MACHINE LEARNING (BITS F464)
ASSIGNMENT #2

Ethnicity/Nationality Identification

A Report on

Establishing the Ethnicity/nationality of person by just looking at his facial image

GROUP 4
HARJAS (2015A7PS0106P)
MANIK (2015A3PS0203P)

1. Introduction

Deep Learning is a subfield of machine learning concerned with algorithms inspired by the structure and function of the brain called artificial neural networks. The advantage of deep learning models is their ability to perform automatic feature extraction from raw data also known as feature learning. A popular paper in this field [1] states that “deep learning allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction.” These methods have improved the state-of-the-art in speech recognition, visual object/image recognition, object detection and many other domains such as drug discovery and genomics. Deep learning discovers the intricate structure in large data sets by using the backpropagation algorithm to indicate how a machine should change its internal parameters that are used to compute the representation in each layer from the representation in the previous layer.
Deep Learning is focused on training deep (many layered) neural network models using the backpropagation algorithm. The most popular techniques include:

- Long Short-Term Memory Recurrent Neural Networks.
- Convolutional Neural Network

RNN, unlike feedforward neural networks, use their internal memory to process random sequences of inputs. It uses time series information (i.e. what I spoke last will impact what I will speak next.) and are ideal for text and speech analysis. CNN is a type of feed-forward artificial neural network - are variations of multilayer perceptrons which are designed to use minimal amounts of preprocessing.They use connectivity pattern between its neurons is inspired by the organization of the animal visual cortex, whose individual neurons are arranged in such a way that they respond to overlapping regions tiling the visual field and are ideal for images and videos processing.

2. Problem Statement

The problem basically is to identify a person’s ethnicity given a image of his face. The major hints which we infer from the problem statement is that we need to extract out the features of a person’s face and then try to set scores of these features which will then help us to differentiate between ethnicities. Some of the observable features which could be taken as an hypothesis would be the face color, eye size, hair color etc. Thus our problem statement can be formulated as training a machine learning model with images labeled with the ethnicities so that it learns how to classify them based on the extracted features and then predicts any facial image to belong to its correct class/ethnicity.

3.Data Description

The data set “Face Place” has been taken from the CNBC wiki datasets. The dataset was originally created by Tarrlab at Brown University [2]. The dataset consists of multiple images of more than 250 people with different races with multiple views as well as emotions(some participants gave images with different haircuts, with or without spectacles also). In total there are about 6000 images distributed in four classes. The training dataset consists of 80% of the images totaling to 4480 images and the testing dataset consists of the rest 1120 images.
The images are in jpeg format, 250x250 72 dpi 24 bit color.

Class	Number of Images
African-American	1199
Asian	1525
Caucasian	2000
Hispanic	878

Table 1: Distribution of images for each class/ethnicity

Figure 1: A sample from the dataset used

4. Model Description

4.1 Theoretical Description

We have built a CNN for this problem. Convolutional Neural Networks can learn extremely complex mapping functions when trained on enough data. All CNN models follow a similar architecture, as shown in the figure 2. We perform a series convolution and pooling operations, followed by a number of fully connected layers. If we are performing multiclass classification the output is softmax. CNN is trained the same way like ANN, backpropagation with gradient descent. Due to the convolution operation it's more mathematically involved. A CNN model can be thought as a combination of two components: feature extraction part and the classification part. The convolution + pooling layers perform feature extraction. For example given an image, the convolution layer detects features such as two eyes, long ears, four legs, a short tail and so on. The fully connected layers then act as a classifier on top of these features, and assign a probability for the input image being a dog. The convolution layers learn such complex features by building on top of each other. The first layers detect edges, the next layers combine them to detect shapes, to following layers merge this information to infer that this is a nose.

