Multiclass Food Image Classification Based on Inception-v3 Transfer Learning Model: A Design

Prof. Kanchan V. Warkar[1]

M.Tech CSE Department
Bapurao Deshmukh College
Of Engineering Sewagram

Anamika B. Pandey[2]

M.Tech. CSE Department
Bapurao Deshmukh
College of Engineering Sewagram

---***---

ABSTRACT Multiclass Food Classification is one of the applications of visual object recognition in the field of computer vision. Convolutional neural networks (CNN) are at the heart of most state-of-the-art computer vision solutions for a wide range of tasks. In this paper, we propose a new approach based on Deep Learning for food image recognition. Based on the Inception V3 model of the Tensor flow platform, we use the transfer learning technology to retrain the last layer of the famous Inception V3 architecture developed by Google for our distribution approach. methods based on geometric transformation were implemented to improve the number of training images. Our method shows promising results with an overall accuracy of approximately 92% in accurate recognition of food images in preventing the over-fitting problem.

INDEX TERMS Food Image Recognition; Object Detection; Deep Learning; Classification; TensorFlow; Inception-v3; Transfer Learning; Computer Vision; Convolutional Neural Network

I.INTRODUCTION

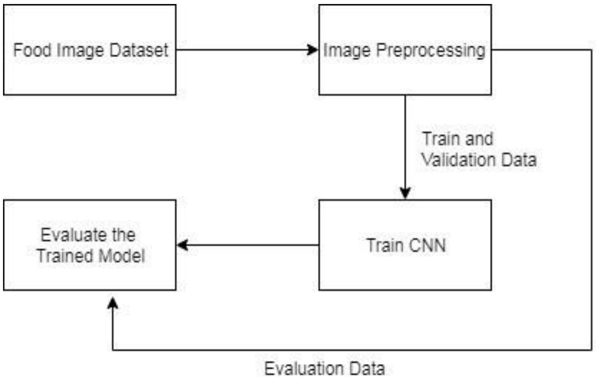
With the advent of social media, thousands of photos are uploaded to social media, and the rise of mobile devices puts cameras in everyone's pocket. Suddenly, users were no longer able to log into their social media accounts without having a desktop device and it started quickly. Instagram-based social networks such as Instagram and Snapchat have emerged to live up to this new reality on the same phone or tablet you use to engage and take photos on social media platforms on the go. Twitter responds to user behavior by giving you more opportunities to center your images in front. Marketers know how important social media marketing is.

The information that overflows on the Internet, social media platforms, is enormous. This data represents the challenges and opportunities we strive to market effectively, protect our image, and enter the era of information overload. The potential hidden in this constantly growing pool of online images. For the brand, this means accessing more data than ever before, especially image-based data. Social media users fully embrace the concept of sharing photos instead of text or text. Snapchat has 184 million users per day. In 2017 (up from 46 million in early 2014), Instagram increased from 800 million in September 2017 to 1 billion monthly users identifies this growth trend the ability to analyze, analyze and utilize image recognition technology had to exist in the future. Most of his digital marketing was dominated by visual data. Otherwise, the brand would have missed an entire pile of valuable data. Media monitoring may not be captured. If you miss a great opportunity to learn and communicate with your customers, artificial intelligence and image recognition make it easier for marketers to find visuals on social media without explicit textual mention. Food images are being uploaded on social media and with food identification, social media can group people based on their food choices. Social media platform for advertising target audiences. Computer vision and image processing techniques are currently being used in many fields. Image: Food identification is a challenging task because food images have less variation within the classroom. Data Based Classification It is possible to use it in a cheaper device. Today, there are inexpensive smartphones with high computing power that are capable of processing high-definition image data, so the model described in this article can be made on smartphones.

