K. K. Wagh Institute of Engineering Education and Research, Nashik
Department of Computer Engineering

Sign-to-Speech Conversion Using CNN

PID : 27

Group Members:

1) Shubham Zope (Group Leader)
2) Abhijit Watpade
3) Sanket Sonar
4) Sameer Rathod

Project Guide: **Prof. N. M. Shahane**

PROBLEM STATEMENT

- ➔ To create a computer software and train a model using CNN which takes an image of hand gesture of American Sign Language and shows the output of the particular sign language in text format converts it into audio format.

MOTIVATION

- ➔ Interaction Barrier for D&M Peoples.

- ➔ If there is   a common interface that converts the sign language to text the hand gestures can be easily understood by the other people.

- ➔ Develop User Friendly Human Computer Interface.

OBJECTIVE

- ➔ To create a completely functional product for the people who are not able to hear, so that ,they can get connected to world easily .

- ➔ To use and understand technologies like OpenCV, Matplotlib,Keras,Deep Learning,Python , Heroku host etc.

- ➔ To create a Web based project for detecting and understanding American sign language using Machine Learning concepts.

LITERATURE SURVEY

- **Data Acquisition :**
  - Use of sensory devices(Gloves)
  - Vision based approach(Computer Camera)

- **Data Pre-processing and Feature Extraction :**
  - Gaussian blur filter, Threshold filter, Gray filter
  - Colour segmentation

Gesture Classification :

- Hidden Markov Model(HMM)
  - It deals with dynamic aspects of gesture. Gestures extracted from video images by tracking skin-color blobs
- Naive Bayes classification(naive)
  - It deals with static aspects of gesture. Performs well with small datasets
- Convolution Neural Network(CNN)
  - It construct a skin model to extract the hand out of an image and then apply binary threshold to the whole image

METHODOLOGY

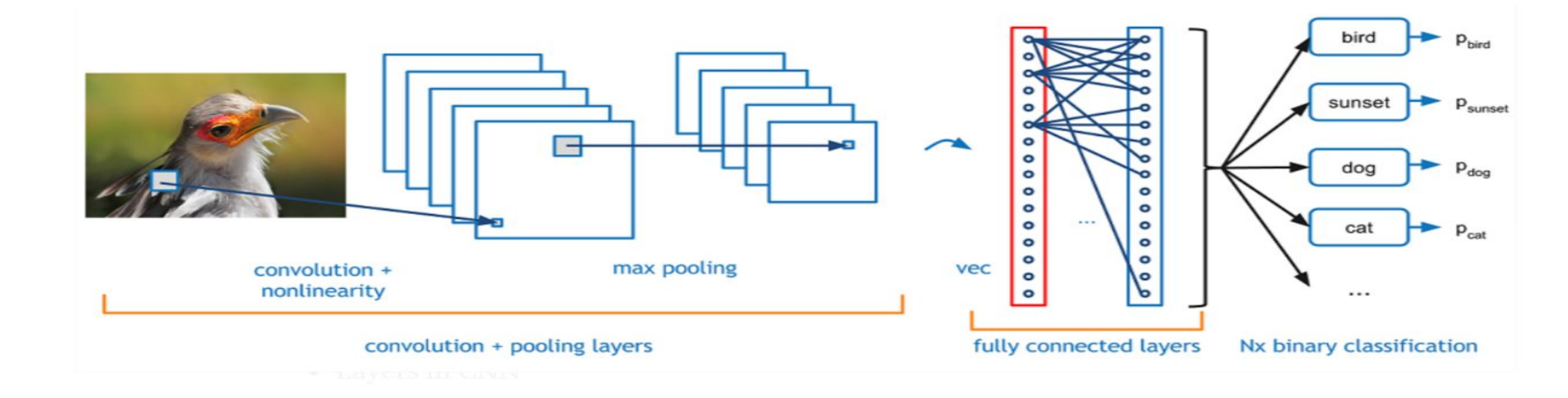
- The project can be made using vision based approach. All the signs are represented with bare hands and so it eliminates the problem of using any artificial devices for interaction.

- **DATA SET :-** We didn't found any database which is in a format which will help us to train our model quickly, so , we have decided to create our own database which will contain the black and white formated images , using OpenCv filters , rather than using the RGB dataset.

- **GESTURE CLASSIFICATION :-**
  - Apply gaussian blur filter and threshold to the frame taken with opencv to get the processed image after feature extraction.
  - This processed image is passed to the CNN model for prediction and if a letter is detected then the letter is printed and taken into consideration for forming the word.
  - Space between the words are considered using the blank screen.

