ijRASET

International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL
FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

www.ijraset.com

Call: ⓒ08813907089 | E-mail ID: ijraset@gmail.com

Real Time Sign Language Recognition System for Hearing and Speech Impaired People

Tanmay Petkar[4], Tanay Patil[3], Ashwini Wadhankar[1], Vaishnavi Chandore[5], Vaishnavi Umate[6], Dhanshri Hingnekar[2]

[1, 2, 3, 4, 5]*Department of Computer Engineering, Bapurao Deshmukh College of Engineering, Sevagram*

Abstract: Sign Language is globally used by more than 70 million impaired people to communicate and is characterized by fast, highly articulate motion of hand gesture which is difficult for verbal speakers to understand. This limitation combined with the lack of knowledge about sign language by verbal speakers creates a separation where both parties are unable to effectively communicate, to overcome this limitation we propose a new method for sign language recognition using OpenCV (A python library) which is used for pre-processing images and extracting different skin toned hands from the background. In this method hand gesture are used to make signs which are detected by YOLOv5 algorithm for object detection which is the fastest algorithm till date while Convolutional-Neural-Networks (CNN) are used for training gesture and to classify the images, and further we proposed a system which translates speech into sign language so that the words of the verbal speaker can be transmitted to the deaf/mute. This automated system first detects speech using the JavaScript Web-Speech API and converts it into text because the recognized text is processed using the Natural Language Toolkit and aligns token text with the sign language library (sign language videos) videos according to well-known text and finally shows a compiled output which is displayed through avatar animation for a deaf / dumb person. The proposed system has various advantages like Portability, User-friendly Interface and Voice Module. The software is also very cost-effective which only needs a laptop camera or webcam and hand gesture, system accuracy is compared to high-quality methods and is found to be the best.
Keywords: Sign language, Hand gesture, OpenCV, Python library, Image processing, YOLOv5, Object detection, Pose estimation, Machine learning.

I. INTRODUCTION

World Health Organization (WHO) survey states that above 6% of the world's population is suffering from hearing impairment these people use sign languages to communicate that are visual representation of thoughts through hand gestures, facial expressions and body movements which is difficult to understand by verbal speakers as sign language is completely independent language from its counterpart of verbal language. Also, for sign languages they have their own grammar and syntax. There are different types of sign language based on location and language status. For example, American Sign Language, Bangla Sign Language, Indian Sign Language, etc. Also, the same gesture can be interpreted differently depending on the variations used by the user in the same sign language. Sometimes one gesture represents the whole word at a time can represent only one alphabet or number, sometimes in addition, sometimes a combination. The facial expressions of the hands are used for communication.

There are three types of sign language. Are as follow:
1) Non-manual features: Tongue, facial expression, body pose, and hand gesture - all of them are used to communicate.
2) Word level sign spelling: Each gesture presents a whole word.
3) Finger vocabulary: One gesture represents the alphabet/numbers.

Our proposed system is aimed towards a fully functional system that can recognize the signs or gestures and display output accordingly. OpenCV is the huge open-source library for computer vision and image processing that can be used to process images and videos to identify objects faces or even hand-writing of a human. As it is stated that OpenCV can be used to process images, it is used in our proposed system for image pre-processing and extracting the different skin toned hand gestures from the background. YoloV5 is the best and fastest object detection algorithm till date. YOLO algorithm first takes data to be trained in xml files and distributes its action format framewise for every particular frame in their txt files, then for the training model it employs convolutional neural networks to detect objects in real-time.

A Deep Learning method called Convolutional Neural Network is very helpful in identifying various features of the images in the spatial domain. Pixels of the images are treated as neurons and then processing is done by neuron-by-neuron. The varying number of kernels are applied at different layers of the convolutional neural networks for extracting shapes of fingers. Towards the end, it classifies the images into various groups based upon the features.

In this manner every neuron is linked to a neuron of the next layer and previous layer as well and forms fully connected layers in the network. So here pre-processed images present in the dataset are fed to the convolutional neural network formed and then the model gets to train and tested. Once tested with the dataset image it can identify the signs performed in real-time.