Transfer learning is the reuse of a pre-trained model for a new problem, it is very popular nowadays in deep learning because it can train deep neural networks with relatively little data, and it is very useful in data science because of most real problems. , you don't have millions of data points marked to train these complex models. Let's take a look at what transfer learning is, how it works, why and when to use it. Includes several resources for models that have been previously trained in learning transfers for example, when you train the classifier to predict whether an image contains food, you can use the knowledge gained during training to recognize drinks, for example, if you trained a simple classifier to predict, if the image includes a backpack, you can use the knowledge gained by the model during training. Recognizes other objects such as sunglasses. In passing on knowledge, we try to basically apply what we have learned in one task to improve the generalization in another. We transfer the weights that the network learned in "task A" to the new "task B". The idea is to use models that have learned from business in a new business with a lot of data training cards available and with little data. Instead of starting with learning processes from the beginning.

Transfer learning is mainly used for natural language processing tasks such as computer vision and emotion analysis due to the large amount of computing power required.

Transfer learning is actually a machine learning technology. No, but we can think of it as . For example, design methodology in areas such as active learning. It is not the exclusive part or research area of machine learning. Nevertheless, it is very popular in combination with neural networks that require large amounts of data and processing power.



Based on the original food-101 [1] dataset with 101 food categories. All images are rescaled to a maximum side length of 512 pixels. Use a subset of the four food categories [Chicken Curry, Hamburger, Omelet, Waffle]. See Figure 1 for this assignment. The data consists of three main subfolders: training, validation, and testing. The training data consists of 1000 images per class, with up to 500 validation images and up to 500 test images per class. The dataset has not been (intentionally) cleaned up and therefore contains some noise. It is mainly displayed in dark colors and in some cases has the wrong label. The dataset is not complete, which makes the problem even more difficult. However, it uses the assigned label).We have taken 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.

II. EXPERIMENT

This section focuses on the construction process of the food classification model. The construction process of the model divided into few steps.

2.1. Convolutional Neural Network

A convolutional neural network is network architecture for deep learning . CNNs are deep artificial neural networks that are primarily used to classify images cluster them by similarity and perform object recognition within scenes. A CNN is comprised of one or more convolutional layers followed by one or more fully connected layers as in a standard multilayer neural network. It learns directly from images. A CNN can be trained to do image analysis tasks including classification, object detection, segmentation and image processing.

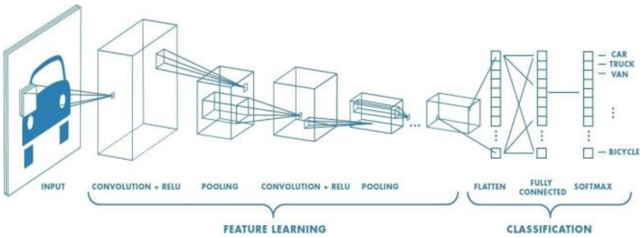


Figure 1. CNN Architecture

The CNN architecture of the neural network explicitly assumes that the input is an image so that you can encode certain properties into the architecture. This makes the implementation of forward functions more efficient and significantly reduces the number of parameters in the network. Layer description with the convolutional neural network described below.

- Convolutional layer: Convolution is the core building block of a convolutional network, doing most of the computationally difficult tasks. Three metadata (depth, stride and padding)

- 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 of the initial representation.

- 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.

- Dropout Layer: A dropout layer randomly sets the input elements to zero with a given probability. The CNN's concatenation (FC) layer represents a feature vector for the input. This feature vector holds the information that is important to the input.

- 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.

2.2 Inception V3 Model

Inception V3 is the 2015 model of Google's image recognition Inception architecture. The InceptionV3 code utilizes TF-Slim, which looks like a kind of TensorFlow abstraction library that makes writing convolutional neural networks(CNN)more compact and easier. ImageNet on a subset of the Google ImageNet database used in the Large Visual Recognition Challenge (ILSVRC). The model is trained with over 1 million images and can be categorized into 1000 objects such as mouse, keyboard, pencil and animal. As a result, the Inception model has learned the rich features and representations of a wide range of images.

Table 1 shows the model architecture.

