Multi-culture Sign Language Detection and Recognition Using Fine-tuned Convolutional Neural Network

K VIGNESHWAR
Assistant Professor, Guru Nanak Institute of Technology, CSE Department, Hyderabad

ABSTRACT: The speech and hearing-impaired community use sign language as the primary means of communication. It is quite challenging for the general population to interpret or learn sign language completely. A sign language recognition system must be designed and developed to address this communication barrier. Most current sign language recognition systems rely on wearable sensors, keeping the recognition system unaffordable for most individuals. Moreover, the existing vision-based sign recognition frameworks do not consider all of the spatial and temporal information required for accurate recognition. A novel vision-based hybrid deep neural net methodology is proposed in this project for recognizing American Sign Language (ASL) and custom sign gestures. The proposed framework aims to establish a single framework for tracking and extracting multi-semantic properties, such as non-manual components and manual co-articulations. Furthermore, spatial feature extraction from the sign gestures is deployed using a Hybrid Deep Neural Network (HDNN) with atrous convolutions. The temporal and sequential feature extraction is carried out by employing attention-based HDNN. In addition, the distinguished abstract feature extraction is done using modified autoencoders. The discriminative feature extraction for differentiating the sign gestures from unwanted transition gestures is done by leveraging the hybrid attention module. The experimentation of the proposed model has been carried out on the novel multi-signer ASL and custom sign language dataset. The proposed sign language recognition framework with hybrid neural nets, specifically using HDNN, yields better results than other state-of-the-art frameworks. Additionally, a detection module is incorporated using Flask web, allowing manual input of images/signs for real-time recognition.

I. INRODUCTION:

Sign Language (SL) is the basic means of interaction for the speech-impaired and hard-of-hearing populace. Like any other language, Sign Language seems to have its under-lying structure and grammatical rules that allow users to communicate and express themselves adequately. Further-more, the SL is usually expressed through manual components such as Hand motion, Hand position and non-manual articulations such as eye gaze, facial expression, lip movement, etc. The manual and non-manual components together form the multi-semantic feature components. Mastering an SL requires substantial effort for the hearing community, which demands developing a Sign Language Recognition system (SLR). Recently developed Sign Language datasets include American, Arabic, German, Chinese, Turkish, Bhutanese, Russian, and Indian Sign languages (ISL). SLR has been intensively investigated to help hearing people comprehend sign language and make the everyday lives of the speech-impaired and hard-of-hearing community more convenient. The SLR frameworks aim to detect and recognize sign language performed by the sign interpreter from a visual medium. Various concerns in developing an SLR include signer-dependent variations, local and global elements, feature extrication from heterogeneous backgrounds, large vocabulary and scalability, multi-modality, occlusion, and movement epenthesis. Although countless research on SLR has been undertaken, most issues remain unresolved. Most of the developed SLRs leverage wearable sensors, color-coded gloves, or multiple depth sensor cameras to capture the data, making the signer very uncomfortable conveying the sign gestures in real-life scenarios.

II. LITERATURE SURVEY

S. Sharma and S. Singh discussed that an efficient sign language recognition system (SLRS) can recognize the gestures of sign language to ease the communication between the signer and non-signer community. In this project, a computer-vision based SLRS using a deep learning technique has been proposed. This project has primary three contributions: first, a large dataset of Indian sign language (ISL) has been created using 65 different users in an uncontrolled environment. Second, the intra-class variance in dataset has been increased using augmentation to improve the generalization ability of the proposed work. Three additional copies for each training image are generated in this project, by using three different affine transformations. Third, a novel and robust model using Convolutional Neural Network (CNN) have been proposed for the feature extraction and classification of ISL gestures. The performance of this method is evaluated on a self-collected ISL dataset and publicly available dataset of ASL. For this total of three datasets have been used and the achieved accuracy is 92.43, 88.01, and 99.52%. The efficiency of this method has been also evaluated in terms of precision, recall, f-score, and time consumed by the system. The results indicate that the proposed method shows encouraging performance compared with existing work.

A. Tunga, S. V. Nuthalapati, and J. Wachs explained that sign language recognition (SLR) plays a crucial role in bridging the communication gap between the hearing and vocally impaired community and the rest of the society. Word-level sign language recognition (WSLR) is the first important step towards understanding and interpreting sign language. However, recognizing signs from videos is a challenging task as the meaning of a word depends on a combination of subtle body motions, hand configurations, and other movements. Recent pose-based architectures for WSLR either model both the spatial and temporal dependencies among the poses in different frames simultaneously or only model the temporal information without fully utilizing the spatial information.