Therefore, it is of great concern to pave the way for the deaf to not only reduce the gap for deaf people to translate into sign language but also to enable them to become self-reliant through the platform of self-education and learning sign language. This system proposes an easy-to-use, timely and comprehensive online forum for deaf people designed as a web app that works as an effective way to communicate and learn from them. The model has a four-stage structured task: Speech acquisition using PyAudio, Conversion to text using the JavaScript Web-Speech API (uses text token and NLP concepts for processing), text comparisons and visual signals library (video set of sign language data) from sign language, to composing similar videos according to the sequence of the processed text and presentation to the deaf person.

II. LITERATURE SURVEY

Tanuj Bohra et al. proposed a two-way real-time sign language conversion program based on image processing for in-depth reading using computer vision. Procedures such as hand detection, skin colour separation, medium blur and frame detection are performed on images in the database for best results. CNN model trained with a large database of 40 classes and able to predict 17600 test images in 14 seconds with 99% accuracy.

Joyeeta Singha and Karen Das proposed the Indian Sign Language Recognition Program in a live video. The program consists of three stages. The pre-screening process involves skin filtering and histogram matching. Eigen-values and eigen-vectors are considered in the output factor category and the Eigen value that measures the Euclidean distance to be divided. The Dataset contained 480 images of 24 ISL symbols signed by 20 people. The system was tested on 20 videos and gained 96.25% accuracy.

Muthu Mariappan H. and Dr. Gomathi V have designed a real-time sign language recognition system as a portable unit that uses contour detection and an incomprehensible algorithm for c-means. Outlines are used to see the face, left hand and right hand. While the k means algorithm is incomprehensible it is used to divide the input data into a specific number of clusters. The program was used on a database containing video recordings of 10 signers for a few words and sentences. It was able to achieve 75% accuracy.

Salma Hayani et al. proposed a CNN-based Arabic sign language recognition program, persuaded from LeNet-5. The database contained 7869 images of Arabic numerals and letters. Various tests are performed by changing the number of training sets from 50% to 80%. 90% accuracy is achieved with 80% training database. The author also compared the results obtained with machine learning algorithms such as KNN (closest neighbor) and SVM (support vector machine) to demonstrate system performance. This model was based on image only and can be extended to video-based identification.

Kshitij Bantupalli and Ying Xie built an American sign-language video recognition system based on Convolution Neural Networks, LSTM(Long Term Short Memory) and RNN(Recurrent Neural Network). A CNN model called Inception was used to extract local features from frames, LSTM long-term dependence and RNN to extract temporary features. Various tests were performed for different sample sizes and the database contains 100 different markers performed by 5 signers and a high accuracy of 93% was achieved. Sequences are then added to LSTM for longer durations. SoftMax layer output and max pooling layer are provided in the RNN architecture to extract temporary features in the SoftMax layer.

Mahesh Kumar put forward a system that can identify 26 sign language gestures in Indian Sign Language based on Linear Discriminant Analysis (LDA). Pre-processing measures such as skin separation and environmental performance are used in the database. The separation of the skin is done using the Otsu algorithm. Discrimination line analysis is used to exclude the feature. Each gesture is presented as a column vector in the training phase and then customized with respect to the median gesture. The algorithm detects eigenvectors of the variance matrix for median gesture. In the recognition phase, the subject vector is usually relative to the median gesture and then displayed in the gesture space using the eigenvector matrix. The Euclidean range is calculated between these speculations and all known assumptions. A small number of these comparisons were selected.

Suharjito et al. attempted to use a sign language recognition system with the I3inception model using the transfer learning method. The public data set LSA64 is used in 10 words with 500 videos. For training the database is distributed in a 6: 2: 2 ratio, 300 training videos, 100 verification and 100 test sets. The model has good training precision but very low validation accuracy.