Type	Patch size / Stride	Input size
Conv	3 x 3 / 2	299 x 299 x 3
Conv	3 x 3 / 1	149 x 149 x 32
Conv padded	3 x 3 / 1	147 x 147 x 32
Pool	3 x 3 / 2	147 x 147 x 64
Conv	3 x 3 / 1	73 x 73 x 64
Conv	3 x 3 / 2	71 x 71 x 80
Conv	3 x 3 / 1	35 x 35 x 192
3x Inception	As in figure 2	35 x 35 x 288
5x Inception	As in figure 3	17 x 17 x 768
2x Inception	As in figure 4	8 x 8 x 1280

Pool	8 x 8	8 x 8 x 2048
Linear	Logits	1 x 1 x 2048
SoftMax	Classifier	1 x 1 x 1000

Fig 1: Shows Inception v3 Architecture

The output size of each module is the input size of the next module. CNN can be used in three ways. We have used the Transfer learning approach. Transfer learning is based on the concept, that the knowledge for solving one type of problem can be used to solve a similar problem. Using Inception V3 model for image classification by retraining it is one kind of transfer learning. For the inception part of the network, we have 3 inception modules at the 35 × 35 with 288 filters each as shown in figure 2.

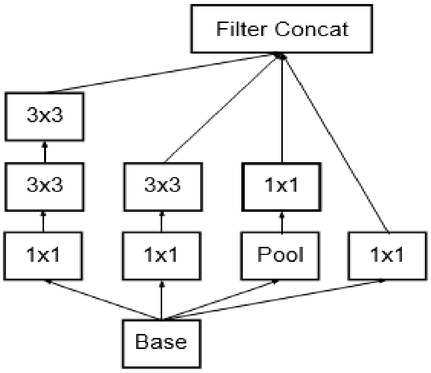


Figure 2. Inception module of 35 x 35 grid with 288 filters

This is reduced to 17 × 17 grid with 768 filters using technique known as grid reduction as described in section 5 of [12]. This is followed by 5 instances of the factorized inception module as shown in figure 3.

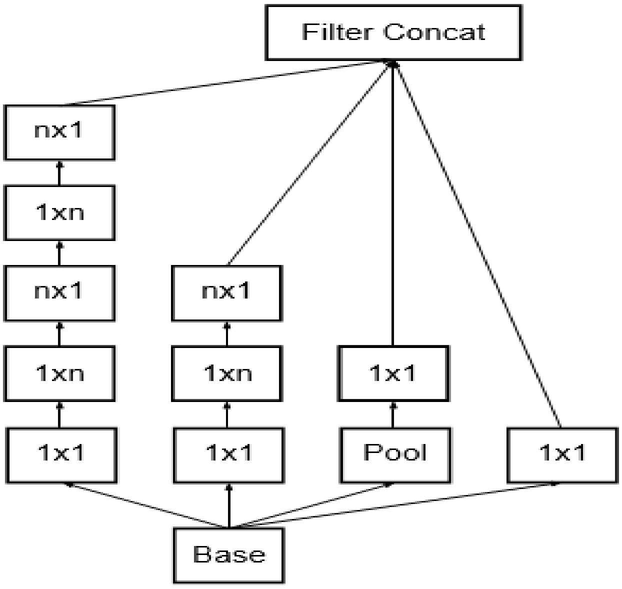


Figure 3. Inception modules after the factorization of the n × n convolutions.