S. Jiang, B. Sun, L. Wang, Y. Bai, K. Li, and Y. Fu, discueesd in their paper sign language is commonly used by deaf or speech impaired people to communicate but requires significant effort to master. Sign Language Recognition (SLR) aims to bridge the gap between sign language users and others by recognizing signs from given videos. It is an essential yet challenging task since sign language is performed with the fast and complex movement of hand gestures, body posture, and even facial expressions. Recently, skeleton-based action recognition attracts increasing attention due to the independence between the subject and background variation. However, skeleton-based SLR is still under exploration due to the lack of annotations on hand keypoints. Some efforts have been made to use hand detectors with pose estimators to extract hand key points and learn to recognize sign language via Neural Networks, but none of them outperforms RGB-based methods. To this end, we propose a novel Skeleton Aware Multi-modal SLR framework (SAM-SLR) to take advantage of multi-modal information towards a higher recognition rate. Specifically, we propose a Sign Language Graph Convolution Network (SL-GCN) to model the embedded dynamics and a novel Separable Spatial-Temporal Convolution Network (SSTCN) to exploit skeleton features. RGB and depth modalities are also incorporated and assembled into our framework to provide global information that is complementary to the skeleton-based methods SL-GCN and SSTCN. As a result, SAM-SLR achieves the highest performance in both RGB (98.42%) and RGB-D (98.53%) tracks in 2021 Looking at People Large Scale Signer Independent Isolated SLR Challenge.

Y. Saleh and G. F. Issa explained sign Language is considered the main communication tool for deaf or hearing impaired people. It is a visual language that uses hands and other parts of the body to provide people who are in need to full access of communication with the world. Accordingly, the automation of sign language recognition has become one of the important applications in the areas of Artificial Intelligence and Machine learning. Specifically speaking, Arabic sign language recognition has been studied and applied using various intelligent and traditional approaches, but with few attempts to improve the process using deep learning networks. This paper utilizes transfer learning and fine tuning deep convolutional neural networks (CNN) to improve the accuracy of recognizing 32 hand gestures from the Arabic sign language. The proposed methodology works by creating models matching the VGG16 and the ResNet152 structures, then, the pre-trained model weights are loaded into the layers of each network, and finally, our own soft-max classification layer is added as the final layer after the last fully connected layer. The networks were fed with normal 2D images of the different Arabic Sign Language data, and was able to provide accuracy of nearly 99%.

III. METHODOLOGY

The study of this paper is to develop a Sign Language Recognition (SLR) system that can correctly understand both bespoke sign gestures and American Sign Language (ASL) at a reasonable cost. We aim to improve recognition precision by optimising feature extraction using HDNN with atrous convolutions, attention-based techniques, and improved autoencoders. Furthermore, the project's Flask web-based detection module will enable real-time identification, enabling manual input for more widely accessible and useful applications.

DISADVANTAGES OF EXISTING SYSTEM:

- This might be problematic when working with dynamic components of sign language when a thorough comprehension of temporal patterns is necessary.
- May find it difficult to generalise, which might result in decreased performance if it doesn't have enough labelled training data.
- They might not be as suitable for managing data that is sequential.

