A Convolution Neural Network for Classification of Indian Faces

Samarth Sarin
Mukesh Patel School of Technology Management and Engineering
Narsee Monjee Institute of Management Studies
Mumbai, India

Siba Panda
Mukesh Patel School of Technology Management and Engineering
Narsee Monjee Institute of Management Studies
Mumbai, India

ABSTRACT
This paper is a modest attempt to compare different Indian faces into three categories that are North Indian, East Indian and South Indian. We have not taken West Indian into consideration as there are no distinct features in those faces. An artificial neural network is proposed for face classification in this paper. Images were collected from online resources and pre-processed before they were feed in the neural network. Experiments were conducted based on extracting features from the faces of Indian People belonging to different regions. The accuracy of face classification was up to 81% and the optimal choice of hyperparameters of the neural network for this task is discussed.

Keywords
Indian face Classification, North Indian, East Indian, South Indian deep learning, convolution neural network.

1. INTRODUCTION
Image pre-processing and object recognition is becoming a boom area of research both for the academicians and practitioner. Neural Networks are becoming fundamental for the classification of images because of their ability to learn and their high accuracy. With an increase in computational devices available they are widely used in the area of research.

Facial Recognition is a topic in which research is at its highest point. People are using different algorithms for face recognition like support vector machine and artificial neural networks which have been outperformed by convolution neural networks based on the ability of feature extraction and also the accuracy of the models. Convolutional Neural Networks (CNN) works basically on feature extraction. Hence it's powerful as it extracts useful features from the images and then tries to search for similar features on the new prediction image which is the unseen image by the model. Image processing is defined as the manipulation of digital images with the use of computers. Image processing involves basic operations such as reshaping, resizing, changing intensity, and changing colors etc. Further, there are two types of classifications. Supervised and unsupervised. The process of image classification involves 3 basic steps. The first is data collection and pre-processing of data. Second is training the neural network on that data and then evaluating the model on testing data. In the next section, we will review kinds of literature based on convolution neural networks.

India is one of the most culturally diverse nations of the world is home to hundreds of cultures, languages, and traditions. Distinct cultural groups of India are North Indian, East Indians, and South Indians.

It is a rationale to study the culture difference between these three regions of India. Faces are one of the key features for the identification of a person belonging to different regions in India. People from the East are mostly having flatter faces with small eyes and pale skin complexion whereas North Indians have more of a beard and mustaches on the face also Sikh males have Turbans on their head. South Indians have dark complexion and dots on their foreheads. These features add up for features that will help us in the classification of faces. Convolution neural networks are strong in extracting these features from the images hence are preferred for classification of Images.

There has been a great development in hardware used for deep learning which includes GPUs and TPUs which drastically reduces the processing time for deep learning. The images of Indian faces were collected keeping in mind the distinct features that will differentiate faces from one region to another.

West Indians were not considered in this classification as there is no distinct feature found in people of west India. Hence that category was not considered.

This paper is divided into different sections. Section I deals with Literature review, Section II deals with Methodology. In section III we have explained the data preparation of the dataset. Section IV deals with the Network Architecture of the study and the last section presents the concluding part of the paper.

2. LITERATURE REVIEW
Muthukrishnan Ramprasath and M.Vijay Anand (2018) explain how computer vision is an upcoming field in Artificial Intelligence Each Image is built with a multi-dimensional array of pixels and each pixel represents different values. A grayscale image is having only 1 channel hence only 1 value for each pixel whereas in a colored image there are 3 channels (R, G, B) which gives 3 value for each pixel. Hyper Spectral Images is a process used for extracting features from an image using CNN. It used different pooling layers to extract the most important feature from the image. It is inspired by Human Brain.

MNIST digits dataset is used to perform CNN. They have created a 4 layer CNN architecture in which 2 are convolution layer,1 is MaxPooling Layer and last layer i.e the 4th layer is a Dense layer with sigmoid activation function for binary classification. In their model, the Relu activation function is used in the first three layers which gives output value when it has a positive input value and returns 0 when the value is negative.

Softmax activation was used in dense layer for predictions which returns an array of probability for the particular observation belonging to that class. The model run with a batch size of 128 epochs 5 and no. of classes 10 with input

dimensions of 28*28 gives an accuracy of 98% which is very high as compared to any other classification model.

