International Journal of Engineering Research and Sustainable Technologies
Volume 2, No.4, Dec 2024, P 26 -33
ISSN: 2584-1394 (Online version)

A REAL TIME INDIAN SIGN LANGUAGE RECOGNITION USING TENSORFLOW

[1]Sanjusaran K, [2]Shakthipriyan S, [3]Supreetraju RU
[1,2,3]Department of CSE
[1,2,3] Dr.MGR Educational and Research Institute, Chennai, India
*** Corresponding author email address**: sanjusarankmgr123@gmail.com

DOI: https://doi.org/10.63458/ijerst.v2i4.98 | ARK: https://n2t.net/ark:/61909/IJERST.v2i4.98

Abstract

Communication is the exchange of information, ideas, or emotions, typically through spoken or written language. However, for individuals who are deaf or mute, traditional communication methods may not be effective. Instead, they rely on sign language—a visual form of communication using gestures and movements. Unfortunately, many people are unfamiliar with sign language, creating a barrier between those who use it and those who do not. Machine learning offers a promising solution to this challenge. By training a model to recognize and translate sign language gestures into spoken or written language, we can bridge this communication gap. This study proposes a real-time Sign Language Recognition (SLR) system using transfer learning with TensorFlow. Our approach involves capturing Indian Sign Language (ISL) gestures via a webcam and continuously training a deep learning model for accurate, real-time recognition. To enhance usability, we integrate a text-to-audio translation feature, converting recognized gestures into spoken language. This functionality allows individuals proficient in sign language to communicate seamlessly with those who are not, fostering inclusivity and accessibility. By leveraging machine learning, our system aims to break communication barriers and create a more inclusive society.

Keywords: Sign Language Recognition, Machine Learning, Transfer Learning, Tensor Flow, Real-time Systems, Communication Barrier, Deaf Communication, Dumb Communication, Indian Sign Language, Gesture Recognition, Text-to-Audio Translation, Accessibility, Assistive Technology, Communication Aid, Human-Computer Interaction.

1. Introduction

Communication is the foundation of human interaction, enabling the exchange of information between individuals. Effective communication occurs when the intended message is successfully transmitted and understood by the recipient. It takes various forms, including formal and informal exchanges, oral and written communication, non-verbal cues, feedback mechanisms, and visual aids.

Formal communication follows structured channels, while informal (grapevine) communication happens spontaneously in social and professional settings. Oral communication involves spoken interactions, either in person or via technology such as voice and video calls. Written communication includes letters, emails, and notices, while non-verbal communication relies on gestures, facial expressions, and body language. Additionally, feedback and visual communication enhance message clarity, and active listening ensures better comprehension.

For individuals with hearing impairments, non-verbal communication—particularly sign language—serves as a crucial alternative to spoken communication. However, a lack of widespread sign language proficiency among the general population creates barriers to interaction, limiting accessibility and inclusivity. Addressing this challenge requires innovative solutions that bridge the communication gap between individuals with and without hearing impairments.

2. Related work

Sign Language (SL) enables individuals with hearing impairments to communicate through structured hand gestures, facial expressions, and body movements. With over 300 distinct sign languages worldwide, fluency in these languages remains limited, creating communication barriers. To bridge this gap, technology-driven solutions are essential.

This study proposes a real-time Sign Language Recognition (SLR) system using the TensorFlow Object Detection API, integrated with web and mobile applications for real-time text-to-audio conversion of Indian Sign Language (ISL). Additionally, blind individuals can benefit from this system by using earphones connected to a mobile application that captures ISL gestures via a smartphone camera and converts them into speech, significantly improving accessibility.

The subsequent sections explore related work in SLR, data acquisition and processing, system methodology, experimental evaluation, and future research directions. The list of papers reviewed through Table 1.

