JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

INDIAN SIGN LANGUAGE RECOGNITION USING DEEP LEARNING

[1]Kokila B,[2]Raju K ,[3]Shibhu G
[1]Assistant Profesor,[2]Student,[3]Student
Department of Computer Science and Engineering,
Sri Ramakrishna Institute of Technology (An Autonomous Institution), Coimbatore, India.

***Abstract*—Human action recognition is detection of the action and movements caused by an individual which will be classified using structure, position, behaviour etc. Sign language is an action based natural language which comes under human action recognition. Sign language is the primary mode of communication between hearing and the vocally impaired population. The government of India has enacted the Rights of Persons with Disabilities Act 2016 (RPwD Act 2016) which recognises Indian Sign Language (ISL) as an important communication medium for communicating with hearing impaired people. This also insists on the need for sign language interpreters in all Government organisations and public sector undertakings in order to abide by RPwD Act 2016. In recent years, due to the widespread usage of various sensors, action recognition is becoming more popular in many fields such as person surveillance, human-robot interaction etc. Using action recognition, recognising Indian sign language gestures (alphabets, numerals and specific words) will be done in real time using deep learning LSTM model. In this project video will be collected in real time through a web camera which will be taken as an input for the model. Given input which is collected as key points from video frames will be passed through the model. The model predicts the output which is the translation of the given input video. User interface is created for hearing impaired individuals which will act as a sign language interpreter. This project is focusing to achieve the maximum accuracy with no loss.**

***Keywords*—Indian sign language, Deep learning, MediaPipe framework, RNN**

I. INTRODUCTION

One of the most precious gifts of nature to human beings is the ability to express himself by responding to the events occurring in his surroundings. Every normal human beings sees, listens and then reacts to the situations by speaking himself out. But there are some unfortunate ones who are deprived of this valuable gift. This creates a gap between the normal human beings and the deprived ones. Indian Sign Language is used in the deaf community all over India.

But ISL is not used in deaf schools to teach deaf children. Teacher training programs do not orient teachers towards teaching methods that use ISL. There is no teaching material that incorporates sign language. Parents of deaf children are not aware about sign language and its ability to remove communication barriers. ISL interpreters are an urgent requirement at institutes and places where communication between deaf and hearing people takes place but India has only less than 300 certified interpreters. Activity recognition aims to recognise the actions and goals of one or more agents from a series of observations on the agents' actions and the environmental conditions. Since the 1980s, this research field has captured the attention of several computer science communities due to its strength in providing personalised support for many different applications and its connection to many different fields of study such as medicine, human-computer interaction, or sociology. Due to its multifaceted nature, different fields may refer to activity recognition as plan recognition, goal recognition, intent recognition, behaviour recognition, location estimation and location-based services.

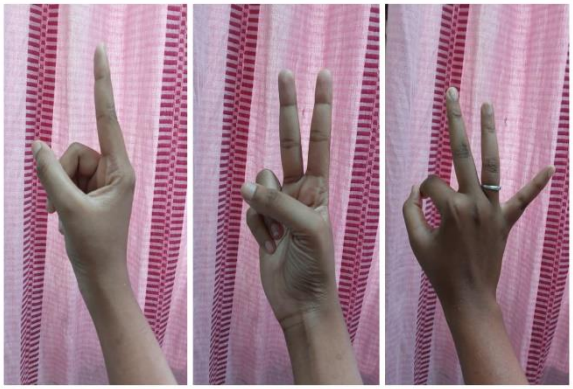

Fig. 1.1 Indian Sign Language Gestures of A(1),B(2),C(3)

The organisation of the paper is as follows. Section II contains related works which contains the deep learning model used. Section III contains modules in our project and our approach towards it. Section IV contains result followed by Section V with conclusions and future work.

II. RELATED WORKS

A. Recurrent Neural Network

Recurrent Neural Network(RNN) is a generalization of feedforward neural network that has an internal memory.RNN works effectively with great results. Comparing with other feedforward neural network RNN resolves a problem which is not considering the output of the previous input which makes it great to work with. RNN is the best for NLP, sequence detection etc. But RNN also has a limitation of not classifying the important and not so important memories, which results in storing less memory of the previous output because it contains all the not so important outputs. Sign Language recognition needs sequential tracking and memorising of the previous frames of the input, which by using RNN may result in less accuracy in the prediction.

B. Long Short Term Memory

Long Short-Term Memory (LSTM) networks are a modified version of recurrent neural networks, which makes it easier to remember past data in memory.Limitations faced by RNN is resolved in LSTM model. LSTM has a great architecture as shown in figure [2] with forget gate, input gate and output gate, which classifies necessary and unnecessary information. Considering the advantages of LSTM which suits our project, we chose LSTM model to go with.

