Rajiv Gandhi Institute of Technology
Chola Nagar, Bengaluru, Karnataka - 560032

Sign Language Interpreter using Deep Learning

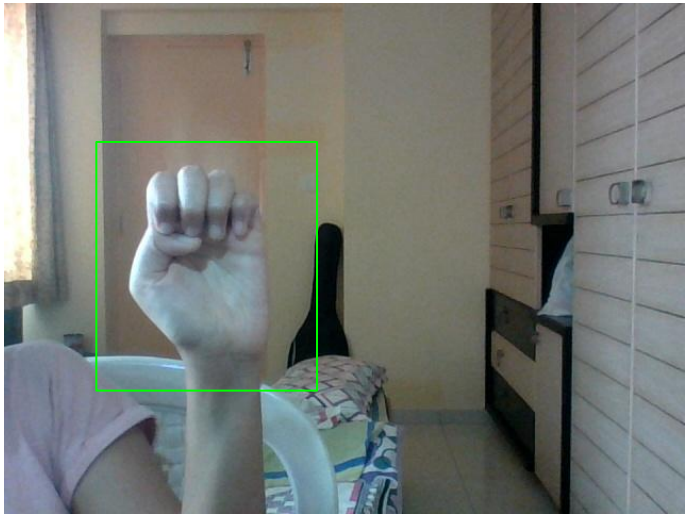
Under the guidance of : Asst. Prof. Pushplata Dubey

- Bahaduri Prachiti Jagdish - 1RG17CS009
- Mohammed Azim - 1RG17CS030
- Mohammed Touseef - 1RG17CS031
- Prajwal B Mani - 1RG17CS037

IMPLEMENTATION OF MODULES

Dataset Creation

Python Script to capture



Labelimage script

E.0a0a0ed8-ae48-11eb-ba12-00f48dddcbd4.xml

D: > Capture Dataset > capture datasets > E > E.0a0a0ed8-ae48-11eb-ba12-00f48dddcbd4.xml

```xml
<annotation>
    <folder>E</folder>
    <filename>E.0a0a0ed8-ae48-11eb-ba12-00f48dddcbd4.jpg</filename>
    <path>D:\Dataset\E\E.0a0a0ed8-ae48-11eb-ba12-00f48dddcbd4.jpg</path>
    <source>
        <database>Unknown</database>
    </source>
    <size>
        <width>640</width>
        <height>480</height>
        <depth>3</depth>
    </size>
    <segmented>0</segmented>
    <object>
        <name>E</name>
        <pose>Unspecified</pose>
        <truncated>0</truncated>
        <difficult>0</difficult>
        <bndbox>
            <xmin>108</xmin>
            <ymin>142</ymin>
            <xmax>266</xmax>
            <ymax>344</ymax>
        </bndbox>
    </object>
</annotation>
```

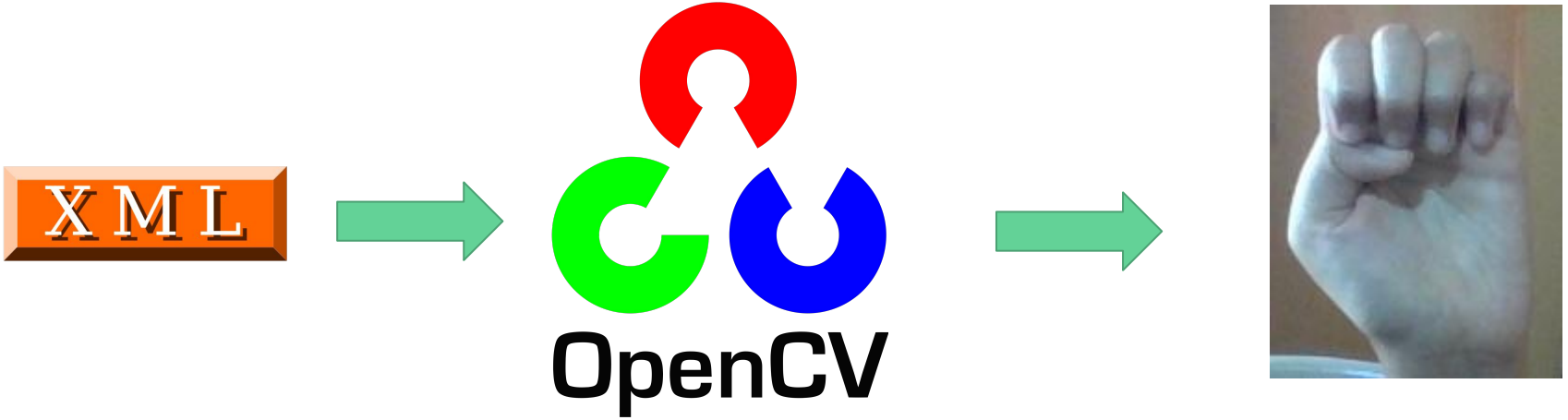

XML
OpenCV

Image PreProcessing:

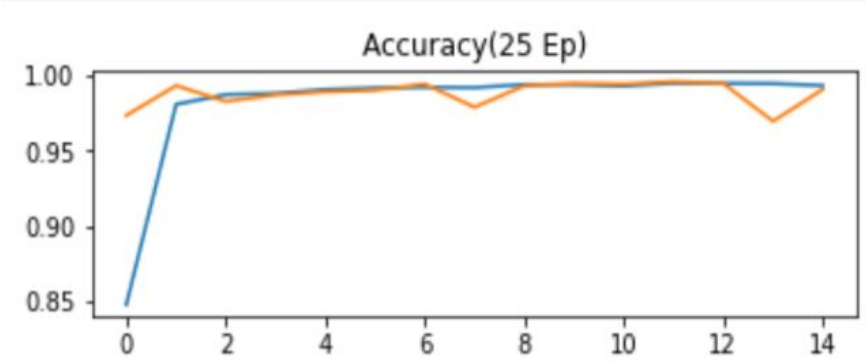
- We use computer algorithms to perform image processing on digital images so that our models has better image data to learn from.
- We have used OpenCV library for preprocessing
- For data augmentation we have used tensorflow and keras ImageDataGenerator
- Some of augmentation we applied are zoom, shift, flip and rescale

CNN Model :

- We have a VGG16 CNN architecture model that is trained on ImageNet weights and our dataset using Tensorflow and Keras
- EarlyStopping keras callback is used so the model doesn’t overfit for the data

```
In [7]: model.summary()
```

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
vgg16 (Functional)           (None, 2, 2, 512)         14714688
_________________________________________________________________
flatten (Flatten)            (None, 2048)              0
_________________________________________________________________
dense (Dense)                (None, 256)               524544
_________________________________________________________________
dense_1 (Dense)              (None, 40)                10280
=================================================================
Total params: 15,249,512
Trainable params: 15,249,512
Non-trainable params: 0
```



- We have used two activation functions that is “relu” and “softmax”
- Adam optimizer
- Categorical_crossentropy loss function
- Early stopping callback is monitored on the validation loss
- Accuracy of 99%

Other Models:

- Image Classification
  - VGG16 (Online Dataset)
  - VGG16 (Our own custom Dataset)
  - VGG19
  - ResNet
- Object Detection
  - SSD MobileNet FPN 320X320
  - SSD MobileNet FPN 640X640
  - ResNet FPN 1024X1024

Language correction model (Language model):

- The main role of this model is to check for spelling mistakes and give us suggestion
- This is built on Phoneme and Symspell where Phoneme uses Soundex algorithm and Symspell uses edit-distance concept.
- The model is deployed as an API(Application programming interface)

Text to Speech:

- We have used google text-to-speech(gtts) api to convert the text to audio which does its job perfectly.

Audio Preprocessing :

In order for the neural network to process the auditory information in the same manner as it does visual information so we do audio preprocessing.

We have used TarsosDSP module in android

Few preprocessing steps are noise cancellation,normalization and noise injection

Speech to Text:

For speech to text we have used google-cloud-speech and Speech Recognition libraries

Giphy API:

- Giphy has a category for sign language which is recorded by Robert with more than 1.3K gif’s for regular words and expressions.
- Using there Api to make our App simple, flexible and access there abundant resource
- At last all the gifs are stored in database and retrieved when it is needed and displayed to the user

Deploying the model:

- Tensorflow lite quantization converting our model from float 64 to float 32 and int 8 data type.
- Deployment is on android using tensorflow lite and tensorflow gpu support.
- Our model can even be deployed on raspberry pi

What and all signs can our app predict?

A	B	C	D	E
F	G	H	I	J
K	L	M	N	O
P	Q	R	S	T
U	V	W	X	Y
Z	Airplane	Bathroom	Bye	Hello
Help	Hurt	I Love You	More	No
Play	Something	Sorry	Thank you	Yes

Android App:

- The app was built on Android studio using java
- >API level 19 (KitKat)
- Few App features
  - Swipe Motion Gesture
  - No extra file storage
  - Easy install
  - Dark Mode

Our Objectives

1. Create a robust android application that aids in communication for deaf people.
2. Use deep learning with transfer learning techniques to build a neural network model that as to ability to learn pattern in and classify the images.
3. An error correction model for identifying the pattern mismatch and correcting it for the audio or text input format.
4. To build a model that has higher accuracy with less bias and overfitting problems.
5. To enhance the model to classify at least 20 classes or more .

DEMO OF THE PROJECT

CONCLUSION

A sign language is a natural mode of communication used by the deaf community. India has large population of speech and hearing impaired but a very small number of certified sign language interpreters are available. Research in hand gesture recognition has gained attention with advancement in the field of computer vision. A Sign Language Interpreter (SLI) decodes and understands the information conveyed by signs. SLI can be a major breakthrough in helping a common people to communicate with the deaf and can help in bridging this communication gap. A SLI can be designed based on video/image processing and deep learning techniques which requires a standard dataset, determination of an optimal feature set, and an appropriate classification technique.

THANK YOU