Karan Chauhan and Shrwan Ram (2018) have worked on cat and dog classifiers with different combinations of activation functions and have compared the results between Tensorflow and Keras. Four different activation functions are used namely softmax, sigmoid, Tanh and ReLu. According to them, the best result was obtained with ReLu in hidden layers and the sigmoid in the classification layer as it's a binary classification problem. The accuracy with this combination was highest i.e 90.54%.

Tibor, Patrik, Richard, Miroslav, peter (2017) have worked on the Animal Recognition system based on CNN. Currently, there is no robust model for animal detection. They have created a dataset of 500 images (5 classes having 100 images each). They have applied different image recognition methods such as Principle component analysis (PCA), Linear Discriminant Analysis (LDA), Local Binary Patterns Histograms (LBPH) and support vector machines (SVM) and Finally CNN. They have compared all the algorithms. In their analysis, they have firstly used PCA is used on this dataset first. PCA mainly helps to reduce the number of features. These new features are called principal components. In CNN colored image of 32x32 was fed to the model with 2 convolution layers having ReLu activation function, 2 Maxpooling layers connected to Dense layers of 256 neurons and the connection to 5 neuron layer having softmax activation function. 6 different training splits were performed and all splits were fed to all models for comparison. Split A (90-10), B(80,20), C(70,30), D(60,40), E(50,50), F(40,60). For split, A CNN performed best with 98% accuracy followed by LBPH having 88% accuracy. In B split also CNN had the highest accuracy with 92%, followed by LBPM 84%. In all the following splits CNN performed the best with accuracy decreased in every split, with a decrease in size of training data and an increase in the size of testing data the accuracy decreased for all models.

Yiyu Hong and Jongweon Kim (2017) have worked on art identification using Convolution neural networks. Each painting is distorted by various operations because of lens, light, scaling and rotation. The input layer undergoes different pre-processing techniques like PCA or normalization which makes the computation simpler for the neural network. The processed data is passed to the convolution neural network according to which the weights matrix is calculated. Dataset of 100 images(which images) was downloaded from google and each image was distorted by different techniques which added 300 more images for each downloaded image making the dataset of 30k images out of which 25k was used for training and 5k were used for testing each having size of 256x256. Pixels of these images were then scaled down between -1 and 1. They have randomly initialized Weights and Bias are randomly initialized. ReLu is the activation function used for convolution layers with max-pooling and Adam optimizer to minimize the loss. Various different models with different

hyperparameters were tried and tested for comparison between different models. In the end, the results of CNN is compared with the SIFT algorithm. Their results show that CNN is outperforming the SIFT algorithm by 13.6%.

Sonali Sawardekar and Prof. Sowmiya Raksha Naik (2018) have worked on Facial Recognition using efficient LBP and CNN. There are six output categories. CFS+ANN was used which gave an accuracy of 91.2%. A database of 2105 images

was used. Accuracy with KNN was noted down to be 77.27%. CNN had an accuracy of 80% where are the accuracy of LBP+CNN was noted as 90%.

Srdjan Sladojevic, Marko Arsenovic, Andras Anderla, Dubravko Culibrk, and Darko Stefanovic (2016) have worked on the detection of plant diseases using CNN. People have used the detection of different diseases for improving crop health and the use of pesticides. Different approaches commonly used for plant disease diagnostics, for instance, double-stranded ribonucleic acid (RNA) analysis, nucleic acid probes, and microscopy. People have worked on different methods for the detection of disease in plants such as Shape detection, Colour detection, Texture classification. All the dataset was collected from the internet along with various diseases and plant names. One class was created with all diseased leaves images and the other was created with healthy leaves. Due to a lack of dataset images, Image augmentation was used to increase the total size of the dataset. Afterimage augmentation the total size of the dataset became 33,469 images with 30880 in training data and 2589 in testing data. The main reason for Image Augmentation was to increase the dataset and would help in reducing overfitting with all the variations. Finally with all the augmentation and different diseases as different classes we now have 15 classes in output variable having 14 diseases and 1 with healthy leaf images. Transfer learning was used with the Caffe model to build CNN. The last layers were removed from the original model and a new custom layer was added instead with a custom softmax activation function of 15 neurons. After fine-tuning the parameters the accuracy achieved was about 96.3%.