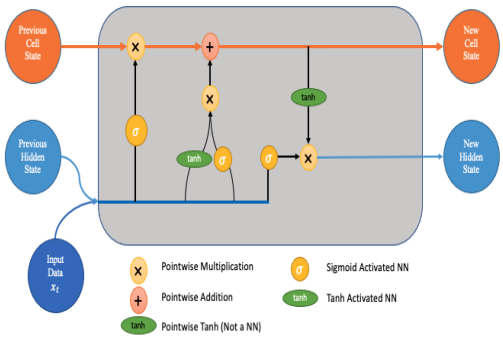

Fig. 2.1 Long Short Term Memory Architecture

C. MediaPipe framework

MediaPipe is a framework provided by Google. MediaPipe Hands is a high-fidelity hand and finger tracking solution. It employs machine learning (ML) to infer 21 3D landmarks of a hand from just a single frame. MediaPipe provides the tracking of the hand gestures with x, y and z coordinates respectively represents the location of the hand.

III. LITERATURE SURVEY

In paper[1], Thai Sign Language Recognition: Application of Deep Neural Network was published in IEEE, in the year 2020. This paper uses RNN neural network, MediaPipe framework, LSTM model which predicts only the Thai sign language.

In paper[2], Hand gesture classification was published in ICICS in the year 2018. The methodology of the project is CNN.

In paper[3], Real time sign language interpreters published in ICICS in the year 2017. The methodology of the project is based on sensory gloves.

In paper[4], Automatic action recognition was published in IEEE in the year 2017. This project uses ANN, which only detects isolated classification. Background disturbance for hand detection is the disadvantage.

In paper[5], the SSVM classifier was published in ICICS in the year 2018. This paper has a sign language dataset trained with 13 languages with 80% accuracy. The disadvantage of the project is only the right hand gesture is implemented.

In paper[6], Indian Sign Language Recognition was published in IEEE in the year 2010. Discrete Wavelet Transform (DWT) based feature extraction and nearest neighbor classifier is used to recognize the sign language is the methodology used in this paper.

In paper[7], Dynamic Gesture Design and Recognition for Human-Robot was published in ICICS in the year 2013.

B. Dataset preprocessing

Collected dataset was down MediaPipe hands which will collect both right and left hands. Some sign languages doesn't require both the hands in that case no key points will be generated. To improve the efficiency and to clean the dataset,

The methodology of the project is Motion History Image (MHI) and Convolutional Neural Networks (CNN). Only 10 classifications are trained in this project.

In paper[8], Arabic sign language recognition was published by IEEE in the year 2017. Methodology of this paper is 3D CNN. Arabic sign language gestures are only detected for isolated classification.

IV. MODULES

In Sign language recognition classification is mainly focused on hand gestures and the position of all key points are graphed.

A. Dataset collection

Indian sign language gestures are Our dataset is collected using OpenCV and MediaPipe. Each video will contain 30 frames as a limit and 50 videos per sign language. While collecting frames using OpenCV the frame is processed with MediaPipe to mark landmarks which graphs the hand gesture. From the drawn landmarks coordinates are collected as x, y and z co-ordinates. Collected key points are processed into single array per frame and saved as numpy array, which is our dataset.

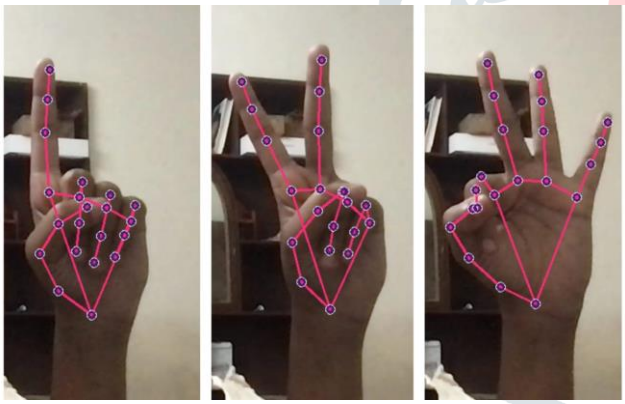

Fig. 3.1. Indian sign language gestures with landmarks marked

assigned in the model is categorical accuracy. Model is finally trained with the categorical accuracy of 96% and accuracy score of 97%.

wherever other hand is not recored, an empty numpy is generated with 0 as value. Collected dataset is labeled using dictionary, where labels are saved in integer format.

