International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

IJIEMR Transactions, online available on 10th Apr 2023. Link :http://www.ijiemr.org/downloads.php?vol=Volume-12&issue=Issue 04

10.48047/IJIEMR/V12/ISSUE 04/83

Title LSTM DEEP LEARNING APPROACH FOR INTERPRETING SIGN LANGUAGE

Volume 12, ISSUE 04, Pages: 682-689

Paper Authors

Mr.Md.Shakeel Ahmed, CH MSVN LAKSHMI DEDEEPYA,GUDIPALLI SUHRUDHA,GANJI KAVYA SRI, GUDETI LAVANYA

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

LSTM DEEP LEARNING APPROACH FOR INTERPRETING SIGN LANGUAGE

Mr.Md.Shakeel Ahmed[1],Associate Professor, Department of IT,
Vasireddy Venkatadri Institute of Technology, Nambur, Guntur Dt., Andhra Pradesh.

CH MSVN LAKSHMI DEDEEPYA[2],**GUDIPALLI SUHRUDHA**[3],**GANJI KAVYA SRI**[4],**GUDETI LAVANYA**[5]
UG Students, Department of IT,
Vasireddy Venkatadri Institute of Technology, Nambur, Guntur Dt., Andhra Pradesh.
[1] cdedeepya78@gmail.com, gudipallisuhruda176@gmail.com,
kavyaganji4@gmail.com, lavanyag12345678@gmail.com

Abstract

The Interpreting of Sign Language is a challenging task due to its complexity and variability of hand gestures. In recent years, deep learning models such as CNN had shown promising results in solving the problem. But they are of high and are best suitable for recognition of alphabets than words. This paper proposes LSTM Deep Learning model to recognize hand gestures. Our project's major goal is to eliminate communication barriers between people who are deaf or hard of hearing and regular people by translating their gestures or actions into text(word form) that is clear to everyone. In order to accomplish this, we are employing a camera-based computer system called OpenCV that has been trained using a neural network to recognise and translate signs[1]. In our study, Media Pipe Holistic is used to recognise the facial, hand, and stance movements that make up a gesture. We then use this information to create Long Short Term Memory (LSTM) layers, which are part of the neural network that learns and identifies these signs. This neural network is used to recognise different movements, with the output appearing as text on the screen.

Keywords: LSTM, OpenCV, Neural Network, Mediapipe Holistic, Deep Learning.

Introduction

Over the world, people who are deaf or have trouble speaking prefer to communicate through sign language. To communicate with them, one should learn sign language. For learning signs, there aren't many study resources available. Learning sign language as a result is a very challenging procedure. Most of the currently available solutions for learning sign language rely on pricey external sensors. With the collection of a dataset and the use of several feature extraction techniques, our study intends to advance this subject by classifying them into sign language gestures. This project is designed to discover signs from set of signs and giving text as output. Using computer vision to recognise sign language might be successful or unsuccessful depending on the situation. A systematic set of movements with each gesture having a distinct meaning is said to exist in sign language[2]. Computers may learn to understand sign language with the use of deep learning and image categorization, which can then be translated by humans.

To recognise sign language gestures, Long Short Term Memory (LSTM) neural networks can be used. Dynamic sign language gestures recorded with a web camera will make up the video dataset used. The recorded video will go through pre-processing, and the cleaned input will then be used.

The outcomes are attained by retraining and testing a Long Short Term Memory (LSTM) neural network model on this dataset. By creating software that can instantly predict motions, this initiative hopes to close the communication gap. Certain common motions are identified and rendered in text form for ease of explanation by ordinary people.

2. Literature Survey

Deep Learning has been used in several studies for sign language recognition.

A Deep Learning-based Indian Sign Language Recognition System with a training accuracy of 99.93% and testing and validation accuracy of 98.64%, this system was successfully trained on all ISL static alphabets. The disadvantage of the work that is missing is context analysis and facial expression.

