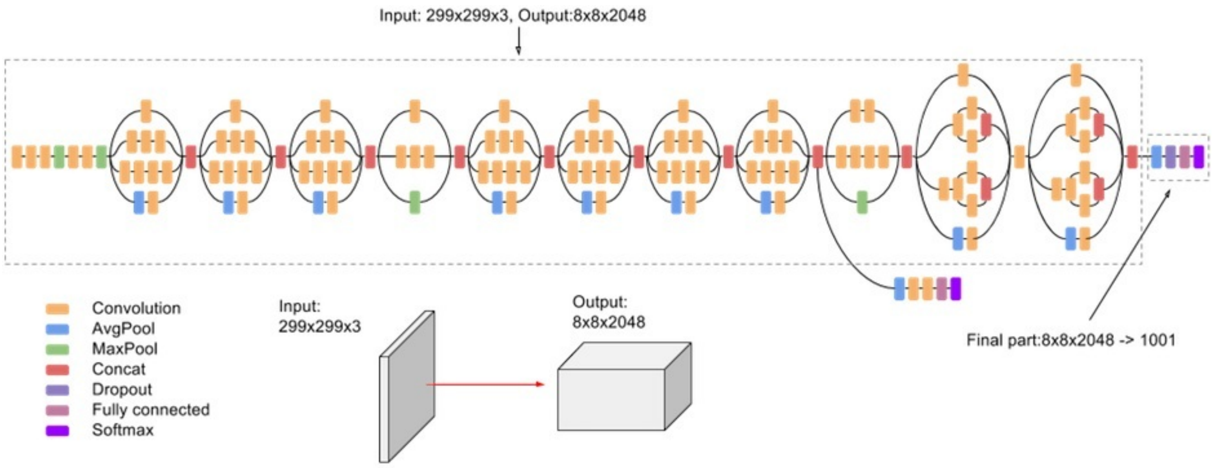

Figure 4.4: Architecture of Inceptin V3.

Chapter 5

Experimental Results

This section describe and analyze the results we got from our experiments. At the beginning we re-sized all of our images to $224 * 224$ do that the images fit into the neural network model that we built. Then we applied the pre-processing that were described in Methodology section. After that we trained our model with training and validation set for 100 epochs. We used Adam optimizer with learning rate 0.01 and 0.01 decay. After finishing the training we used the evaluation set of our dataset to evaluate the trained model.

For transfer learning we re-sized our images to $299 * 299$ to fit images into Inception V3 neural network. We applied the same pre-processing that were applied in previously. After that we trained our model with the same training and validation set and evaluated our model with same evaluation set. In this approach we used SGD optimizer with initial learning rate of 0.01 and 0.9 momentum. We used a learning scheduler to set learning rate to 0.002 after 15 epochs. We trained the neural network for 32 epochs.

After getting the result we compared our results with accuracy and loss curve. Then we also compared our approach with the approach proposed in [21].

5.1 Evaluation of Model

Our dataset was divided into three parts: training, validation and evaluation. We used training and validation parts of the dataset while training the model and we used evaluation part of our dataset during the evaluation of our model. We resized the images of evaluation part in 224 x 224 x 3. We evaluated the accuracy of the model by true positive (TP), true negative (TN), false positive (FP) and false negative (FN) after classification.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

5.2 Obtained Results

The results obtained by running different models with food-11 dataset are given in Table 5.1.

Table 5.1: Test results of models

Model	**Accuracy**
Proposed CNN	74.70%
Inception V3 (Transfer Learning)	92.86%
Proposed CNN in [17]	70.12%

We can see that our model have better accuracy than the model proposed in the same dataset in [17]. We also see that Inception V3 has an accuracy of 92.86%. Since inception v3 model has been already pre-trained on imagenet it transfered its learning in our dataset. So inception v3 has better result as expected.

We can also see the loss and accuracy curve of our proposed model in figure 5.1 and 5.2. We see that the gap between validation and train accuracy decreased after several epochs so there is low overfitting. We also see that the loss curve decreased after several epochs.

From the accuracy and loss curve from inception v3 model we also see that there was a big gap between validation and training accuracy but it reduced after some epochs. The loss curve also reduced after several epochs.

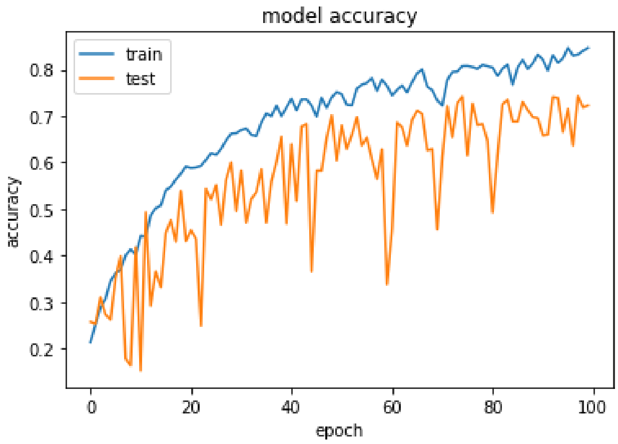

Figure 5.1: Accuracy curve of proposed model.