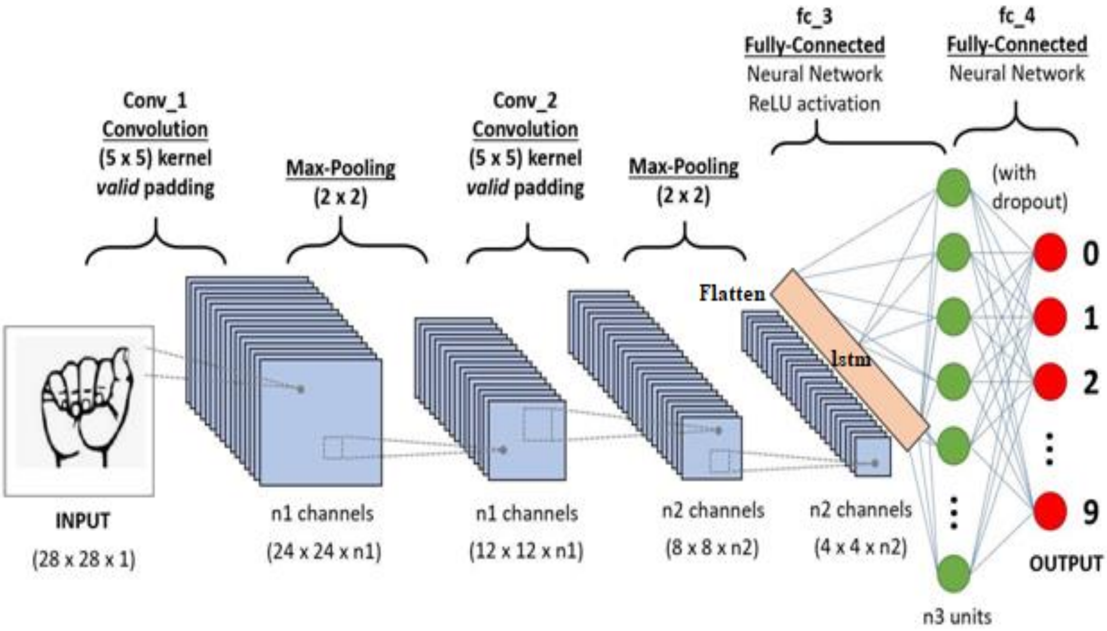
PROPOSED SYSTEM

The suggested system presents a novel method for recognising sign language, tackling the difficulties encountered by the community of people with speech and hearing impairments. Our approach is to offer the general public an accessible and cost-effective means of learning and interpreting sign language.Unlike current systems that depend on wearable sensors, our method makes use of vision-based technology and improves both spatial and temporal information to achieve accurate recognition. The system also includes an easy-to-use Flask web interface for manual picture entry and real-time sign language recognition.

ADVANTAGES OF PROPOSED SYSTEM:

- Awareness of complicated patterns and linkages within multi-sign gestures.
- This adaptability helps to increase accuracy, particularly when it comes to capturing the lively motions and co-articulations that are characteristic of sign language. Identifying indications given by diverse people and in a range of environments, strengthening and broadening the system's applicability.

SYSTEM ARCHITECTURE:



MODULES:

Dataset: The system's Multi-Cultural Sign Language Detection and Recognition functionality is intended to operate with a customized dataset. The collection contains extra signs like 'del' (delete), 'nothing', and 'space' in addition to signs that correspond to the English alphabet from A to Z. Additionally, there are signs for words and phrases like "work," "today," "time," "thank you," "rain," "practice," "love," "internet," "home," "help," "hello," "happy," "game," "friends," "food," "family," "delete," "chocolate," "brave," and "beautiful." There are a total of 20 classes for the Custom dataset and 29 classes for ASL.

Importing Necessary Libraries: The libraries required for model construction and training are imported, and Python is the selected programming language. These libraries include conventional libraries like pandas, numpy, matplotlib, and tensorflow, as well as Keras for model construction, scikit-learn for data partitioning, and PIL for image processing.

Retrieving and Preprocessing Images: Images are taken from the designated folder (train_folder) together with the labels that go with them. Preprocessing operations performed on the photos include edge detection, scaling to (64, 64), and numpy array conversion.

Building the HDNN Model: Using Keras, a Hierarchical Deep Neural Network (HDNN) model is built. The model architecture is composed of dense layers, recurrent layers (LSTM), flatten layers, and convolutional layers (Conv2D) with max-pooling. The last dense layer generates 29 nodes corresponding to the 29 classes. Another HDNN of the final dense layer utilises the softmax activation function for multiclass classification and produces 20 nodes matching to the 20 classes.

Training the Model: Categorical cross-entropy loss and the Adam optimizer are used to construct the model. The fit function is then used to train it on the prepared dataset. Accuracy and loss graphs are shown, and the training process is tracked.

Saving the Trained Model: The tensorflow Keras model.save() method is used to save the trained HDNN model into an HDF5 file called "HDNNsign.h5." Deploying the model in settings prepared for production requires completing this step. Furthermore, the pickle library is mentioned as a tool for storing the model as a.H5 file; however, the given code lacks the real implementation.

