**TELUGU CHARACTER RECOGNITION**

**Final Report**

**In fulfillment of the requirements for the**

**CS 342 ARTIFICIAL NEURAL NETWORKS**

**At NIIT University**

**Under: Prof PRASHANT SRIVASTAVA**

****

**Submitted by**

|  |  |  |
| --- | --- | --- |
| **S.No** | **NAME** | **Enrol No.** |
| **1** | **Bolam Sreekar Reddy** | **020** |
| **2** | **Tavva GNRSN Prudhvith** | **142** |
| **3** | **Pisupati VNSSK Chaitanya** | **085** |

**1 INTRODUCTION**

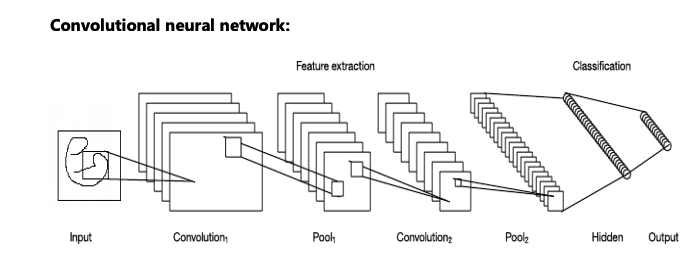
Neural Networks are recently being used in various kinds of pattern recognition. Handwritings of different persons are different therefore, it is very difficult to recognize the handwritten characters. Handwritten Character recognition is an area of pattern recognition that has become the subject of research during the last some decades. Neural network is playing an important role in handwritten character recognition. Handwriting recognition is the ability of a computer to receive and interpret intelligible handwritten input from sources such as paper documents, photographs, touch screens and other devices. It can be online or offline. In this context, online recognition involves conversion of digital pen-tip movements into a list of coordinates, used as input for the classification system whereas offline recognition uses images of characters as input.

**2 MOTIVATION**

Although a lot of work has been reported for handwriting recognition in English and Asian languages such as Japanese, Chinese etc., and very few attempts on Indian languages like Hindi, Tamil, Telugu, Kannada etc. In this paper, I am developing handwritten character recognition algorithm for Telugu [South Indian language] with high recognition accuracy and minimum training and classification time. So, with this trend in mind we tried to find an application of the course CS-342 Artificial Neural Networks and incorporate it into our project. Some of the earlier works apply shallow learning with hand-designed features on both online and offline datasets. Examples of hand-designed features include pixel densities over regions of image, character curvature, dimensions, and number of horizontal and vertical lines.