In Inception V3 architecture, we have n = 7 for the 17 × 17 grid

This is reduced to 8 × 8 × 1280 grid with the grid reduction technique. At the coarsest 8 × 8 level, we have two Inception modules as shown in figure 3, with a concatenated output filter bank size of 2048 for each tile

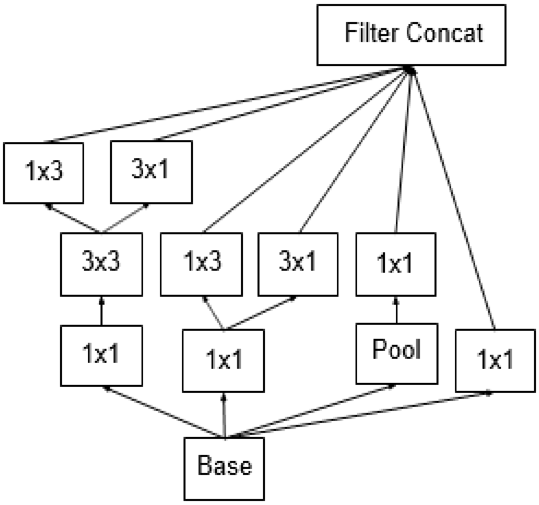


Figure 4. Inception modules with expanded the filter bank outputs.

This architecture is used on the coarsest (8 × 8) grids to promote high dimensional representations.

2.3 Dataset Collection

Based on the original food-101 [1] dataset with 101 food categories. All images are rescaled to a maximum side length of 512 pixels. Use a subset of the four food categories [Chicken Curry, Hamburger, Omelette, Waffle].The data consists of three main subfolders: training, validation, and testing. The training data consists of 1000 images per class, with up to 500 validation images and up to 500 test images per class. The dataset has not been (intentionally) cleaned up and therefore contains some noise. It is mainly displayed in dark colours and in some cases has the wrong label. The dataset is not complete, which makes the problem even more difficult. However, it uses the assigned label).

2.4 Data Augmentation

We artificially expand the dataset to avoid overfitting. This data will create some variance that may occur when someone else takes a new web or real-life data. After gathering data for each class, we increase the dataset in 5 different methods, the following methods are:

1.Rotate left -30 degree

2.Rotate right +30 degree

3.Flip horizontally about Y axis

4.Shear by a certain amount

5.Rotate left +90 degree

2.5 Proposed Inception Model

The Inception V3 model is a deep neural network so it is very challenging for us to train it directly with a low configured computer because it will take at least a few days to train the model. TensorFlow library provides us with the ability to retrain the final layer of Inception model for new categories using transfer learning. We employ the transfer learning approach that keeps the parameters of the previous layer and removes the last layer of the Inception V3 model for the retrain purpose. The number of output nodes in the last layer is equal to the total number of categories in the dataset.

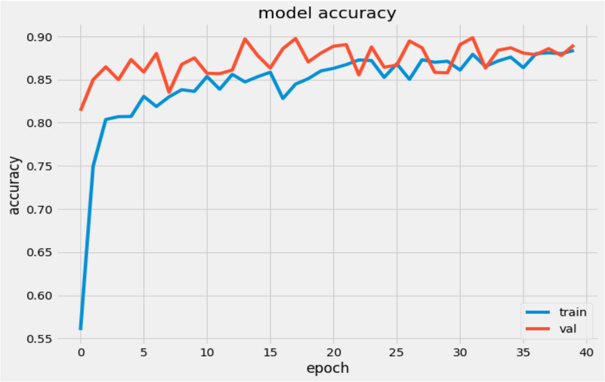
2.6 Training the model

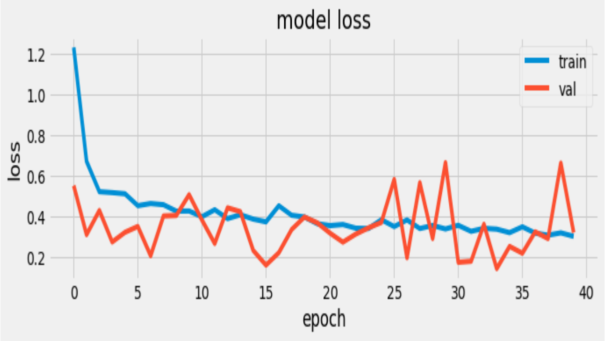
After creating bottleneck files for all data, the main training of the network's final layer started. The training process runs efficiently by feeding the cache value into the layer for each image. The truth label for each image is also fed into the Ground Truth node. We can see a series of step outputs after the training process, each showing validation accuracy, training accuracy, and cross entropy. The role of training process is to make the cross- entropy as small as possible by keeping an eye on whether the loss keeps trending downwards and neglecting the short-term noise. This model script runs 4,000 training steps by default. Each step randomly selects 4 images from the training set, obtains their bottlenecks from the cache, and feeds them into the final layer for predictions. Those predictions are then compared to the actual labels to update the weights of the final layer through a backpropagation process. As the training process goes on, we can see the reported accuracy improved. After all the training steps completed, the script runs a final test accuracy evaluation on a set of images that are kept for testing purpose.