IV. IMPLEMENTATION

Convolutional Neural Networks (CNNs):

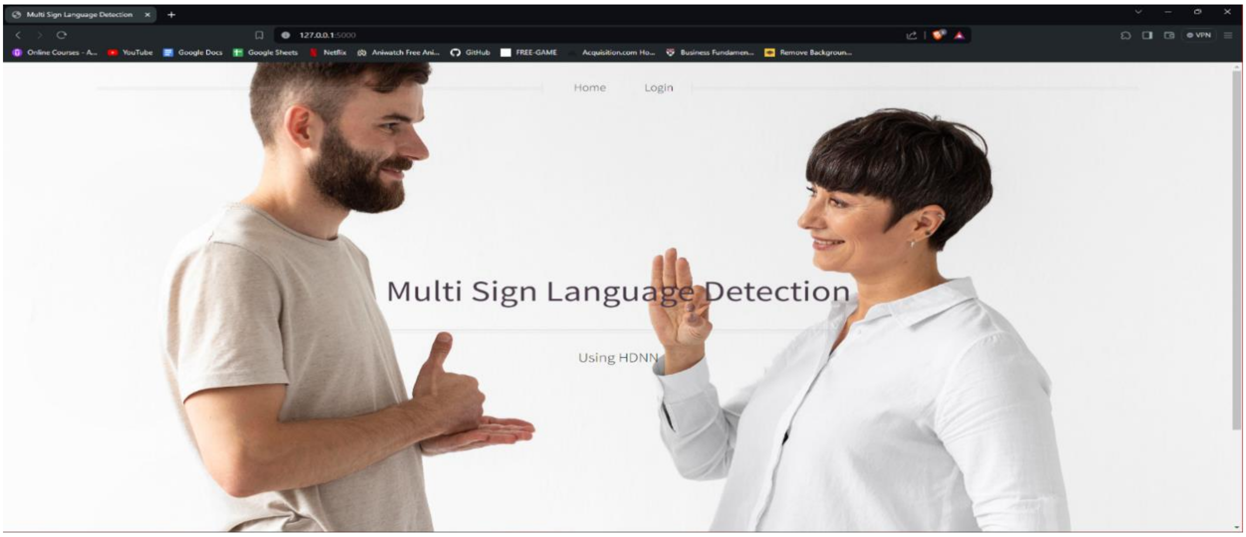
The existing technique utilizes a conventional approach to sign language recognition, with a focus on a dataset of Indian sign language (ISL). The dataset, although sizable, lacks diversity and is collected in a controlled environment, limiting the system's generalization ability to real-world, uncontrolled scenarios. Augmentation is applied with three basic affine transformations, but it does not extensively enhance intra-class variance.The feature extraction and classification of ISL gestures rely on methods that are less robust compared to modern deep learning techniques. Specifically, the model uses conventional methods instead of advanced neural networks like Convolutional Neural Networks (CNNs), resulting in lower accuracy and efficiency. The evaluation metrics used are relatively basic, and the overall performance of the system is constrained by its reliance on the usage of less sophistication.

Hybrid Deep Neural Network (HDNN) based Technique:

A unique vision-based hybrid deep neural network is used in the suggested approach to enable robust sign language recognition. Both bespoke sign gestures and American Sign Language (ASL) are recognised by this system. A Hybrid Deep Neural Network (HDNN) with atrous convolutions improves spatial information extraction and offers a thorough comprehension of the sign gestures. Unlike conventional LSTM, our method extracts temporal and sequential features efficiently by using attention-based HDNN. Modified autoencoders are used to extract distinguished abstract features. In order to extract discriminative features and separate undesirable transition movements from sign gestures, the hybrid attention module is implemented. Our suggested approach outperforms previous state-of-the-art frameworks, as demonstrated by the testing on a multi-signer ASL and custom sign language dataset. By using Flask web, real-time sign language identification with human input is made possible, improving the system's usability and accessibility.