Hand Gesture Recognition for Sign Language Using 3DCNN - The use of 3DCNN for hand gesture identification is investigated in this paper. Six more cutting-edge techniques from the literature were compared to the ways the authors had suggested. They fared better than four of them and equally well as the other two techniques. It does not, however, function for a live video broadcast. **Video-based isolated hand sign language recognition using a deep cascaded model -** To develop a deep-based model for systematically recognizing hand signs, the authors combined SSD, CNN, and LSTM using RGB Videos. With the use of this model, hand sign recognition's complexity and accuracy were both increased. It offered quick processing when faced with an unpredictable environment, such quick hand movements. The detection's precision may be increased with more data. **A Modified-LSTM Model for Continuous Sign Language Recognition using Leap motion -** The Leap motion sensor is used in this paper to offer a unique architecture for continuous-SLR. For the recognition of sign words and sentences, a modified LSTM architecture has also been put forth. On the signed sentences and isolated sign words, respectively, average accuracy rates of 72.3% and 89.5% have been noted. By adding more training data for better model learning, the recognition

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

performance can be enhanced. **Deep Convolutional Neural Networks for Sign Language Recognition** - The authors suggested a CNN architecture for categorizing motions used in selfie sign language. The suggested CNN architecture shows less training and validation loss, but the database isn't openly accessible.

3. Problem Identification

The Problem Identification for Recognition of Sign Language using LSTM model can be summarized as follows:

1) How do I start building Sign Language Recognition System?

Building a Sign Language Recognition System Using LSTM Deep Learning Model that identifies signs and display corresponding text of the sign on the screen[3]. The main focus of our system is to develop a method that makes use of Media Pipe Holistic to recognize the facial, hand, and stance gestures made by the person, then uses this information to create Long Short Term Memory (LSTM) layers, which create the neural network that learns and identifies these signs.

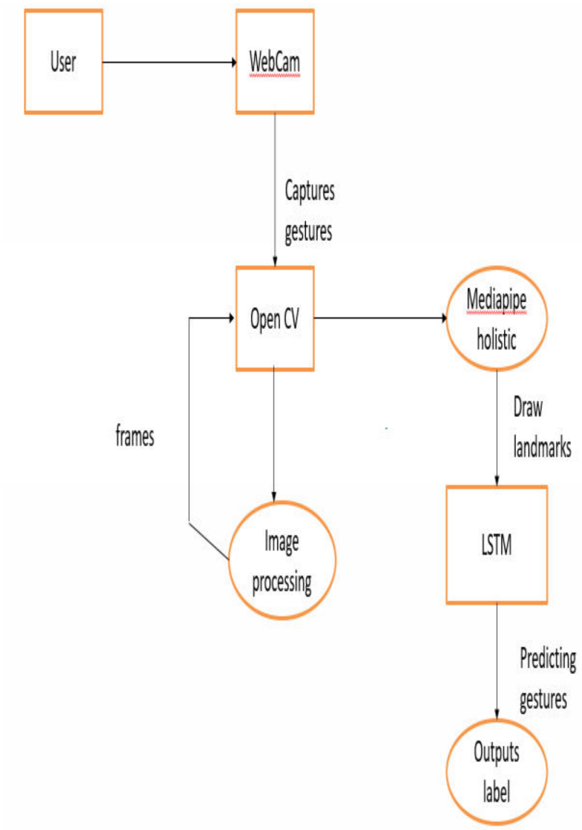
4. Methodology

Long short-term memory (LSTM) is frequently suggested as a solution to difficulties with cost, complexity, and short-term memory. Such a system is used to choose the best time delay and choose when to forget specific information. In order to detect and recognize sign language motions from a video source and to create the associated English word, the methodology suggested in this study leverages LSTM. The video source, which consists of ISL sign language gesture sequences, is used as an input in this technique.

The following is the step-by-step description of how to build the sign language recognition system using LSTM model:

1) Install and import all required packages.
2) Define a function Media pipe holistic to identify face, hands and pose movements of the person.
3) With the help of media pipe we can draw and style the landmarks of the person gesture.
4) OpenCV (Open Source Computer Vision) library is a python library used for image processing and computer vision tasks.
5) Next we will set the array of actions that system needs to identify based on a particular gesture.
6) Now we will train the model for this array of actions.
7) Next we assign labels(integers) for these actions.
8) We use TensorFlow to achieve a good level of accuracy even with a limited dataset.

9) Finally, we will test the model which produce efficient results i.e. display correct text as output for the corresponding action.



