Vellore Institute of technology

Sign Language Character Recognition

Soft Computing (SWE1011)

Guided by: Jagadeesh Kannan R
Slot: D1+TD1

Submitted by:
Artham Abhinay
16MIS1027
Shantanu Gupta
16MIS1109

Abstract

Our project is an effort towards studying the challenges in classification of characters in American Sign Language(ISL). A lot of research has been done in the corresponding field of American Sign Language(ASL).. Lack of standard datasets, occluded features and variation in the language with locality have been the major barriers which has led to little research being done in ISL&ASL. Our project aims at extending a step forward in this field by using the dataset of the deaf, and then use various feature extraction techniques to extract useful information which is then input into various supervised learning techniques. Currently, we have reported categorical crossentropy results for the different approaches, and the difference from the previous work done can be attributed to the fact that in our validation, the validation set correspond to images of a person different from the persons in the training set.

Contents

1 **Introduction** 2

2 **Motivation** 2

3 **Challenges** 3

4 **Dataset** 3

5 **Methodology** 4

6 **Results** 6

7 **Observation & Conclusions** 7

8 **Future Work** 7

9 **Acknowledgement** 7

1 Introduction

According to the World Federation of the Deaf (WFD), there are 70 million deaf people using sign language as their first language; many hearing people also use sign language as their first or second language. In the U.S. alone, there are one half to two million people using American Sign Language (ASL) in the 1990s . There is a huge barrier between the Deaf community and people that do not understand or know little about sign language. A sign language recognition system would help break this barrier, we aim to solve this problem using state of the art computer vision and Convolutional Neural Network.

Figure 1: American Sign Language

2 Motivation

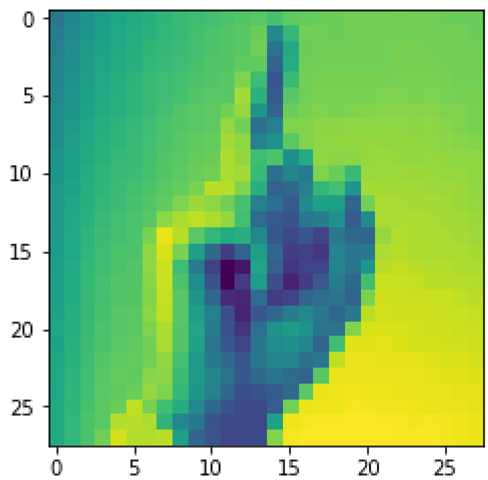
Communication is one of the basic requirement for survival in society. Deaf and dumb people communicate among themselves using sign language but normal people find it difficult to understand their language. Extensive work has been done on American sign language recognition but Indian sign language differs significantly from American sign language. ASL uses single hand for communicating. In addition to this, lack of datasets along with variance in sign language with locality has resulted in restrained efforts in ASL gesture detection. Our project aims at taking the basic step in bridging the communication gap between normal people and deaf and dumb people using Indian sign language. Effective extension of this project to words and common expressions may not only make the deaf and dumb people communicate faster and easier with outer world, but also provide a boost in developing autonomous systems for understanding and aiding them.

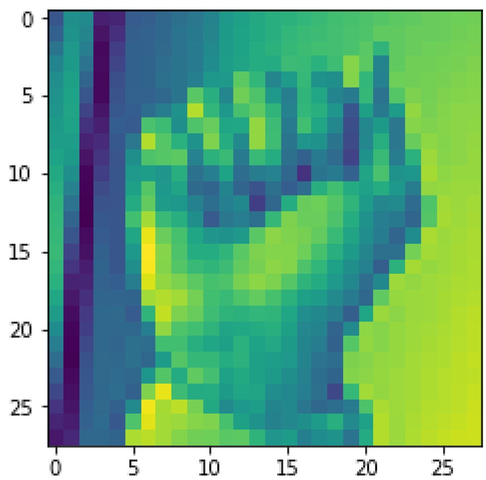
3 Challenges

The American sign language is widely used sign language for the deaf all around the world but there is a deficiency of proper datasets for training and tested the sign language. The thresholding of the real time captured signs was a difficult task to accomplish.

4 Dataset

Previous Datasets : We couldn't find any reliable datasets on the internet as most of the videos of the sign language we found on the internet were of people who were demonstrating what sign language looked like and not those who actually spoke the language. Also the dataset found at was taken from one of the team members and so was not reliable. We finally found a dataset from Kaggle which was the standard MNIST dataset used for American sign language recognition.





5 Methodology

Having collected the dataset, we divided our approach to tackle the classification problem using Convolutional Neural Network(CNN).

We divided our MNIST dataset acquired from Kaggle into 70:30 training and testing dataset ratio. The dataset contains a total 27,455 similar with a header row of label ,pixel 1,pixel 2 ,…..pixel784 which represent a single 28x28 image with grayscale values between 0-255.

- The first stage we use the traditional CNN model i.e. Convolution layer + Pooling layer and ReLU is the activation function used in this CNN model. The CNN network model image is given below.
- After the CNN, we passed our featured images into the dense layer with 128 neurons, with a dropout value of 0.20
- The last layer is also a dense layer with classes ranging from 0-24, where 0 corresponds to ‘A’ and 24 is ‘Y’. For this layer we use the Softmax activation function.