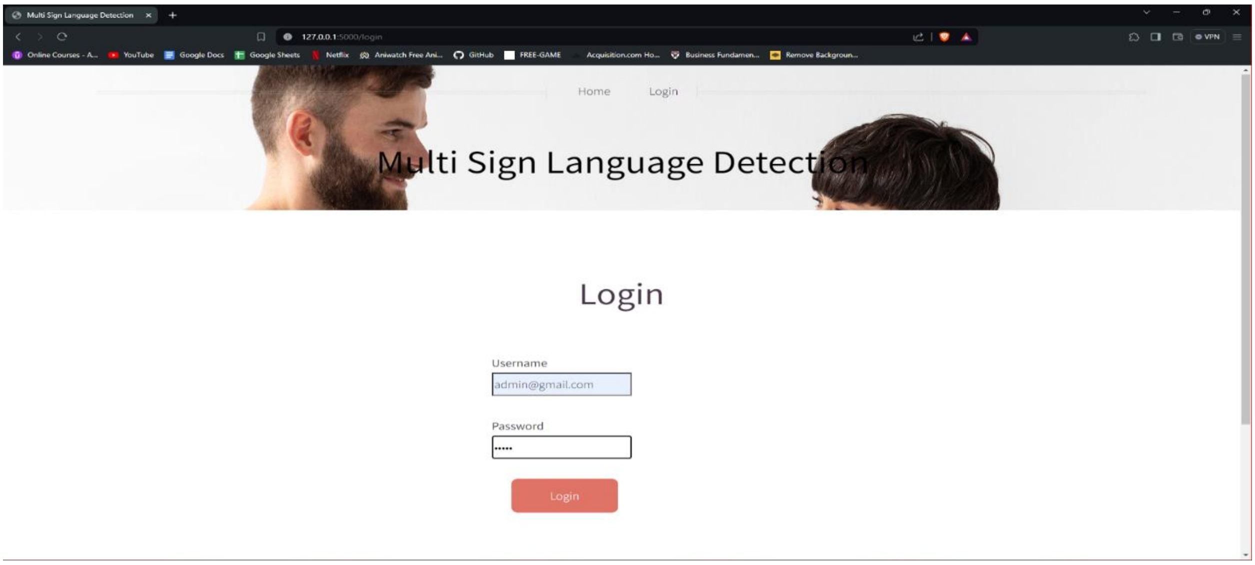
V.EXPERIMENTAL RESULTS

HOME PAGE



EXPLAINATION: Upon executing the program and pasting the web address into the browser, the homepage is the initial loaded page. It serves as the primary point of interaction with the website or web application, indicating the initiation of browsing activities.

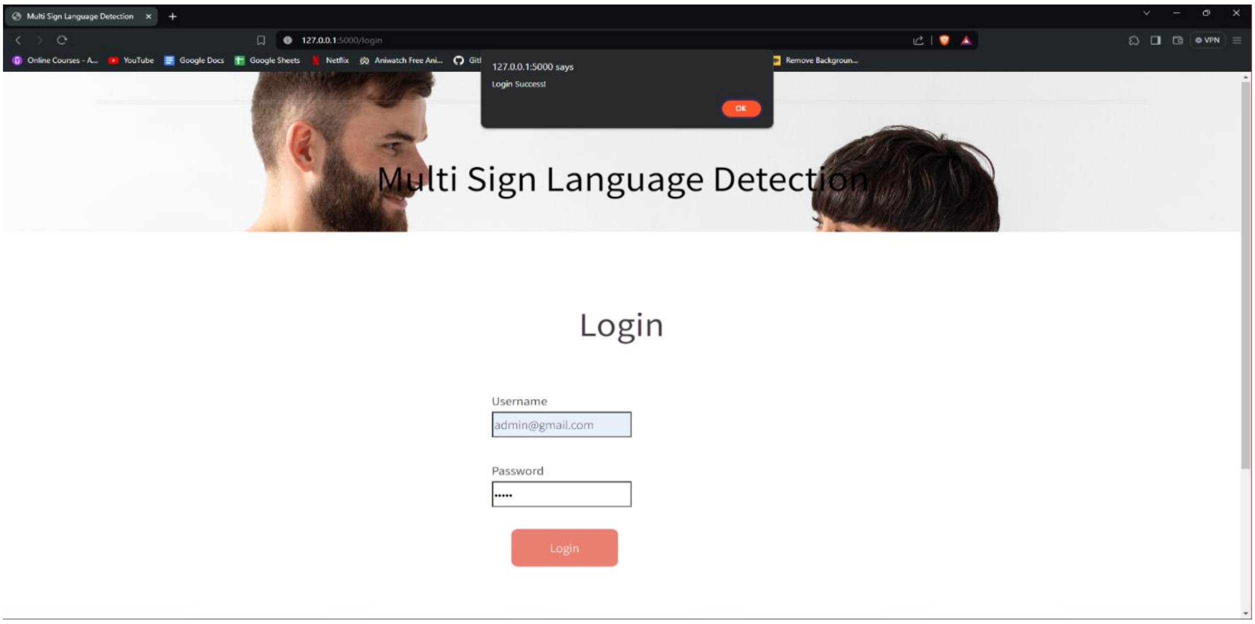
USER LOGIN PAGE



Username: admin@gmail.com

Password: •••••

EXPLAINATION: Following the click of the login button, the Login page becomes visible, prompting the entry of details. This sequence illustrates the transition from initiating the login process to providing necessary information for authentication. It underscores the procedural nature of user interaction in accessing secured areas of a website or application.