**3. WORK DONE**

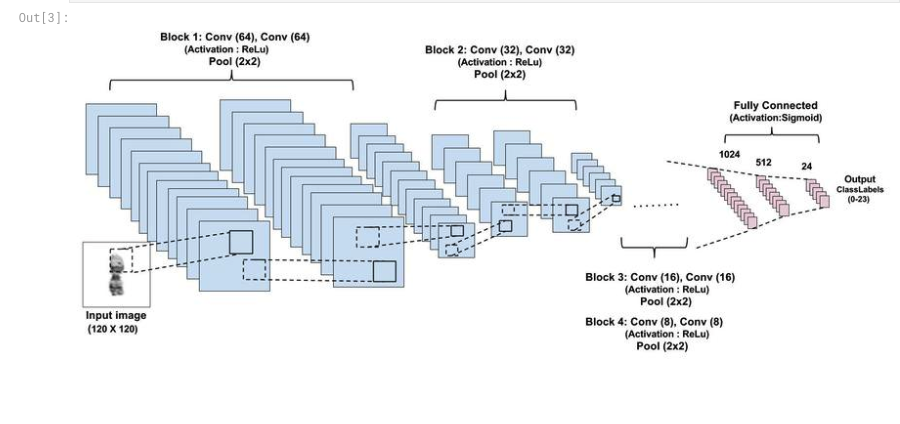
**Algorithm**



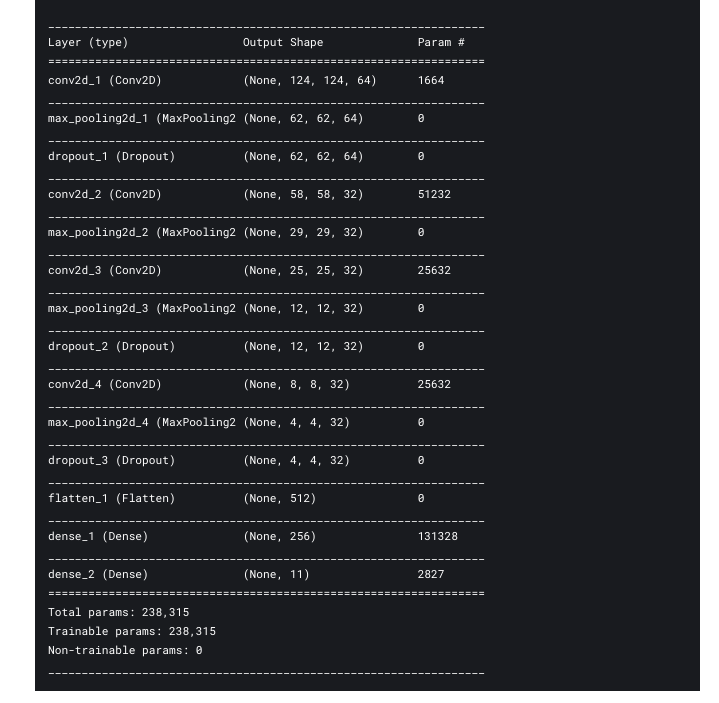
The diagram shows an overall architecture of CNN that consists with two main parts: feature extraction and classification. In the feature extraction layers, each layer of the network receives the output from its immediate previous layer as its input, and passes the current output as input to the next layer. The CNN architecture is composed with the combination of three types of layers: convolution, max-pooling, and classification. Convolutional layer and max-pooling layer are two types of layers in the low and middle-level of the network. The even numbered layers work for convolution and odd numbered layers work for max-pooling operation. The output nodes of the convolution and maxpooling layers are grouped in to a 2D plane which is called feature mapping. Each plane of the layer usually derived with the combination of one or more planes of the previous layers. The node of the plane is connected to a small region of each connected planes of the previous layer. Each node of the convolution layer extracts features from the input images by convolution operation on the input nodes. The max-pooling layer abstracts features through average or propagating operation on the input nodes.