Softmax function:

$$\sigma(x_j) = \frac{e^{x_j}}{\sum_i e^{x_i}}$$

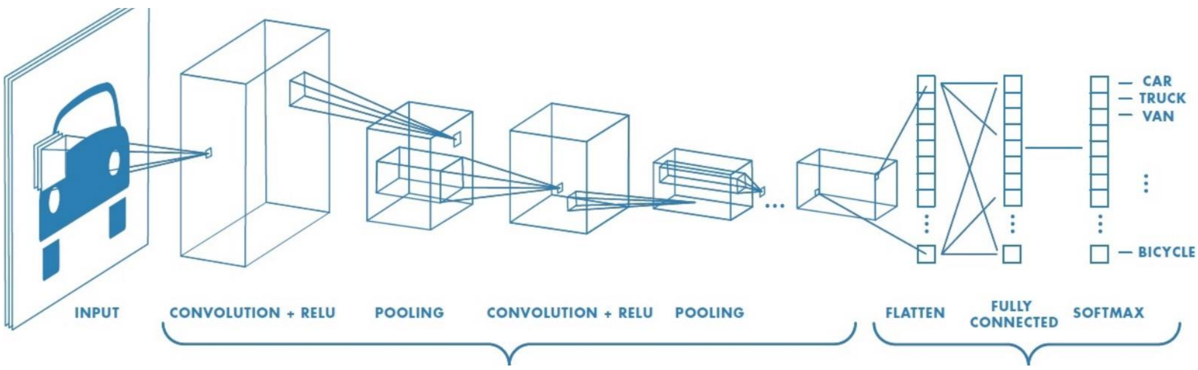
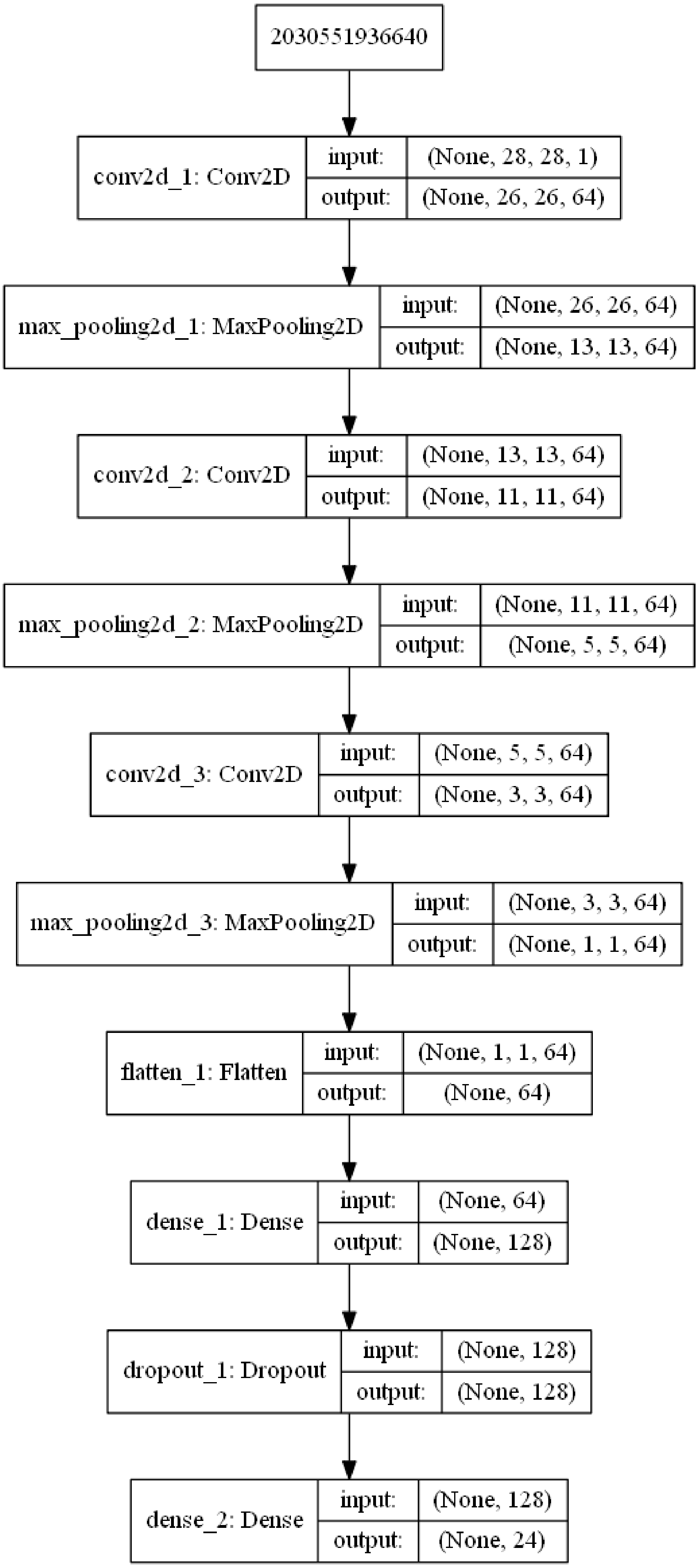