Juan Zamora-Mora et al. introduced CNN-HMM which is a hybrid of sign language recognition. They did experiments on three databases namely RWTH-PHOENIX-Weather 2012, RWTHPHOENIX-Weather Multi Signer 2014 and one SIGNUM signer. The training and certification set has a rating of 10 to 1. After the end of CNN training the SoftMax layer is added and the results are applied to HMM as viewing opportunities.

Mengyi Xie and Xin Ma put forward an end-to-end program using a residual neural network to initiate American Sign Language recognition. The data set contains 2524 images of 36 classes. Data enrichment is used to expand the database to 17640 images. These images are converted to a CSV file format and after inserting hot coding and are provided as embedded in the ResNet50 network for training. The model provides 96.02% accuracy without data development and accuracy improves with data enrichment up to 99.4%.

G. Anantha Rao et al. raises Indian sign language gesture recognition using a convolutional neural network. This application applies to videos taken from the front mobile camera. Database created by making 200 ISL(Indian Sign Language) symbols. CNN training is done on 3 different databases. In the first group, a single set of information sets is provided as input. The second set consists of 2 sets of training data and the third set respectively contains 3 sets of training data. The average visibility of this CNN model is 92.88%.

Aditya Das et al. trained a convolutional neural network using the Inception v3 model of American Sign Language. Data augmentation is applied to photos before training to avoid overcrowding. This model provides more than 90% accuracy in the Sreehari sreejith database of 24 class labels with 100 images per class.

III.METHODOLOGY

Our proposed system is a sign language recognition system that detects a variety of gestures by recording video and converting it into independent sign language labels. Hand pixels are then classified and matched to an image obtained and sent to be compared with a trained model. So, our system is very strong in finding specific character labels. Our proposed system is a sign language recognition system that detects various gestures by video recording and to convert it into independent frames. Then the hand pixels are separated and matched to the image obtained and shipped for comparison with a trained model, so our system is very tight finding specific text labels for characters. The Proposed System consists of Collaborative Communication which allows users to communicate properly due to language or speech barriers, the proposed system also consists of Embedded Voice Module with a User-Friendly Interface. This system can be used by both verbal speakers and sign language users for communication, which is the biggest advantage of this proposed system. The proposed system works on Python with YOLOv5 Algorithm which works with modules like Graphical User Interface for easier use, Training Module to train CNN models, Gesture Module to allow users to create their own Gesture, Word Formation Module to Create a word by combining gesture and the speech module that converts the converted text to speech. Our proposed system is designed to address the problems faced by the deaf people in India. This system is designed to translate each word received as input into sign language. This project translates words based on Indian Sign Language.

1) *Natural Language Processing:* Filler phrases like 'is', 'has', 'changed into', 'these' and so on. Are phrases that infrequently make contributions to the context in signal language conversion. Therefore, the system gets rid of those filler words from the sentence.
2) *Root Words:* Words can be in gerund form, plural form or adjective form. The proposed device will dispose of those types of phrases and locate root phrases from those phrases. These primary words could be helpful in powerful conversion of sign language.
3) *Dataset:* The gadget has a large dataset of Indian Sign Language phrases to map from speech to diagnosed text or textual content. So, it is going to be beneficial for all the deaf humans in India. It facilitates people to understand maximum of the speech or textual content.

The Objectives of our Proposed System are: -
a) To design a system for speech/hearing impaired that provides a better way to confer in public.
b) To implement a plan to reduce the communication gap between two types of people
c) To provide a Universal Sign Language Recognition System which can be implemented using this idea.
d) To give a system which helps communicate with people from different regions around the world.
e) To develop a system that helps people in learning Sign Language.

A. System Architecture

In this paper, we have mentioned below an architecture which describes the complete understanding about how it works.
Our Proposed System works in both ways: Sign-Language to Text Conversion and Text to Sign-Language Conversion. Let's discuss the working of Sign-Language to Text Conversion. We have created our own unique dataset by recording and saving gestures through a laptop camera or webcam with the help of OpenCV. After successful creation of the dataset, we need to train the dataset using TensorFlow which helped us in achieving accuracy of 90% and predicting the text accurately.