**SCREENSHOTS OF THE WORK DONE**

### **Defining the model and implementation**

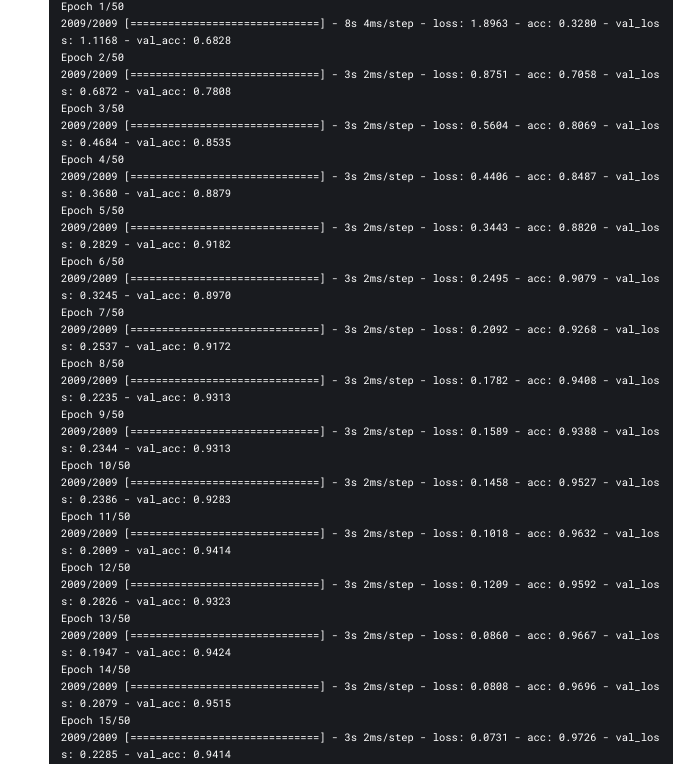
****

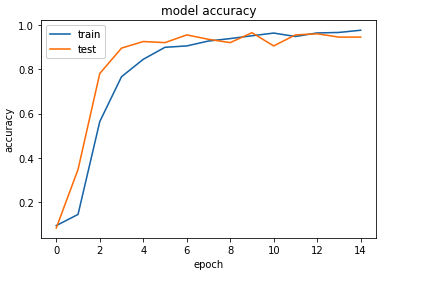
#### **Converting 11 classes into one-hot encoding**

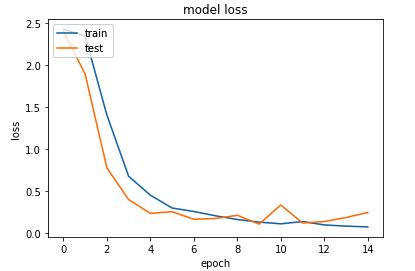
****

### **Training**

* Splitting the data for training and validation to improve accuracy.

****

****

****

I got test accuracy of 91% and training accuracy of 97% on Telugu character dataset with 15 epochs, if we increase number of epochs, then the accuracy will increase further.

**4 APPLICATION OF THE PROJECT**

**4.1. Problem Statement**

Handwriting recognition has been one of the active and challenging research areas in the field of image processing and pattern recognition. Since 1929, number of character recognition systems have been developed and are used for even commercial purpose also. Several applications including mail sorting, bank cheques processing, reading aid for blind, document reading and postal address recognition require offline handwriting systems. Working in Postal service need us to decode and deliver something like 30 million handwritten envelopes every single day. The challenges are to do mail-sorting that ensure all those millions of letters reach their destinations.

**4.2. Constituents**

Character recognition complexity varies among different languages due to distinct shapes, strokes and number of characters. Telugu, a South Indian language that ranks third by the number of native speakers in India. Fifteenth in the Ethnologic list of most-spoken languages worldwide and is the most widely spoken Dravidian language in the world. About 800 million people use Telugu as their speaking and writing purpose. Telugu script has 18 vowels and 36 consonants, of which 13 vowels and 35 consonants are in common usage. Of all the Indic scripts, the Telugu script has the largest number of vowels and consonants. Handwritten character recognition can be online or offline. In this context, online recognition involves conversion of digital pen-tip movements into a list of coordinates, used as input for the classification system whereas offline recognition uses images of characters as input. Some of the earlier works apply shallow learning with hand-designed features on both online and offline datasets. Examples of hand-designed features include pixel densities over regions of image, character curvature, dimensions, and number of horizontal and vertical lines.

**5 Conclusion**

It is well known that the deep convolutional neural networks are very good at classifying image data. There were many experiments done on handwritten character recognition using convolutional neural networks for English alphabets, numbers, Chinese, Arabic and some of the Indian languages like Hindi, Devanagari script etc., But there is very less contribution on Telugu language character recognition. Due to very less contribution, the data available on internet is not so good. I didn’t notice that initially and faced many problems with repeated data. I found that the problem was in data, then I deleted the repeated ones and made the data clear. Then we found difficult in tuning the algorithm. I tried different optimizers with different learning rates and tried a model changing number of layers and number of filters and filter size. Finally, we ended up with a model that gives 91% accuracy.

**6 FUTURE DEVELOPMENT POSSIBILITIES**

This algorithm used only single characters of Telugu language. But there are many characters with ‘vattu’ and ‘gunintham’. As there are very less contributions on Telugu language characters, I couldn’t find the dataset consisting of all these characters. But the characters used in this project are like vowels and consonants in English. Due to time constraints I didn’t generated dataset with all those characters. In future, I want to extend this algorithm from character recognition to text recognition by creating my own dataset.