127.0.0.1:5000 says
Login Success!
OK

Multi Sign Language Detection

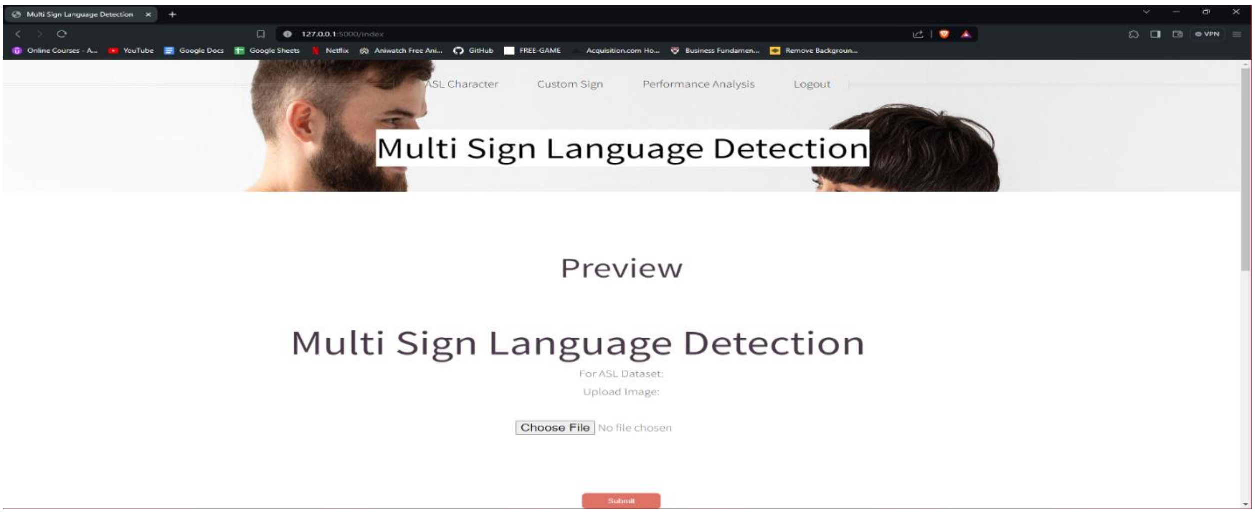
Login

Username
admin@gmail.com

Password
•••••

Login

SIGN UPLOAD PAGE



ASL Character Custom Sign Performance Analysis Logout

Multi Sign Language Detection

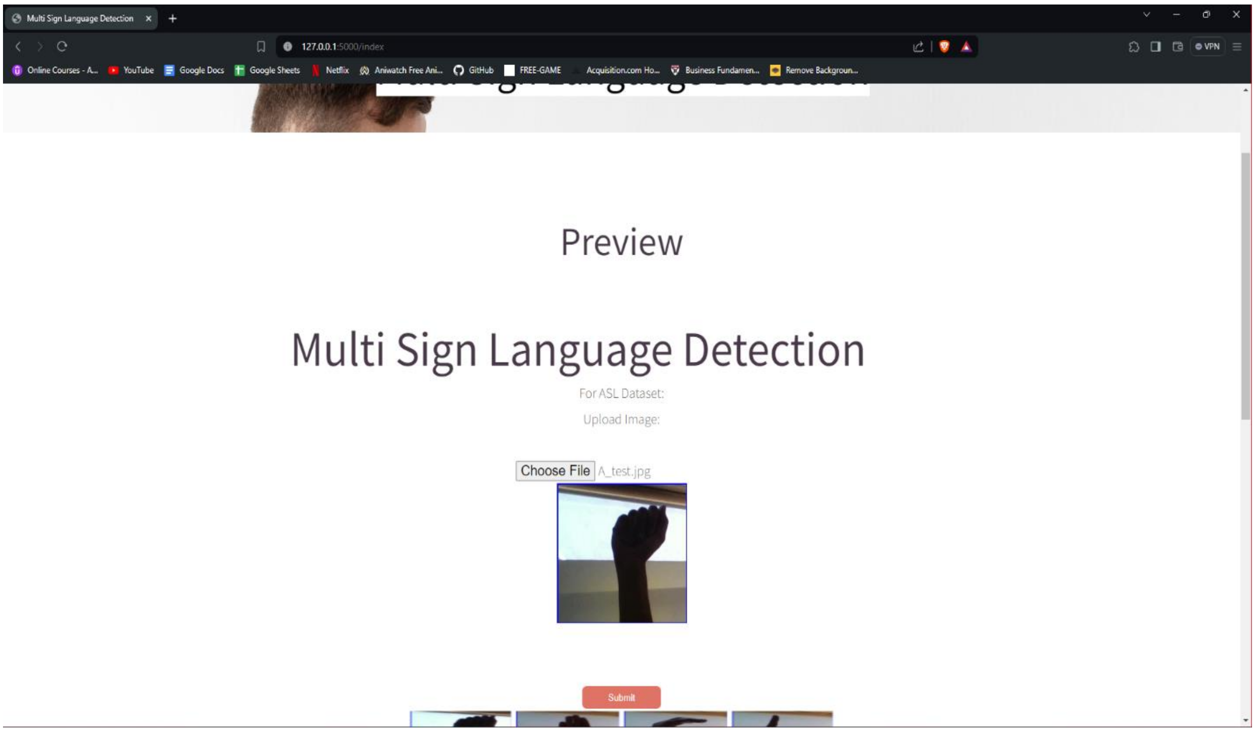
Preview

Multi Sign Language Detection
For ASL Dataset:
Upload Image:

Choose File No file chosen

Submit

EXPLAINATION: Upon successful login, a page allowing selection of the ASL image is presented. The user then chooses the desired image before submitting it. This process illustrates the