Now, if any verbal speaker wants to establish communication between the impaired people, the Text-to-Sign-Language Conversion proves to be helpful in converting the basic sentences or words into sign language. To make it more accessible and easier to understand we've created an Avatar using Blender 3D tool and animated the equivalent gestures for the alphabets and words. Creation of our own unique dataset of the Animated gestures which would help us to translate the text input given by the user into its equivalent gestures using NLTK. The user can input the text manually or by using live voice functionality.
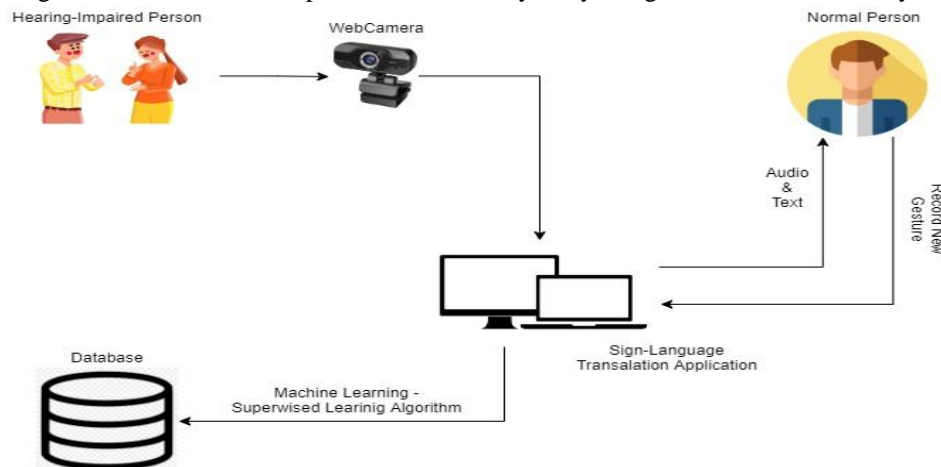

Fig. 1 System Architecture

B. System Design

1) Forms of Input

Our project aims to receive input in multiple formats. Inputs can be of the form:

a)   Text input

b)   Live speech input

2) Speech Recognition

Real time voice is received as input from the microphone of our system. This is done using the JavaScript Web Speech API. The audio thus received is converted into text using the JavaScript Web-Speech API. It is an API that helps in converting audio to text by incorporating neural network models. The audio received using this web-speech recognizer is translated into text. For long audio inputs, the audio is split into smaller pieces based on the occurrence of silence. The fragments are then passed to a web-speech recognizer to effectively convert them to text.

3) Preprocessing of Text

Filler words are used to fill in the blanks in the sentences and are obviously words with a little meaning. They give a little context in the sentence. There are about 30+ filling words in the English language that are not well understood in a sentence. Thus, the system removes the filling words in the sentence and makes them more meaningful. Deleting these words will save system time. The system also removes any punctuation marks in the sentence and simply combines the letters and numbers of the sentences.

4) Text to Sign-Language Conversion

Natural Language Processing, popularly known as NLP, has revolutionized the world of computation by providing a means through which computers can understand human language. When the JavaScript Web Speech API works in the first step, an output text is generated from the input audio signal. The text undergoes tokenization to differentiate each word which may include different forms of the word according to the rules of grammar but the root word remains the same as in grow and grow the same root word grows but in English used differently. Hence, stemming and lemmatization come into the picture. Both these techniques are generalization techniques for text or words and are essential for the initial processing of the text material to make it useful for further processing. Our proposed application uses the Python NLTK or Natural Language Toolkit package to implement these techniques. Stemming and lemmatization help us to remove the infected part of the text or document, leaving only the stem and root of the word. It helps us to get the text at this stage of working from the NLTK package.

5) *Avatarifying the Signs*

To Create the Avatar that displays the signs we've used Blender, which is free and open-source 3D Creation Suite. It Supports all 3D pipeline-modelling, manipulation, animation, simulation, rendering, composing and tracking motion, even video editing and game creation.