Jiantao Liu, Xiaoxiang Yang, Chen Wang, Yi Tao (2018) proposed a CNN for the identification of different types of dolphin species based on echolocations. A convolution network was built with four 1-D Convolution layers and two fully connected layers. Every convolution layer was followed by batch normalization layers, max-pooling layers, and ReLu as an activation function. Dropout layers were set between the layers as well to prevent overfitting. Convolution layer extracts high-level features from the images. As it is a signal processing task we use a 1-D convolution layer. CNN was built with Pytorch and Adam optimizer and a batch size of 64. Hyperparameter tuning was performed with kernel size from the range of 1 – 13 size of the kernel. The best test accuracy was achieved of 99.75 with a kernel size of 13,13. With kernel size of 1 or 3 convergence was slow and sometimes also let to overfitting.

Rajesh Kumar G A, Ravi Kant Kumar, Goutam Sanyal (2017) worked on facial emotion analysis using CNN. There are seven basic emotions of humans: Anger, Disgust, Fear, Happiness, Sadness, Surprise and neutral. The Dataset is provided by FER which holds 39K images each of size 48x48. In their analysis, the first step is to pre-process the images and crop only the face images from the whole image in order to reduce the noise from the image. These images are then resized to 48x48. Haar Cascade is used to detect faces from the Images. With these pre-processing steps, the model dataset was created. 28K observations were used in training and 11K were used in the testing of the data. 3 Convolution layers were used each followed by a max-pooling layer which was then connected to a fully connected dense layer having 3072 neurons which were further connected with dense layers having 7 neurons with a softmax activation function. The accuracy achieved was only 48%.

3. METHODOLOGY

Convolution Neural Networks are made up of neurons with learnable weights and bias. Each neuron gets an input which is then processed by passing it through an activation function and giving an output. The whole network works to reduce the error of the network.

1) Convolution Layer – In the convolution layer, we set a basic size of the kernel which hovers over the image and takes a dot product of each kernel with its underlining image. This value that we get from the dot product is replaced with the middle value of the image pixels covered under that kernel. We prefer taking an odd value of kernel so that we can get a middle pixel. These kernels are initialized randomly and values of these kernels learned over as the training process continues. The number of features extracted in the layer depend on the number of neurons present in a particular layer. The extracted image which contains the features of a particular image is called a feature map. The objective of the convolution layer is to extract high-level features from the images and then feed these features to the next layer.

2) Max Pooling – Pooling operation is often referred to as the downsampling technique. With the pooling operation, we shrink the features and the complexity of the model also reduces. It output the maximum value inside the kernel.

3) ReLu (Rectified Linear Unit) – It is an activation function that is used for mapping of non- linearity into the data. ReLu can be defined as a function that gives output as $max(0,x)$ where x is the input i.e the values less than 0 are output as 0 and values greater than 0 are x itself.

4) Batch Normalization - Batch Normalization layer normalizes the data from the previous layers on each batch. It reduces the impact of earlier layers by keeping the mean and variance fixed making the layers much more independent of each other and increasing the convergence speed.

5) Dropout - The dropout layer randomly turns off some neurons. The main idea mind is to make all the neurons independent of each other and help in reducing overfitting.

Overfitting can be easily caused by Neural networks because of model complexity, which decreases the validation accuracy of the model. Overfitting model starts learning irrelevant features from the data (noise), leading to less generalization of the model hence reducing the validation/test accuracy.

Overfitting can be avoided by using Batch Normalization, Dropout or using pre-trained models.

4. DATASET PREPARATION

The dataset for this task was collected from google images through web scrapping. This dataset also comprises of Indian celebrity images and also famous personalities of India. The dataset is divided into 3 classes that is North Indian, East Indian, and South Indian. The training dataset is having around 1600 images per class and the testing dataset is having around 300 images per class. The images were cropped to extract only the faces using Python, Opencv and face haar cascade.

5. NETWORK ARCHITECTURE

Convolution neural network architecture was used for Indian face classification which will help in extracting different and unique features from the faces and eventually help in classification based on those extracted features.