System Implementation

In this project, the neural network will be constructed using LSTM (Long Short Term Memory). Long-range dependencies are capture able by LSTM. It is capable of maintaining a long-term memory of prior inputs. LSTM is the greatest choice for our project because we are working with video datasets. Here are the explanations for selecting LSTM:

When dealing with continuous and non-static data, LSTM works well.

LSTM is faster to train, takes less data to provide more accurate results, and produces faster detections than other neural networks that rely on CNN[4]. Instantiating the model is the initial step in creating a neural network. To develop our network, we are employing a sequential model. A simple stack of layers with exactly one input tensor and one output tensor per layer is suitable for a sequential model.

Once the sequential model has been created, the LSTM layers are added using the input shape, number of neurons, and activation function as parameters "relu" serves as the activation function.

The activation function in a neural network is in charge of activating the node or output for that input from the node's summed weighted input. Comparatively speaking to other CNN-using neural networks, it produces faster detections.

The LSTM layers are inserted first, then the thick layers. These layers, commonly referred to as the hidden layers, are entirely interconnected. The output layer is shaped exactly like the input layer.

Adam, the optimizer

Categorical cross entropy is the loss.

Metrics: categorical precision

By providing the training data and setting the epochs, the model is finally fitted. A neural network is trained in an epoch using all of the training data for a single cycle.

Prerequisites

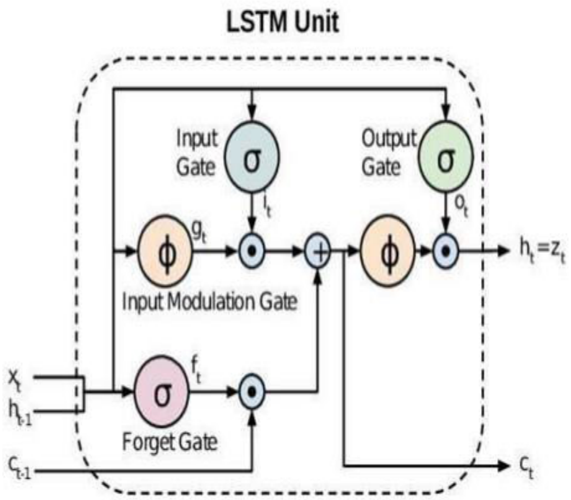
To follow along the reader will need the following:

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

1. Python installed on your system and access to the jupyter notebook.
2. Installing the packages and libraries are : OpenCV, Tensorflow, Sklearn, Matplotlib and Mediapipe in jupyter notebook.

LSTM(Long Short Term Memory):

In very deep neutral networks, there is a problem known as the vanishing gradient problem. Sepp Hochreiter and Juergen Schmidhuber developed Long Short-Term Memory to solve the vanishing gradient issue in RNNs. An improvement to the RNN hidden layer is LSTM.



An LSTM cell contains 3 gates. These gates are utilised in LSTM memory modification. The gradient propagation in the recurrent network's memory is controlled by gates in long short-term memory (LSTM).

1. Forget Gate : Eliminates data from the cellstate that is no longer necessary.
2. Input Gate: The input gate adds more helpful data to the cellstate.
3. Output Gate: The output gate augments the cellstate with additional valuable data. For sequence models, the widely used deep learning algorithm LSTM is used. The success of those apps can be attributed to the LSTM algorithm, which is employed in real-world applications like Apple's Siri and Google's voice search. Recent studies have demonstrated how the LSTM algorithm can enhance the effectiveness of the machine learning model. LSTM is also employed for text classification and time-series predictions.

OpenCV

A computer vision and machine learning software library called OpenCV is available for free download. To facilitate the use of machine perception in commercial goods and to provide a common foundation for computer vision applications, OpenCV was developed. OpenCV makes it simple for businesses to use and alter the code as it is a BSD-licensed product.

In this project, OpenCV was utilised for the following purposes:

videoCapture() - The VideoCapture() method from OpenCV is used to interact with the camera. Here is a task that we can complete: Video can be displayed, read, and saved. display a camera capture that was made.

it's "isOpened() - It returns true if the initialization of video capture has already occurred. The method returns true if the last successful execution of the VideoCapture function Object() { [native code] } or VideoCapture::open() function.

cvtColor() - The cvtColor() function in OpenCV accepts two parameters, image and code, where image is an image whose colour space is to be converted to another colour space and code is a representation of the colour conversion code.