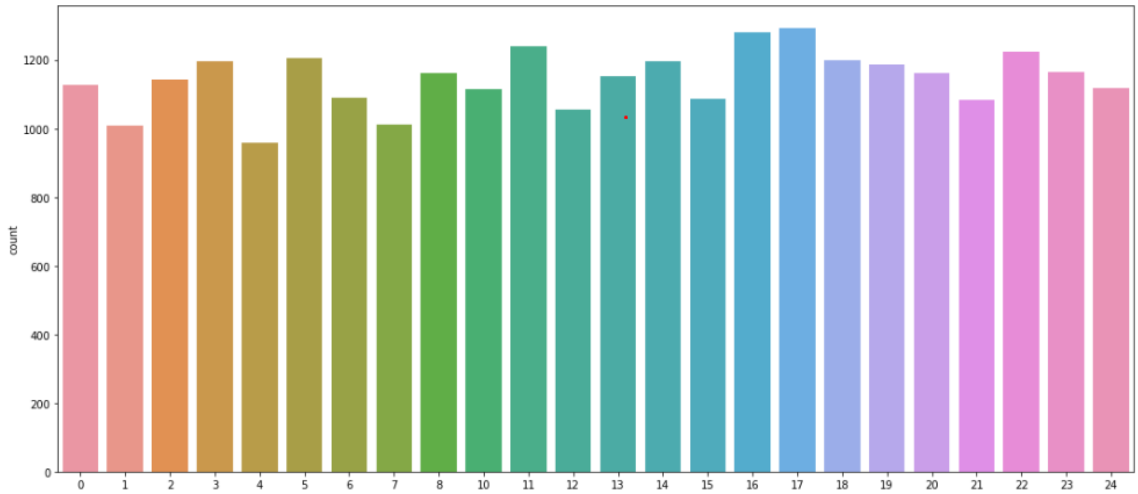
Figure: A CNN architecture.



the CNN model.

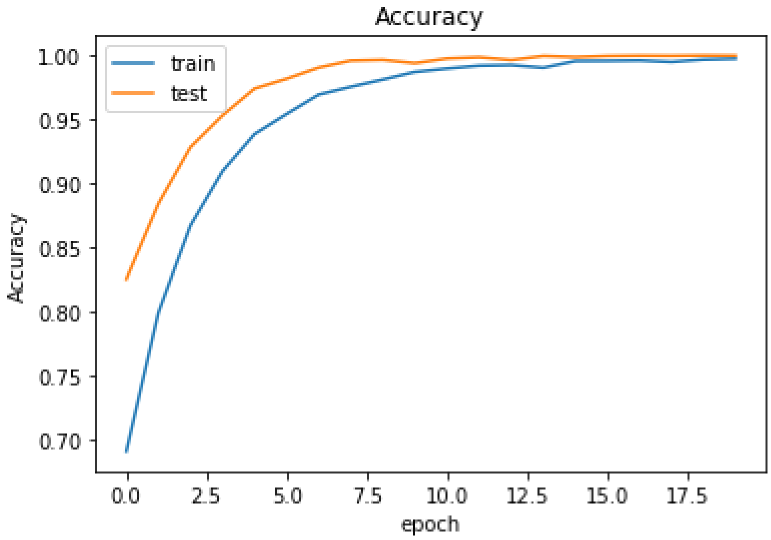
This model performed with 81% accuracy on the training and tested dataset which we divided in the first stage of the project. An alphabet is first classified as one handed, and then depending on the classification, is put into the corresponding model and given a label.

6 Results



This is the distribution of our data, each column represents the number of images existing for an individual alphabet

We received an accuracy of 99% while training…



7 Observations

- There are no signs for the alphabets ‘J’ and ‘Z’, since we should use motion to represent them

- A better dataset (with more variations, so that the model is more robust to the sign variations) would definitely improve the results

8 Conclusions

Although there was some work done priory on ASL, they generated the training and test dataset from the same person which lead to higher accuracies. We did categorical cross entropy by using the images for training and testing the model which gave us an accuracy of 89%. Some images in the dataset were taken in bad illumination which gave noisy images while image segmentation. A better dataset would have helped in finding more precise features and would have resulted in higher accuracies.

9 Future Work

The ultimate aim of the project is to bridge the gap in access to next generation Human Computer Interface.
Currently the project uses a dataset and perform supervised learning to recognize different sign alphabets in American Sign Language. After building a neural network and training the model, we can recognize sign alphabets of real time captured image.
Future scope of the project is to execute unsupervised learning of sign models i.e. to learn sentences and sign from a series of live images in real time.

10 Acknowledgement

We thank Prof. Jagadeesh Kannan R of School of Computer Science Engineering for his valuable support throughout the project guiding us from time to time and looking into the project when it was needed. We also thank python community for their user friendly libraries.

References

[1] https://pythonprogramming.net/introduction-deep-learning-pythontensorflow-keras/

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

[3] https://github.com/loicmarie/sign-language-alphabet-recognizer