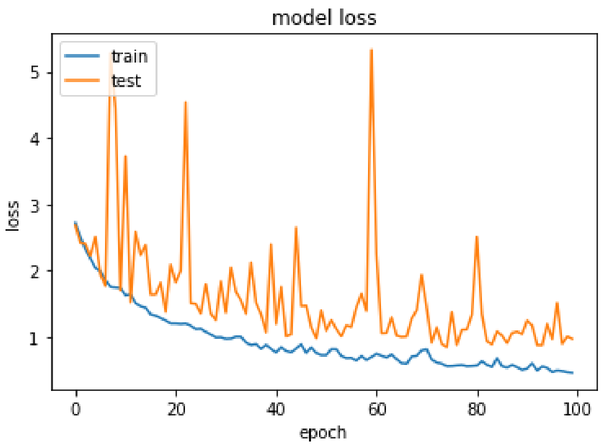

Figure 5.2: Loss curve of proposed model.

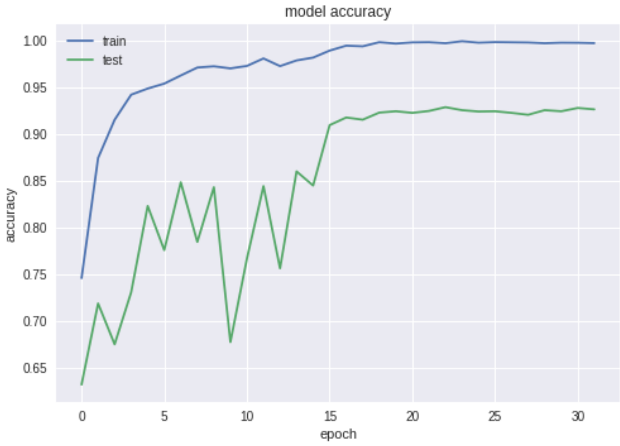

Figure 5.3: Accuracy curve of Inception V3 (Transfer Learning).

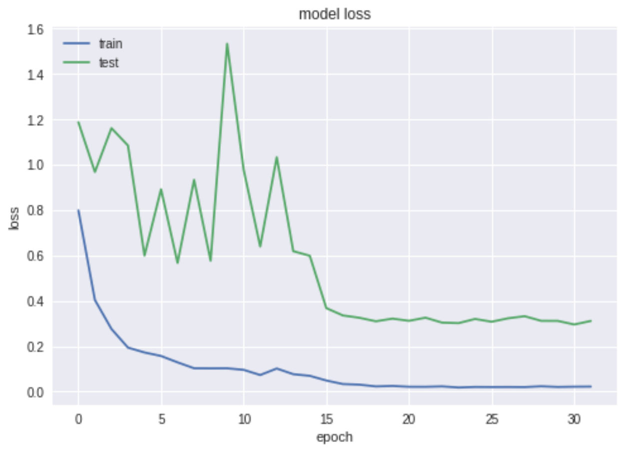

Figure 5.4: Loss curve of Inception V3 (Transfer Learning).

Chapter 6

Concluson and Future Work

Food image classification is a challenging task because there is so much variations even in same class of food image. In this paper we tried to use CNN to classify food images. From our experiment we got 74.70 % accuracy from our proposed model in food classification task. However the there can be still some improvement to reduce the gap between training and test accuracy in our model. Other neural network models such as recurrent neural network [25], dilated convolutional neural network [26], etc can be applied to classify food images. Convolution neural network models take time for computation but once the model is trained it can be easily used for classification. We can also improve our research by applying feature based models which will take less computational time. We will try to make an android application in future to detect food from images. The CNN model should be trained on more category of foods to work efficiently with social media to classify foods because there are thousands more category of foods worldwide. If the CNN can be trained with more category of foods then it will help social media platform to classify foods efficiently even if the food is not labeled in caption. It will also help the restaurants to advertise their targeted audience efficiently. It will also help consumer to choose restaurants with similar kind of food We will also try to build a recommender system that will suggest restaurants to the people according to their food preferences.

Bibliography

[1] W. E. Grimson, "On the recognition of curved objects," vol. 11, no. 6. Washington, DC, USA: IEEE Computer Society, Jun. 1989, pp. 632–643. [Online]. Available: https://doi.org/10.1109/34.24797

[2] D. H. Ballard, "Generalizing the hough transform to detect arbitrary shapes," vol. 13, 1981, pp. 111–122.

[3] R. C. Bolles and R. A. Cain, "Recognizing and locating partially visible objects: The local-feature-focus method," vol. 1, no. 3, 1982, pp. 57–82. [Online]. Available: https://doi.org/10.1177/027836498200100304

[4] D. Scherer, A. Müller, and S. Behnke, "Evaluation of pooling operations in convolutional architectures for object recognition," pp. 92–101, 2010.