post-login navigation flow, emphasizing user interaction in selecting and submitting the appropriate image for further action.



Preview

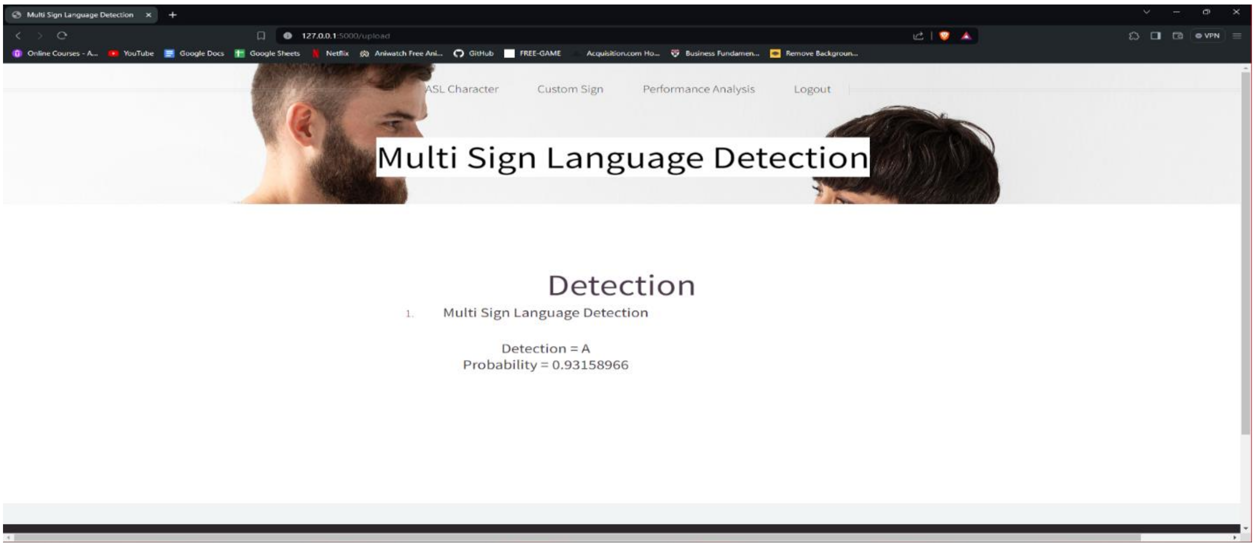
Multi Sign Language Detection

For ASL Dataset:

Upload Image:

A_test.jpg

FINAL RESULT DETECTION PAGE



Detection

1. Multi Sign Language Detection

Detection = A
Probability = 0.93158966

EXPLAINATION: Following detection, the system provides the outcome along with the probability of accuracy. This step highlights the informative feedback loop, offering users insights into the reliability of the detection process.