Numpy

"Numeric Python" or "Numerical Python" is referred to by the acronym NumPy. Fast precompiled functions for mathematical and numerical routines are provided by this open-source Python extension package. NumPy also adds strong data structures to Python, enabling Python programmers to compute multi-dimensional arrays and matrices quickly and effectively.

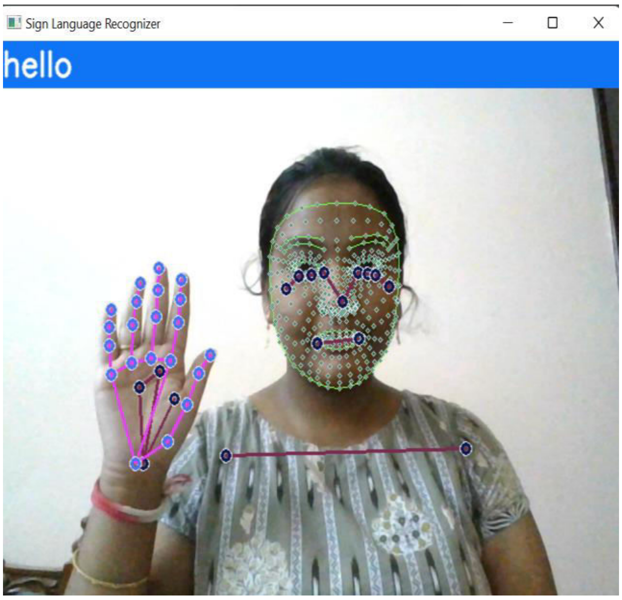
Mediapipe Holistic

One of the pipelines, called Mediapipe Holistic, has face, hand, and pose optimizations that enable holistic tracking and let the model simultaneously recognise hand and body poses in addition to facial landmarks. To identify faces and hands and extract keypoints to send to a computer vision model is one of the main uses for MediaPipe holistic.

The above fig shows Mediapipe landmark detection of hands, pose and face movements.