The basic structure is having five blocks of layers.

The first block is having a Convolution Layer with 32 neurons and a kernel size of (3,3) followed by a Max Pooling Layer and a Dropout Layer. Here the activation function used is ReLu. The second block is exactly the same as the first block. The third block is also having a convolution layer with 64 neurons and kernel size of (3,3) followed by the Max Pooling layer and a Dropout layer. The fourth Layer is exactly similar to the third layer. Then we add a flatten layer which is the fifth layer to make the exacted features into a 1-D array. This array is then passed to the dense fully connected Layer.

The fully connected layers gather all the information from the previous convolution layers and then serve high-level reasoning for the classification of that particular input image.

Since the dense layers contain a lot of parameters hence overfitting can easily happen. Adding a dropout layer is an option that can prevent overfitting.

Image Augmentation is also done with the help of Keras ImageDataGenerator which helps in the better generalization of the model as it brings more variations in our dataset. Finally, we compile our model with Adam optimizer and categorical_crossentropy loss function. This loss function calculates how far the actual class is from the predicted class in order to calculate the error. This helps in the tuning of the weights through backpropagation.

The input image is of size 64x64 and is colored hence is having 3 channels (RGB). After the first block, the input image for the second block is 32x32. Input for the third, fourth and fifth layers is 16x16, 8x8 and 4x4 respectively. This reduction in dimension is because of the Max Pooling layer.
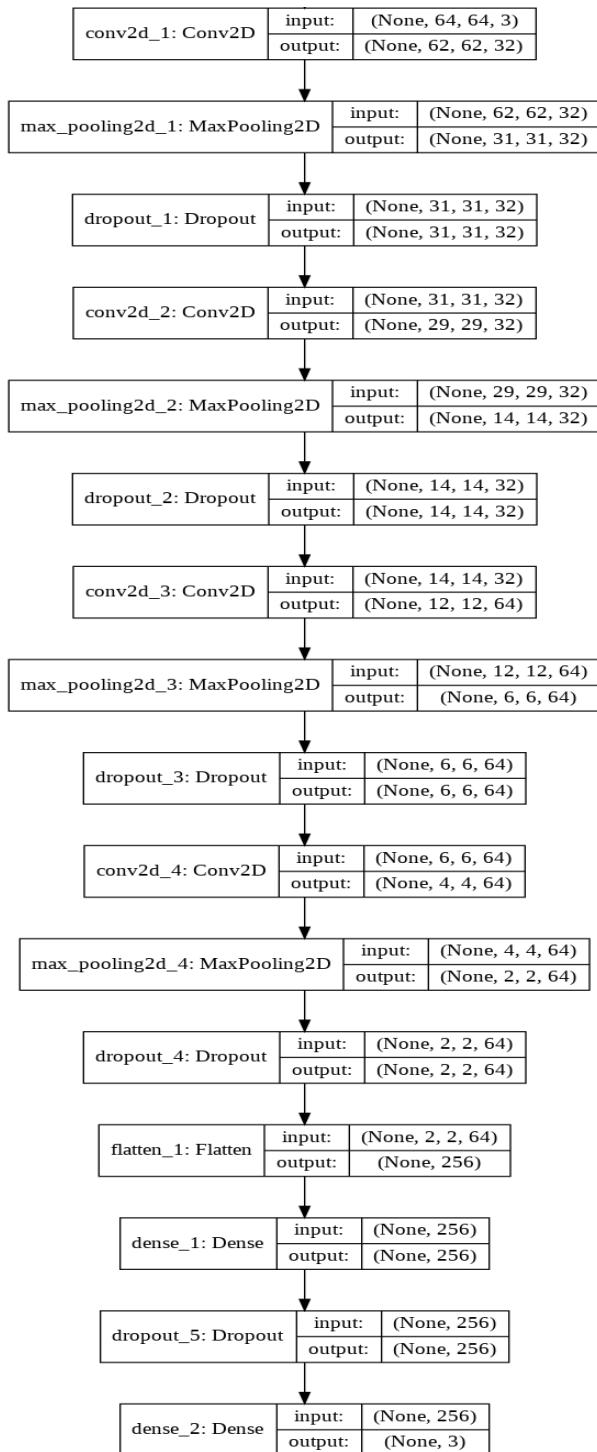
Fig I: Network Architecture

6. IMPERICAL ANALYSIS

The results presented in this section are related to training with the whole database containing both original and augmented images. As it is known that convolutional networks are able to learn features well when trained on larger datasets. The network was trained with the above-mentioned architecture and varied different hyperparameters in order to achieve high accuracy.