6) *Matching with Visual sign word Library*

Talking about the language translation systems available today, there are not even a handful of systems that work to include sign language in them. The reason for this lies in the asymmetric nature of languages in which sign language is a visual-spatial language using postures of different parts of the body such as the hands, arms, face, head and body. In addition, the grammar rules of almost all oral languages are standardized, resulting in the presence of inflection in them. But sign language includes no such rules and, therefore, to convert text into sign language the original has to be analysed using either pre-recorded videos or animations generated by computer avatars. It is difficult to visualize uniformity patterns in sign languages around the world and thus, the target of our model is American Sign Language. Although an exact number is not known, one estimate is between 500,000 and 15 million people. About 10% of the total population in the US suffers from hearing loss and most of these people have ASL as their first language. For each word/character from the processed text received after the second stage of the application, we perform a matching operation using tags in the Visual Sign Word library for the video in its sine database. Whenever a match is found, the matched video is retrieved from the SINE database and moved to the desired location.
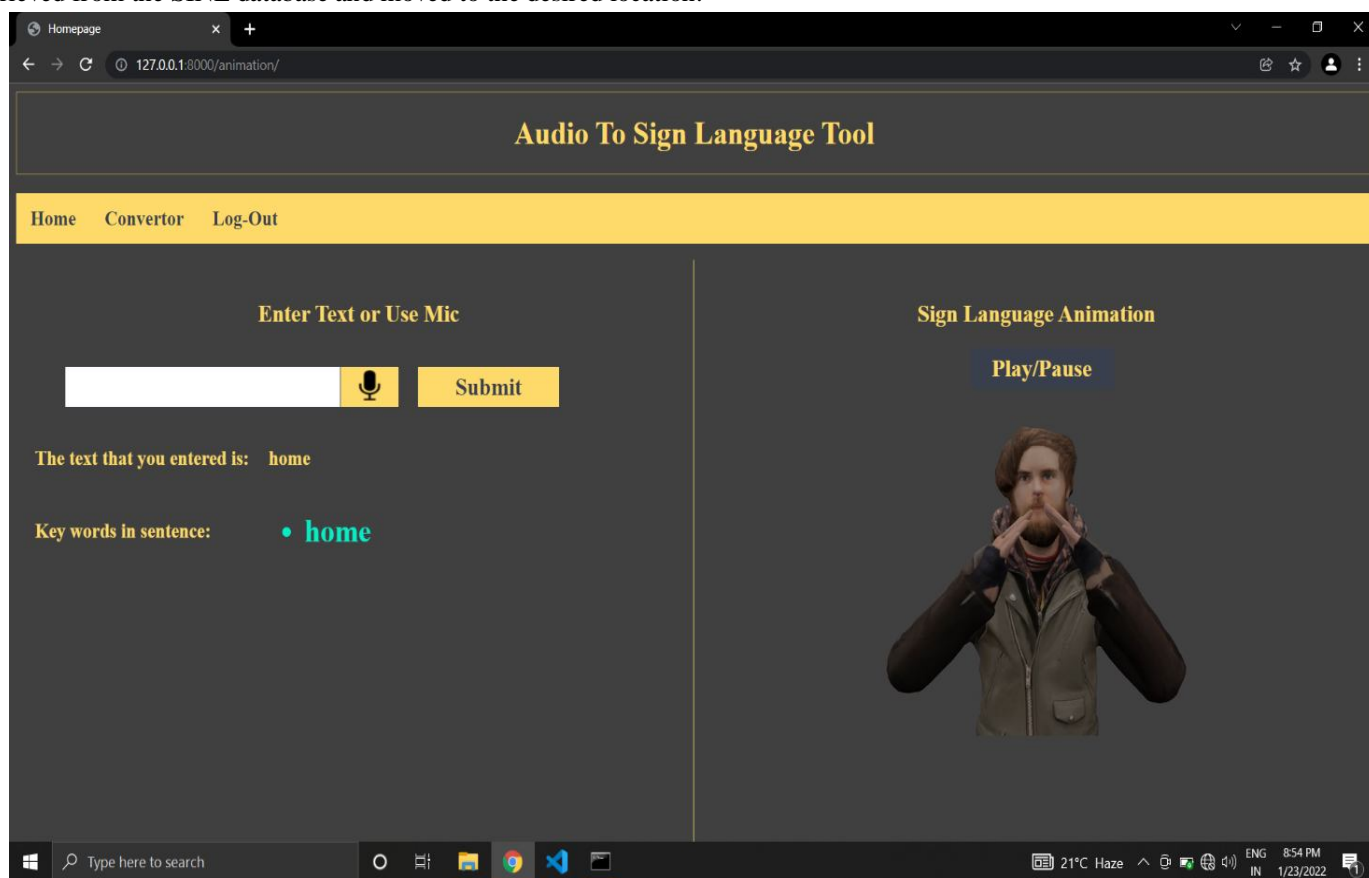