VI. CONCLUSION

A novel Sign Language Recognition system has been developed by creating HDNN architecture. The proposed HDNN framework intends to extract semantic manual co-articulations and non-manual elements, which are the key components necessary for sign recognition. In addition, spatial, sequential, and temporal features are also considered for accurate recognition. Furthermore, abstract and discriminative feature extraction is also carried out to segregate genuine and non-useful gestures. The genuine gestures are then used for recognizing the Sign gesture representations, thereby reducing the computation overhead. The experimentation of the proposed HDNN framework has been deployed on the newly created Custom data and ASL Dataset. The results generated represent a good and efficient performance.

FUTURE ENHANCEMENT

As a part of future work, we would like to extend our study toward continuous sign sentence recognition. We would also like to design a framework for handling the segmentation ambiguities and moment epenthesis in continuous sign language recognition. The isolated word gesture recognition framework could also be integrated to enhance sign spotting from continuous sign video stream for recognition sign sentences from continuous sign gestures. We also intend to increase the dataset and publicly publish it for further research.

REFERENCES

[1] Y. Saleh and G. F. Issa, ‘‘Arabic sign language recognition through deep neural networks fine-tuning,’’ Int. J. Online Biomed. Eng., vol. 16, no. 5, pp. 71–83, 2020.

[2] X. Jiang, M. Lu, and S.-H. Wang, ‘‘An eight-layer convolutional neural network with stochastic pooling, batch normalization and dropout for fingerspelling recognition of Chinese sign language,’’ Multimedia Tools Appl., vol. 79, nos. 21–22, pp. 15697–15715, Jun. 2020.

[3] O. Sevli and N. Kemaloğlu, ‘‘Turkish sign language digits classification with CNN using different optimizers,’’ Int. Adv. Researches Eng. J., vol. 4, no. 3, pp. 200–207, Dec. 2020.

[4] K. Wangchuk, K. Wangchuk, and P. Riyamongkol, ‘‘Bhutanese sign language hand-shaped alphabets and digits detection and recognition,’’ Ph.D. dissertation, Dept. Comput. Eng., Naresuan Univ., Phitsanulok, Thailand, 2020.

[5] R. Elakkiya and E. Rajalakshmi. ISLAN. Mendeley Data. Accessed: Jan. 8, 2021. [Online]. Available: https://data.mendeley.com/datasets/rc349 j45m5/1

[6] W. Sandler, ‘‘The phonological organization of sign languages,’’ Lang. Linguistics Compass, vol. 6, no. 3, pp. 162–182, Mar. 2012.

[7] R. Elakkiya, ''Retraction note to: Machine learning based sign language recognition: A review and its research frontier,'' J. Ambient Intell. Human-ized Comput., vol. 12, no. 7, pp. 7205–7224, Jul. 2022.

[8] S. Diwakar and A. Basu, ''A multilingual multimedia Indian sign language dictionary tool,'' in Proc. IJCNLP, 2008, p. 57.

[9] S. K. Liddell and R. E. Johnson, ''American sign language: The phonological base,'' Sign Lang. Stud., vol. 1064, no. 1, pp. 195–277, 1989.

[10] P. Eccarius and D. Brentari, ''Symmetry and dominance: A cross-linguistic study of signs and classifier constructions,'' Lingua, vol. 117, no. 7, pp. 1169–1201, Jul. 2007.

[11] Z. Ren, J. Yuan, J. Meng, and Z. Zhang, ''Robust part-based hand gesture recognition using Kinect sensor,'' IEEE Trans. Multimedia, vol. 15, no. 5, pp. 1110–1120, Aug. 2013.

[12] C. Wang, Z. Liu, and S.-C. Chan, ''Superpixel-based hand gesture recog-nition with Kinect depth camera,'' IEEE Trans. Multimedia, vol. 17, no. 1, pp. 29–39, Jan. 2015.

[13] Z. Liu, X. Chai, Z. Liu, and X. Chen, ''Continuous gesture recognition with hand-oriented spatiotemporal feature,'' in Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCVW), Oct. 2017, pp. 3056–3064.

[14] P. Molchanov, X. Yang, S. Gupta, K. Kim, S. Tyree, and J. Kautz, ''Online detection and classification of dynamic hand gestures with recurrent 3D convolutional neural networks,'' in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 4207–4215.

[15] R. Agarwal, ''Bayesian K-nearest neighbour based redundancy removal and hand gesture recognition in isolated Indian sign language without materials support,'' IOP Conf. Ser., Mater. Sci. Eng., vol. 1116, no. 1, 2021, Art. no. 012126.

[16]O. Aran, ''SignTutor: An interactive system for sign language tutoring,'' IEEE MultiMedia vol. 16, no. 1, pp. 81–93, Mar. 2009.