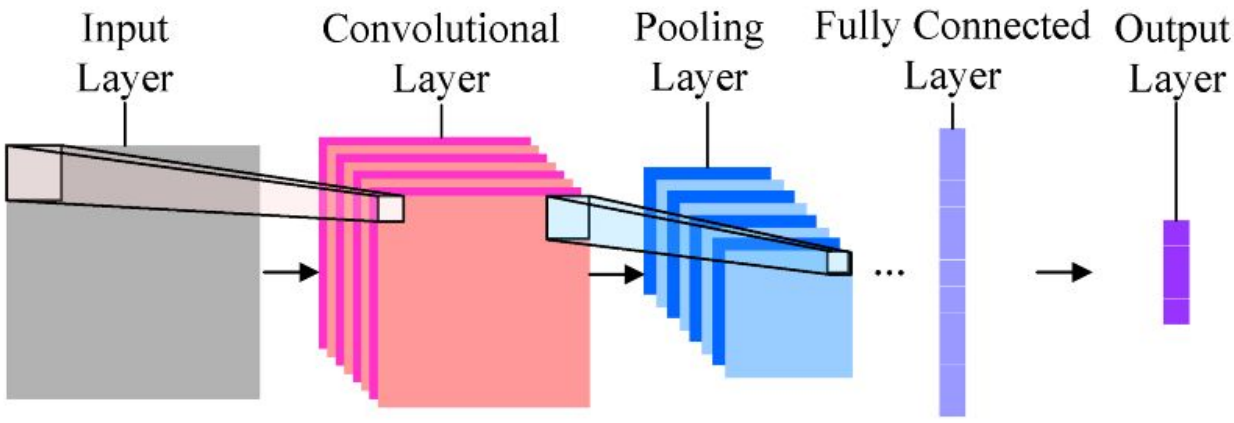

Figure 2: Representation of CNN

4.2 Transfer Learning

We have used transfer learning to first pretrain our model over an already trained model and then we have added our convolution layers at the end which act as the classifiers. The basic concept used here is that the initial layers of the model are basically used to extract out the features of the image they never act as the differentiators. Only the last few or the last layer acts as the differentiator. The pretrained model which we have used were originally trained to classify dogs and cats for the imagenet competition. We have trained our model over two imagenet models- MobileNet and VGG16 and then compared the thus obtained accuracies.

The model has been prepared using *keras*[5] library with *tensorflow* at the backend.The data has been loaded using image_data_generator and then has been preprocessed over the initial layers of the imagenet model.The data then passes through 3 dense layers and one fully connected last layer which acts as the output. The optimizer used in the model is adam optimizer. This optimizer is optimum for data with a large number of features.This optimizer is based on adaptive estimates of lower order moments.The loss function used by the model is categorical cross-entropy since it is a multiclass problem. We have even provided the provision of making a few layers untrainable which helps in increasing the accuracy of our model.

5.Results

We have trained our model with two different imagenet models - MobileNet and VGG16 and then compared their accuracies. It was observed that the model trained best with VGG16.The most probable reason for this could be that the dataset for which VGG16 was originally modeled would have been more similar to our data as compared to the data for which MobileNet was modeled.
For testing purposes, we have divided the dataset of images in the ratio of 80:20 where 80% indicates the number of images used for training the model and 20% are the ones on which our models were tested.

Model Name	Accuracy(%)
VGG16	89.32
MobileNet	83.09

Table 2: Results given by the CNN Model

6.Conclusion

Deep learning is a particular kind of machine learning that achieves great power and flexibility by learning to represent the world as nested hierarchy of concepts, with each concept defined in relation to simpler concepts, and more abstract representations computed in terms of less abstract ones. It has one great advantage of feature engineering which can be very helpful in problems related to images. However the data and computation power required are very high. CNN is a very fundamental deep learning technique. It can help us extract features from the facial images and then classify them in a similar fashion to ANN. Transfer Learning can be used as it requires less data since it already has trained layers in the CNN. Deep learning can be used to model the working of the human brain for image detection(related problems) as the framework underneath and working are very similar.

7. Appendix

Code used for building the machine learning models can be viewed at :
https://github.com/harjas27/Ethnicity-detecction-using-facial-images

8.References

[1] LeCun, Yann & Bengio, Y & Hinton, Geoffrey. (2015). Deep Learning. Nature. 521. 436-44. 10.1038/nature14539.
[2] Face Place dataset created by the Tarrlab at Brown University.
[3]https://towardsdatascience.com/applied-deep-learning-part-4-convolutional-neural-networks-584bc134c1e2
[4] https://towardsdatascience.com/keras-transfer-learning-for-beginners-6c9b8b7143e
[5] https://keras.io/