C. Training and Testing split

In the project for the better training and testing, training is spliced with 80% of the dataset and testing with 10% dataset.

D. Model creation

We have created our sequential model layers. Our layers consist of 3 LSTM layers and 3 Dense layers. Dense Layer is simple layer of neurons in which each neuron receives input from all the neurons of previous layer, thus called as dense. Dense layer interconnect each and every layers' input and output which makes the model work efficiently. The model summary is as following Fig. 4.3

```
array([[[ 0.00000000e+00,  0.00000000e+00,  0.00000000e+00, ...,
          0.00000000e+00,  0.00000000e+00,  0.00000000e+00],
        [ 0.00000000e+00,  0.00000000e+00,  0.00000000e+00, ...,
          0.00000000e+00,  0.00000000e+00,  0.00000000e+00],
        [ 0.00000000e+00,  0.00000000e+00,  0.00000000e+00, ...,
          0.00000000e+00,  0.00000000e+00,  0.00000000e+00],
        ...,
        [ 6.44552767e-01,  8.91422510e-01,  2.63464813e-07, ...,
          4.72082436e-01,  6.94364786e-01, -2.82747634e-02],
        [ 6.45690918e-01,  8.90465379e-01,  2.33635902e-07, ...,
          4.78199661e-01,  6.75708532e-01, -2.53940523e-02],
        [ 6.45231128e-01,  8.88323426e-01,  2.45933109e-07, ...,
          4.78323519e-01,  6.88351214e-01, -1.50280753e-02]],

       [[ 6.48053706e-01,  8.87597084e-01,  2.29817488e-07, ...,
          4.78480399e-01,  6.82501376e-01, -1.60410386e-02],
        [ 0.00000000e+00,  0.00000000e+00,  0.00000000e+00, ...,
          0.00000000e+00,  0.00000000e+00,  0.00000000e+00],
        [ 0.00000000e+00,  0.00000000e+00,  0.00000000e+00, ...,
          0.00000000e+00,  0.00000000e+00,  0.00000000e+00],
        ...,
```

Fig. 4.2. Dataset collection after preprocessing

E. Model Training

Optimization is an important process which optimize the input weights by comparing the prediction and the loss function. Optimiser were using while compiling is Adam optimizer. Only metric

Indian sign language interpreter in realtime with high accuracy and low loss with 21 classifications is achieved in this project. The result of the project, which is real time translation is shown in Fig. 6.3.

```
Layer (type)                 Output Shape              Param #
=================================================================
 lstm_27 (LSTM)              (None, 30, 64)            48896

 lstm_28 (LSTM)              (None, 30, 128)           98816

 lstm_29 (LSTM)              (None, 64)                49408

 dense_27 (Dense)            (None, 64)                4160

 dense_28 (Dense)            (None, 32)                2080

 dense_29 (Dense)            (None, 21)                693

=================================================================
Total params: 204,053
Trainable params: 204,053
Non-trainable params: 0
```

Fig. 4.3. Model Summary

F. Testing in realtime

Live video frames are collected using OpenCV and the frames are processed to extract key points by drawing landmarks using MediaPipe. Using the extracted key points as input, which is passed through the model to be predicted.

V. METHODOLOGY

Hardware requirements of the project are processor greater than 4GB RAM and i3 7th generation and web camera. Software Requirements for the project are Open CV, Django framework, Python, Keras, Tensorflow, Numpy, Sklearn, MediaPipe and Pandas.

Using website interface input video frames are collected using OpenCV. The collected frames are being processed and sent through the model to be predicted. The predicted output text is displayed as text.

VI. RESULTS AND DISCUSSION

This project results in Indian sign language recognition using deep learning in real time. Realtime translation of actions add advantage to the project. The project is integrated in web application using Django framework. The model is trained with 96% accuracy. The accuracy score is as shown in Fig.6.1. The categorical accuracy is 0.9667 and loss of 0.10, which is shown in the Fig.6.2. The aim of the project which is developing an

VII. CONCLUSION AND FUTURE WORKS

Increasing the number of actions to the model is the future plan of the project. This project's future is to be integrated in mobile application and desktop application. Experimenting the project in various models and comparing and choosing the best and integrating with the project.