In the below graph we can see that the accuracy for training and testing starts to deviate as the number of epochs increases

leading to overfitting. Hence we try to decrease the number of epochs that are kept down to 10 only.

Accuracy of 81% is achieved on the validation dataset with 10 epochs. The accuracy for south Indian class is not that high as there is no special feature that was extracted from the faces of South Indians as compared to other classes.

Taking into account the fact we have done this research on our own image dataset there no one has used this up till now hence we don't have any model to make our comparison with. In addition, since no one has used deep learning to differentiate between faces it is impossible to compare it with other examples currently present in the market.

Nevertheless, as a result of extensive research going on in the field of convolution neural networks, Deep Learning techniques have shown better results in pattern recognition, image detection, and object recognition.
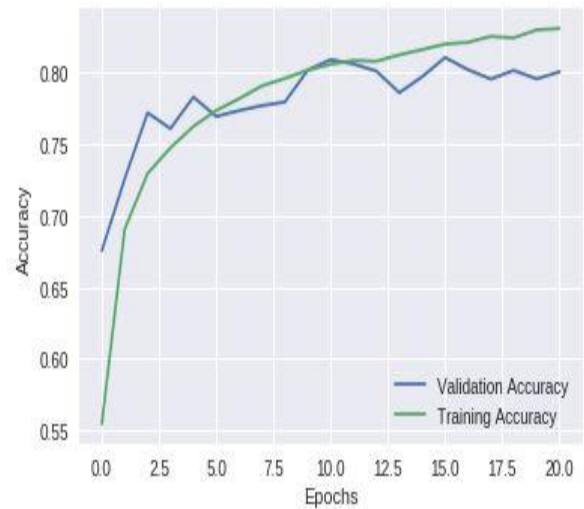
7. CONCLUSION

We have successfully completed the model with face classification of Indian Faces into three categories i.e North Indian, East Indian, and South Indian. We have maximum misclassification in South Indian Class as the model is not able to capture any distinctive feature for South Indian Faces which makes it independent from the other two classes. The final Accuracy of the model achieved is 81%.

8. REFERENCES

[1] Yiyu Hong and Jongweon Kim Art Painting Identification using Convolutional Neural Network International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 4 (2017)

[2] Sonali Sawardekar, Prof. Sowmiya Raksha Naik Facial Expression Recognition using Efficient LBP and CNN International Research Journal of Engineering and Technology (IRJET) Volume: 05 Issue: 06 | June 2018

[3] Srdjan Sladojevic, Marko Arsenovic, Andras Anderla, Dubravko Culibrk, and Darko Stefanovic, Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification, Hindawi Publishing Corporation Computational Intelligence and Neuroscience Volume 2016

[4] Jiantao Liu, Xiaoxiang Yang, Chen Wang, Yi Tao, A convolution neural network for dolphin species identification using echolocation clicks signal, IEEE 2018

[5] Rajesh Kumar G A, Ravi Kant Kumar, Goutam Sanyal, Facial Emotion Analysis using Deep Convolution Neural Network, International Conference on Signal Processing and Communication (ICSPC'17) – 28th & 29th July 2017

[6] Muthukrishnan Ramprasath, M. Vijay Anand, Shanmugasundaram Hariharan Image Classification using Convolution Neural Network International Journal of Pure and Applied Mathematics, Volume 119 No. 17 2018.

[7] Karan Chauhan, Shrwan Ram Image Classification with Deep Learning and Comparision between different convolution neural network structures using Tensorflow and Keras, International Journal of Advance Engineering and Research Development Volume 5, Issue 02, February -2018

[8] Tibor Tmovszky, Patrik Kamencay, Richard Orjesek, Miroslav Benco, Peter Sykora Animal Recognition system based on convolutional neural network digital image processing and computer graphics volume: 15 Issue: 3 September 2017