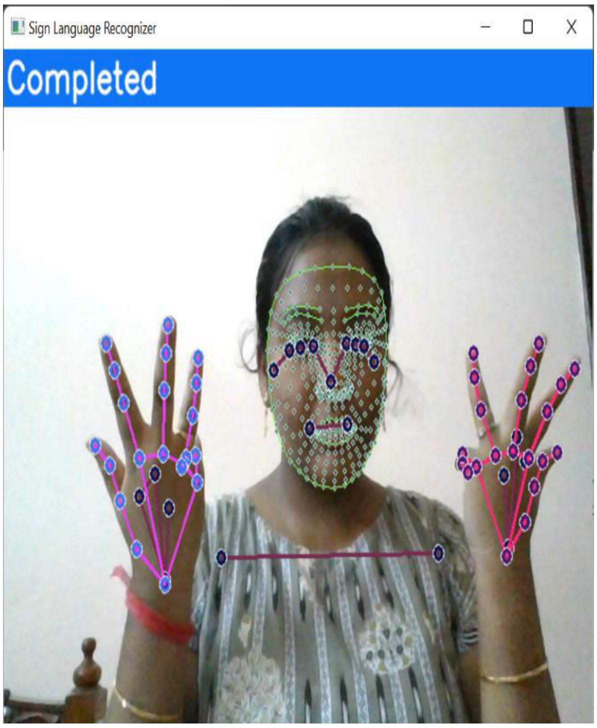
6. Results

Output1: Hello

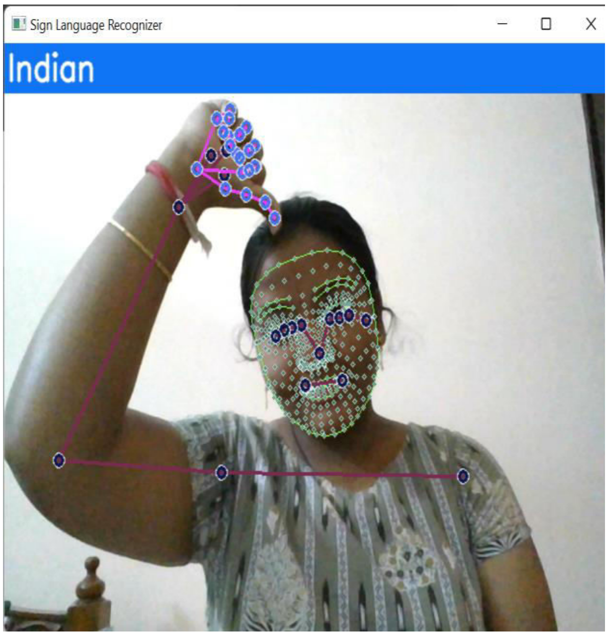


International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

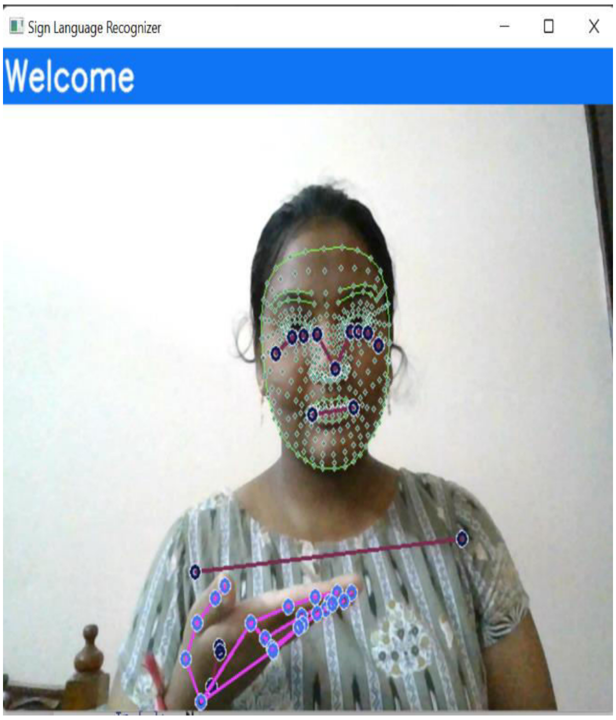
Output2: Completed



Output3: Indian



Output4: Welcome



Output5: Thank you

7. Conclusion

The goal of the Sign Language Recognizer is to close the communication gap between a hearing- or speech-impaired person and a regular person. Our solution uses a video of a person's motions to show the recognised sign on the screen. Even when there is less data, the MediaPipe across LSTM layers produces reliable results. In comparison to other techniques that use layers, such CNN, the neural network built over the LSTM makes detections more quickly[5]. Because it recognises signs from dynamic gestures rather than from static photos, our project is a trustworthy sign language recognizer. Five signs—"Hello," "Thank You," "Welcome," "Indian," and "Completed"—have been identified in our research.

This project can be expanded by teaching the system other signs, and it may eventually be turned into an application.

8. Future Work

This project could be turned into an app that speechless people could use to improve their quality of life. It could also be used by government agencies that are responsible for providing equal service to all citizens, private businesses that want to reach and assist speechless people, and foundations and co-ops that aim to aid people with speech disorders. We can create a model to recognise ISL (Indian Sign Language) sentences at the sentence level. Another area of inquiry is how to translate the series of gestures into audible words. It is possible to create a dependable application that is more accurate and works flawlessly as a translator for sign language.

9. References

1. Nelson, A., Price, K. J., & Multari, R. (2019). ASL reverse dictionary-ASL translation using deep learning. *SMU Data Science Review*, *2*(1), 21.

2. Ong, S. C., & Ranganath, S. (2005). Automatic sign language analysis: A survey and the future beyond lexical meaning. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, *27*(06), 873-891.

3 Sonare, B., Padgal, A., Gaikwad, Y., & Patil, A. (2021, May). Video-based sign language translation system using machine learning. In *2021 2nd International Conference for Emerging Technology (INCET)* (pp. 1-4). IEEE.

4. Lopez-Martin, M., Carro, B., Sanchez-Esguevillas, A., & Lloret, J. (2017). Network traffic classifier with convolutional and recurrent neural networks for Internet of Things. *IEEE access*, *5*, 18042-18050.

5. Duy Khuat, B., Thai Phung, D., Thi Thu Pham, H., Ngoc Bui, A., & Tung Ngo, S. (2021, February). Vietnamese sign language detection using Mediapipe. In *2021 10th International Conference on Software and Computer Applications* (pp. 162-165).