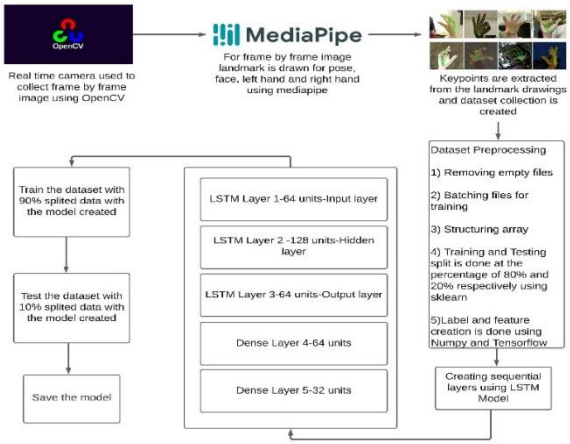

Fig. 5.1. Architecture of the project

```
In [40]: accuracy_score(ytrue,yhat)

Out[40]: 0.9777777777777777
```

Fig. 6.1. Accuracy score

```
Epoch 84/500
25/30 [=================>.....] - ETA: 0s - loss: 0.1115 - categorical_accuracy: 0.9663
```

Fig.6.2. Categorical accuracy and categorical loss

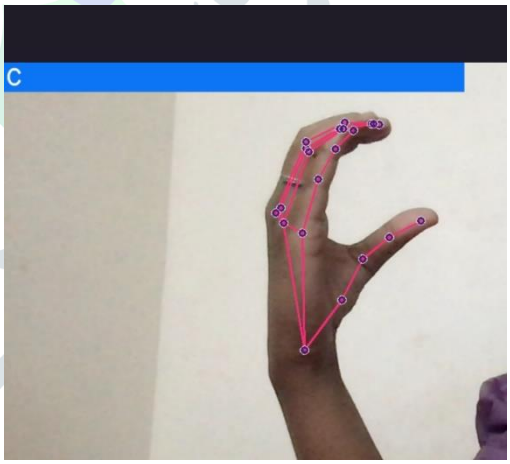

Fig. 6.3. Real time translation of Indian Sign language result

11. Kartik Shenoy, Tejas Dastane, Varun Rao, Devendra Vyavaharkar (2018), "Real-time Indian Sign Language (ISL) Recognition", The Institute of Electrical and Electronics Engineers 2018.
12. M. Burić, M. Pobar, M. Ivašić Kos (2017), "An Overview of Action Recognition in Videos", Manufactured Imports Promotion Organization 2017.
13. P. Shopa, N. Sumitha, and P. S. K. Patra, 'Traffic sign detection and recognition using OpenCV,"

VIII. REFERENCES

1. Anusorn Chaikaew, Kritsana Somkuan and Thidalak Yuyen (2020), "Thai Sign Language Recognition: Application of Deep Neural Network", NCON.
2. Kumar S, Yadav G, Singh HP, Malhotra S, Gupta A (2018), "Hand gesture classification", International Conference on International and Communications Security, 2018.
3. Luis Quesada, Gustavo Lopez and Luis A.Guerrero (2017), "Automatic recognition of American sign language classifiers", The Institute of Electrical and Electronics Engineers 2017.
4. Madhuri Sharma, Ranjna Pal and Ashok Kumar Sahoo (2014), "Indian Sign Language recognition using neural networks and kNN classifiers", 2014.
5. Yun-Jye Lee (2018), "SSVM Classifier", International Conference on International and Communications Security 2018.
6. Haodong Chen, Wenjin Tao and Ming C.Leu(2020), "Dynamic Gesture Design and Recognition for Human-Robot", The Institute of Electrical and Electronics Engineers 2020.
7. El Badawy M, Elons AS, Shedeed HA (2017), "Arabic sign language recognition", International Conference on International and Communications Security.
8. "Real-time American sign language recognition", The Institute of Electrical and Electronics Engineers 1998. Wang hao Liang and Liu Guixi (2000), "Dynamic and combined gestures- recognition based on multi-feature fusion in a complex environment", The Institute of Electrical and Electronics Engineers .
9. Sruthi C. J and Lijiya A (2019), "Signet: A Deep Learning based Indian Sign Language Recognition System", The Institute of Electrical and Electronics Engineers.
10. Geethu G Nath and Arun C S (2017), "Real Time Sign Language Interpreter", Industrial Credit and Investment Corporation of India 2017.
The Institute of Electrical and Electronics Engineers 2015.
14. Pulver, Andrew; Lyu, Siwei (2017)," Neural Networks (IJCNN) - LSTM with working memory", The Institute of Electrical and Electronics Engineers.
15. Majida Ali Abed , Ahmad Nasser Ismail and Zubadi MatizHazi (2010)," Pattern recognition Using Genetic Algorithm",International Journal of Computer and Electrical Engineering.