Fig. 2 Audio to Sign Language with Demo Input

7) *Creating Dataset for Real time Sign Language Recognition*

For this real time sign language recognition, we need to create dataset for every single alphabet or word which we want to get recognised. We used Open CV which is python library for recording the video of gestures after that breaking them into the frames then saving them into a folder with the alphabet name as label. For achieving the more accuracy, we are making 1000 images for every alphabet and words.

8) Training of Model

To train the dataset we are using TensorFlow which is an open-source library primarily developed for machine learning applications. After successful training of the dataset, the trained model file is saved into the project directory and accuracy can be seen exactly after training the module.
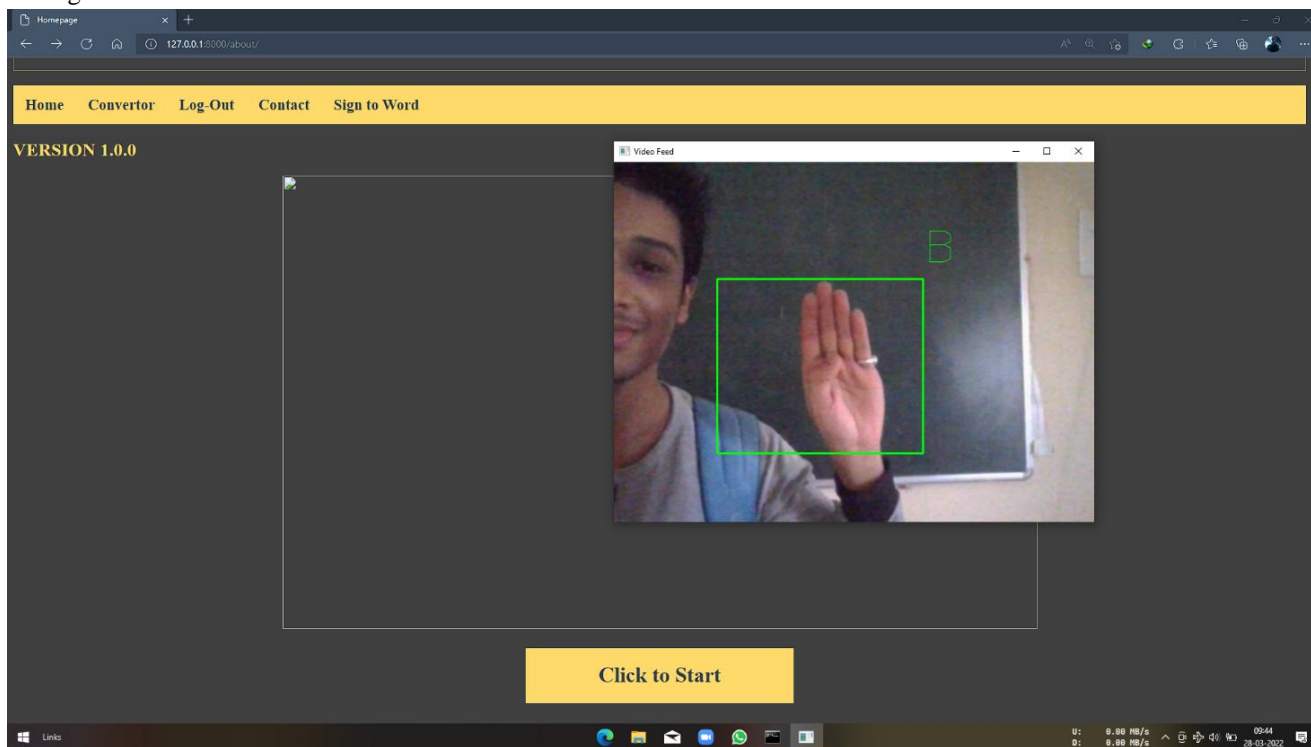