ALGORITHMS

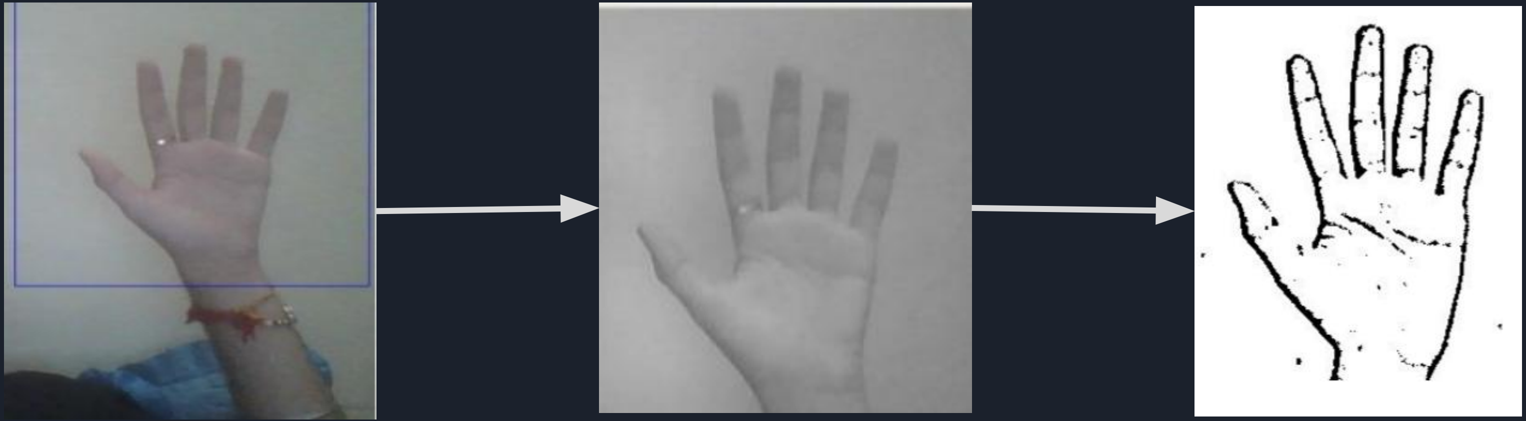
- ➔ **CNN (Convolutional Neural Network)** :- It is most preferably used algorithm for operations upon the images. In simple word what CNN does is, it extract the feature of image and convert it into lower dimension without losing its characteristics .So that working upon it would not take more computational power.
  - ◆ **Layers** :-
    - Convolutional Layers
    - Poling Layers
      - Max Pooling
    - Fully Connected Layer
    - Final Output Layer



➔ **Activation Function** :- We have decided to use **ReLU(Rectified Linear Unit)** as our activation function in each layer. It is a simple function stated as $f(x) = max(0,x)$ for each input pixel **.**The biggest advantage is it adds the nonlinearity to the model.

➔ **Optimizer** :- As our optimizer function we have decided to use Adam optimizer for updating the model in response to the output of the loss function. It is mostly used Gradient Descent optimizer because it use advantages of both stochastic gradient descent algorithms namely adaptive gradient algorithm(ADA GRAD) and root mean square propagation(RMSProp)

DATASET GENERATION FLOW

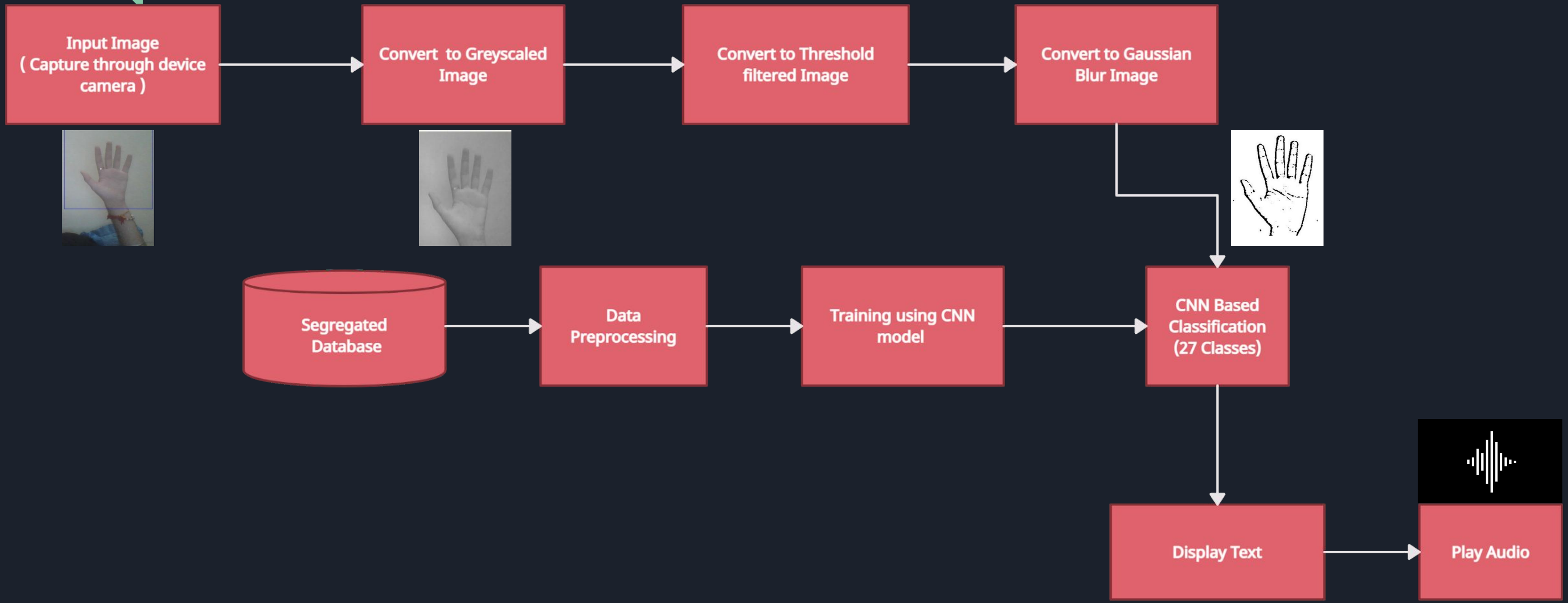


Capturing raw image

Grey scale image

Image post gaussian blur

BLOCK DIAGRAM



Input Image
(Capture through device camera)

Convert to Greyscaled Image

Convert to Threshold filtered Image

Convert to Gaussian Blur Image

Segregated Database

Data Preprocessing

Training using CNN model

CNN Based Classification (27 Classes)

Display Text

Play Audio

CONCLUSION

Thus, we have done with the research part as we have referred many research papers and now creating database and implementing project part will start.

THANK YOU!