[5] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, pp. 1929–1958, 2014. [Online]. Available: http://jmlr.org/papers/v15/srivastava14a.html

[6] C. Bishop, *Pattern Recognition and Machine Learning.* Springer-Verlag, 2006.

[7] Y. He, C. Xu, N. Khanna, C. J. Boushey, and E. J. Delp, "Analysis of food images: Features and classification," in *2014 IEEE International Conference on Image Processing (ICIP)*, Oct 2014, pp. 2744–2748.

[8] D. chen He and L. Wang, "Texture unit, texture spectrum, and texture analysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 28, no. 4, pp. 509–512, July 1990.

[9] Y. He, C. Xu, N. Khanna, C. J. Boushey, and E. J. Delp, "Analysis of food images: Features and classification," in *2014 IEEE International Conference on Image Processing (ICIP)*, Oct 2014, pp. 2744–2748.

[10] Z. Zong, D. T. Nguyen, P. Ogunbona, and W. Li, "On the combination of local texture and global structure for food classification," in *2010 IEEE International Symposium on Multimedia*, Dec 2010, pp. 204–211.

[11] S. J. Minija and W. R. S. Emmanuel, "Food image classification using sphere shaped support vector machine," in *2017 International Conference on Inventive Computing and Informatics (ICICI)*, Nov 2017, pp. 109–113.

[12] L. Bossard, M. Guillaumin, and L. Van Gool, "Food-101 – mining discriminative components with random forests," in *European Conference on Computer Vision*, 2014.

[13] C. Liu, Y. Cao, Y. Luo, G. Chen, V. Vokkarane, and Y. Ma, "Deepfood: Deep learning-based food image recognition for computer-aided dietary assessment," *CoRR*, vol. abs/1606.05675, 2016. [Online]. Available: http://arxiv.org/abs/1606.05675

[14] K. Yanai and Y. Kawano, "Food image recognition using deep convolutional network with pre-training and fine-tuning," in *2015 IEEE International Conference on Multimedia Expo Workshops (ICMEW)*, June 2015, pp. 1–6.

[15] N. Hnoohom and S. Yuenyong, "Thai fast food image classification using deep learning," in *2018 International ECTI Northern Section Conference on Electrical, Elec-*

tronics, Computer and Telecommunications Engineering (ECTI-NCON), Feb 2018, pp. 116–119.

[16] Y. Hua, J. Guo, and H. Zhao, "Deep belief networks and deep learning," in *Proceedings of 2015 International Conference on Intelligent Computing and Internet of Things*, Jan 2015, pp. 1–4.

[17] G. zsert Yiit and B. M. zyildirim, "Comparison of convolutional neural network models for food image classification," in *2017 IEEE International Conference on INnovations in Intelligent SysTems and Applications (INISTA)*, July 2017, pp. 349–353.

[18] D. J. Attokaren, I. G. Fernandes, A. Sriram, Y. V. S. Murthy, and S. G. Koolagudi, "Food classification from images using convolutional neural networks," in *TENCON 2017 - 2017 IEEE Region 10 Conference*, Nov 2017, pp. 2801–2806.

[19] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," vol. abs/1512.00567, 2015. [Online]. Available: http://arxiv.org/abs/1512.00567

[20] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," in *Proceedings of the IEEE*, 1998, pp. 2278–2324.

[21] M. Matsugu, K. Mori, Y. Mitari, and Y. Kaneda, "Subject independent facial expression recognition with robust face detection using a convolutional neural network," vol. 16, no. 5, 2003, pp. 555 – 559, advances in Neural Networks Research: IJCNN '03. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0893608003001151

[22] W. Zhang, K. Itoh, J. Tanida, and Y. Ichioka, "Parallel distributed processing model with local space-invariant interconnections and its optical architecture,"

vol. 29, no. 32. OSA, Nov 1990, pp. 4790–4797. [Online]. Available: http://ao.osa.org/abstract.cfm?URI=ao-29-32-4790

[23] J. Deng, W. Dong, R. Socher, L. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, June 2009, pp. 248–255.

[24] H. Hassannejad, G. Matrella, P. Ciampolini, I. De Munari, M. Mordonini, and S. Cagnoni, "Food image recognition using very deep convolutional networks," in *Proceedings of the 2Nd International Workshop on Multimedia Assisted Dietary Management*, ser. MADiMa '16. New York, NY, USA: ACM, 2016, pp. 41–49. [Online]. Available: http://doi.acm.org/10.1145/2986035.2986042

[25] M. Liang and X. Hu, "Recurrent convolutional neural network for object recognition," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015, pp. 3367–3375.

[26] Y. Li, X. Zhang, and D. Chen, "Csrnet: Dilated convolutional neural networks for understanding the highly congested scenes," *CoRR*, vol. abs/1802.10062, 2018. [Online]. Available: http://arxiv.org/abs/1802.10062