Fig. 3 Sign to Text Converter: Conversion Demo 1

IV. CONCLUSION

In this work, we propose a new posture-guided pooling strategy to extract additional features from 3D convolutional neural networks in the context of world-class sign language recognition. Our research shows that combining features from different levels of the network can improve the overall detection accuracy. As a future direction, we aim to consider phrase-level sign language modelling. We plan to use this function to localize code words in phrase level sign language recognition tasks.

V. ACKNOWLEDGMENT

REFERENCES

[1] Al Amin Hosain , Panneer Selvam Santhalingam, Parth Pathak, Huzefa Rangwala and Jana Koˇseck´a George Mason University, Fairfax, USA at "2021 IEEE Winter Conference on Applications of Computer Vision (WACV)", (3 Jan 2021).

[2] Y. J. Fan, ''Autoencoder node saliency: Selecting relevant latent representations,'' Pattern Recognition, (Apr. 2019).

[3] Muhammad Al-Qurishi, Thariq Khalidand Riad Souissi'sm "Deep Learning for Sign Language Recognition: Current Techniques, Benchmarks, and Open Issues" IEEEAccess (April 2021).

[4] Tasnim Ferdous Dima, MD. Eleas Ahmed's "Using YOLOv5 Algorithm to Detect and Recognize American Sign Language" at 2021 International Conference on Information Technology (ICIT), (01 September 2021).

[5] Soma Shrenika Prof. Myneni Madhu Bala Institute of Aeronautical Engineering's "SIGN LANGUAGE RECOGNITION USING TEMPLATE MATCHING TECHNIQUE" at 2020 International Conference on Computer Science, Engineering and Applications (ICCSEA March 2020).

[6] Necati Cihan Camg¨oz, Oscar Koller, Simon Hadfield and Richard Bowden's "Sign Language Transformers: Joint End-to-end Sign Language Recognition and Translation" at 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), (10 August 2020).

[7] Juan Zamora-Mora Escuela de Ingeniería del Software Universidad Cenfotec San José, Costa Rica, Mario Chacón-Rivas Escuela de Computación Instituto Tecnológico de Costa Rica Cartago, Costa Rica 's "2019 International- Conference on Inclusive Technologies and Education (CONTIE)", (4 July 2019).

[8] Lean Karlo S. Tolentino, Ronnie O. Serfa Juan, August C..Thio-ac, Maria Abigail B. Pamahoy, Joni Rose R. Forteza, and Xavier Jet O. Garcia's "Static Sign Language Recognition Using Deep Learning" at International Journal of Machine Learning and Computing, Vol. 9, No. (6, December 2019).

[9]   Muthu Mariappan H, Dr Gomathi V Department of Computer Science and Engineering National Engineering College Kovilpatti, Tamil Nadu, India's "Real-Time Recognition of Indian Sign Language" at Second International Conference on Computational Intelligence in Data Science (ICCIDS-2019).

[10]   Aditya Das1, Shantanu Gawde1, Khyati Suratwala1 and Dr. Dhananjay Kalbande's "SIGN LANGUAGE RECOGNITION USING DEEP LEARNING ON CUSTOM PROCESSED STATIC GESTURE IMAGES" at 2018 International Conference on Smart City and Emerging Technology (ICSCET Sept 2018).

[11]   H. Qu, T. Yuan, Z. Sheng, and Y. Zhang, "A pedestrian detection method based on YOLOv3 model and image enhanced by retinex," in 2018 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISPBMEI). (IEEE, 2018).