III. RESULT

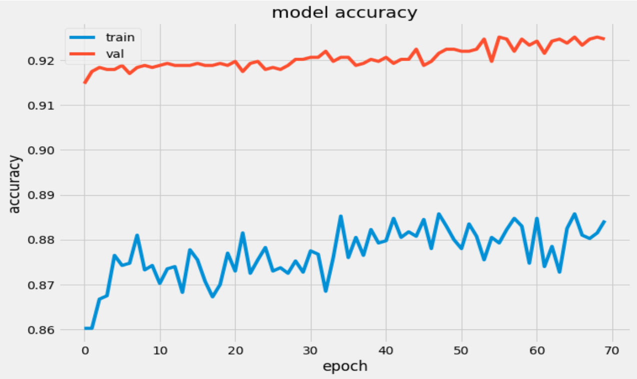
The model (Inception V3 recognition) resulted in an overall accuracy (i.e. the ratio of the number of correctly recognized images to the number of total images) of 92% and 91% for training and test images, respectively

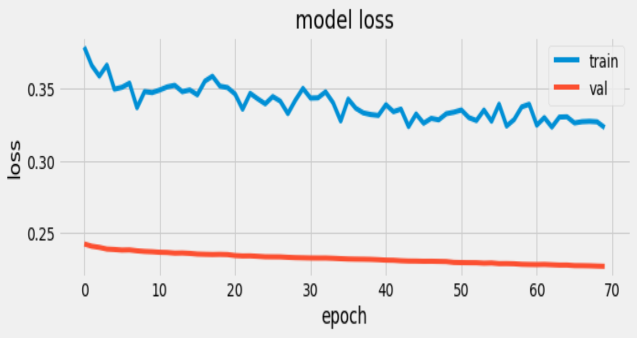
1. Model Before tuning





2. Model After Tuning





IV. REFERENCES

[1] L. Shao, F. Zhu and X. Li, "Transfer Learning for Visual Categorization: A Survey," in IEEE Transactions on Neural Networks and Learning Systems, vol. 26, no. 5, pp. 1019-1034, May 2015, doi: 10.1109/TNNLS.2014.2330900.

[2].C.Szegedy et al., "Going deeper with convolutions," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA,2015,pp. 1-9, doi: 10.1109/CVPR.2015.7298594.

[3]. Neyshabur, B., Sedghi, H., & Zhang, C. (2020). What is being transferred in transfer learning? *ArXiv, abs/2008.11687*.

[4]. Tan C., Sun F., Kong T., Zhang W., Yang C., Liu C. (2018) 'A Survey on Deep Transfer Learning'. In: Kůrková V., Manolopoulos Y., Hammer B., Iliadis L., Maglogiannis I. (eds) Artificial Neural Networks and Machine Learning – ICANN 2018. ICANN 2018. Lecture Notes in Computer Science, vol 11141. Springer, Cham. https://doi.org/10.1007/978-3-030-01424-7_27

[5].Cheng Wang, Lin Hao ; Xuebo Liu ; Yu Zeng ; Jianwei Chen ; Guokai Zhang et. al. C. Wang et al., "Pulmonary Image Classification Based on Inception-v3 Transfer Learning Model," in IEEE Access, vol. 7, pp.146533-146541,2019,doi:10.1109/ACCESS.2019.2946000.