[12]   P. S. Zaki, M. M. William, B. K. Soliman, K. G. Alexsan, K. Khalil, and M. El-Moursy, "Traffic signs 5 detection and recognition system using deep learning,"(2020).

[13]   C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning. thirty-first aaai conf," Artif. Intell, (2017).

[14]   P. Rathi, R. Kuwar Gupta, S. Agarwal, and A. Shukla, "Sign language recognition using resnet50 deep neural network architecture," Available at SSRN 3545064, (2020).

[15]   S.-K. Ko, J. G. Son, and H. Jung, "Sign language recognition with recurrent neural network using human keypoint detection," in Proceedings of the 2018 Conference on Research in Adaptive and Convergent Systems, (2018).

[16]   P. T. Krishnan and P. Balasubramanian, "Detection of alphabets for machine translation of sign language using deep neural net," in 2019 International Conference on Data Science and Communication (IconDSC). (IEEE, 2019).

[17]   P. Liu, X. Li, H. Cui, S. Li, and Y. Yuan, "Hand gesture recognition based on single-shot multibox detector deep learning," Mobile Information Systems, vol. (2019).

[18]   S. Kim, Y. Ji, and K.-B. Lee, "An effective sign language learning with object detection based roi segmentation," in 2018 Second IEEE International Conference on Robotic Computing (IRC). (IEEE, 2018).

[19]   J. Redmon and A. Farhadi, "Yolo9000: better, faster, stronger," in Proceedings of the IEEE conference on computer vision and pattern recognition, (2017).

[20]   A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," arXiv preprint arXiv:2004.10934, (2020).

[21]   C.-Y. Wang, H.-Y. M. Liao, Y.-H. Wu, P.-Y. Chen, J.-W. Hsieh, and I.-H. Yeh, "Cspnet: A new backbone that can enhance learning capability of cnn," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops, (2020).

[22]   S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path aggregation network for instance segmentation," in Proceedings of the IEEE conference on computer vision and pattern recognition, (2018).

[23]   R. Xu, H. Lin, K. Lu, L. Cao, and Y. Liu, "A forest fire detection system based on ensemble learning," Forests, vol. 12, no. 2, p. 217, (2021).

[24]   I. Jindal, M. Nokleby, and X. Chen, ''Learning deep networks from noisy labels with dropout regularization,'' in Proc. IEEE 16th Int. Conf. Data Mining (ICDM), Barcelona, Spain, (Dec. 2016).

[25]   F. Ordóñez and D. Roggen, ''Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition,'' Sensors, vol. 16, no. 1, p. 115, (Jan. 2016).

[26]   I. Goodfellow, Y. Bengio, and A. Courville, ''Autoencoders,'' in Deep Learning, 1st ed. Cambridge, MA, USA: MIT Press, 2016, pp. 502–525.

[27]   M. Al-Hammadi, G. Muhammad, W. Abdul, M. Alsulaiman, and M. S. Hossain, ''Hand gesture recognition using 3D-CNN model,'' IEEE Consum. Electron. Mag., vol. 9, no. 1, pp. 95–101, (Jan. 2020).

[28]   G. Muhammad, M. F. Alhamid, M. Alsulaiman, and B. Gupta, ''Edge computing with cloud for voice disorder assessment and treatment,'' IEEE Commun. Mag., vol. 56, no. 4, pp. 60–65, (Apr. 2018).

[29]   Manasa Srinivasa H S and Suresha H S, "Implementation of Real Time Hand Gesture Recognition," International Journal of Innovative Research in Computer and Communication Engineering, Vol. 3, Issue 5, (May 2015).

[30]   Archana S. Ghotkar and Gajanan K. Kharate, "Dynamic Hand Gesture Recognition and Novel Sentence Interpretation Algorithm for Indian Sign Language Using Microsoft Kinect Sensor," Journal of Pattern Recognition Research 1 (2015) 28-38.

INTERNATIONAL JOURNAL
FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 ◯ (24*7 Support on Whatsapp)