Table 1. List the Survey Papers with Author Details

S.No	Title	Authors	Publication Year
1	The Types of Communication	Kapur, R	2020
2	Sign Language Recognition Application Systems for Deaf-Mute People: A Review Based on Input Process-Output	Suharjito, Anderson, R., Wiryana, F., Ariesta, M.C., Kusuma, G.P	2017
3	Sign language recognition based onhandandbodyskeletaldata.3DTV-Conference	Konstantinidis, D., Dimitropoulos, K., Daras, P	2018
4	Machine Learning Techniques for Indian Sign Language Recognition. Int. Conf. Curr. Trends Comput	Dutta, K.K., Bellary, S.A.S.	2018
5	Sign Language Recognition, Generation, and Translation: An Interdisciplinary Perspective.	Bragg,D.,Koller,O.,Bellard,M.,Berke,L., Boudreault, P., Braffort, A., Caselli, N., Huenerfauth, M., Kacorri, H., Verhoef, T., Vogler, C., Morris, M.R	2019
6	Sign Language Recognition Based on Intelligent Glove Using Machine Learning Techniques	Rosero-Montalvo, P.D., Godoy-Trujillo, P., Flores-Bosmediano, E., Carrascal-Garcia,J.,Otero-Potosi, S., Benitez-Pereira, H., Peluffo-Ordonez, D.H	2018
7	Recent Advances of Deep Learning for Sign Language Recognition. DICTA	Zheng, L., Liang, B., Jiang, A	2017
8	A Real Time Hand Tracking System for Interactive Applications	Rautaray, S.S	2011
9	Hand tracking algorithm based on super-pixels feature	Zhang, Z., Huang, F	2014
10	A feature covariance matrix with serial particle filter for isolated sign language recognition	Lim, K.M., Tan, A.W.C.,Tan, S.C	2018
11	Block-based histogram of optical flow for isolated sign language recognition.	Lim,K.M.,Tan,A.W.C.,Tan,S.C	2016
12	Hidden Markov Model-Based gesture recognition with overlapping hand-head/hand-hand estimated using Kalman Filter	Gaus,Y.F.A.,Wong, F.	2012

S.No	Title	Authors	Publication Year
13	Sign language recognition using image based hand gesture recognition techniques	Nikam, A.S., Ambekar, A.G	2016
14	Arabic sign language recognition using the leap motion controller	Mohandes,M.,Aliyu,S.,Deriche,M	2014
15	Sign language recognition through Leap Motion controller and input prediction algorithm	Enikeev, D.G., Mustafina, S.A	2021
16	Sign language transformers: Joint end-to-end sign language recognition and translation	Camgöz,N.C.,Koller,O.,Hadfield,S., Bowden,R	2020
17	American Sign Language Recognition using Deep Learning and Computer Vision	Bantupalli, K., Xie, Y	2019

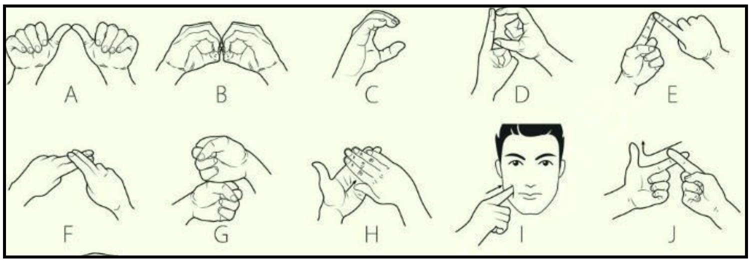

Fig. 1. Indian Sign Language

impairments to engage freely with others. Sign Language Recognition (SLR) systems offer a solution by enabling communication in sign language without requiring fluency. These systems recognize and translate sign language gestures into commonly spoken languages, such as English, facilitating communication between individuals with and without hearing impairments. Extensive [3] research has been conducted in the field of SLR, yet several aspects remain to be addressed. [6] Machine learning techniques play a pivotal role in SLR systems, allowing electronic systems to make decisions based on experiential data. Classification algorithms rely on training and testing datasets, with numerous authors having developed efficient data acquisition and classification methods. [8] Previous work in SLR can be categorized into two approaches based on data acquisition: direct measurement methods and vision-based approaches. Direct measurement methods involve devices such as motion data gloves, motion capturing systems, or sensors, while vision-based approaches rely on RGB image processing to extract spatial and temporal features. Hand detection is a critical aspect of vision-based SLR, often achieved through semantic segmentation and skin color detection. However, challenges such as mistaken recognition of other body parts as hands necessitate advanced techniques like face detection and background subtraction for accurate hand tracking. Various processing methods have been employed in SLR systems, including Hidden Markov Models (HMM), neural networks, [13] Naïve Bayes Classifier (NBC), and Support Vector Machines (SVM). These methods yield different accuracy results, influenced by factors such as dataset size, image clarity, and data acquisition methods. SLR systems can be categorized into isolated and continuous SLR, with the former recognizing single gestures and the latter translating entire sentences. Despite advancements, challenges persist, including the laborious labeling process for isolated SLR, the high cost of devices for data acquisition, and inaccuracies introduced by vision-based methodologies. In this paper, a real-time SLR system is being developed, utilizing a dataset created using Python and OpenCV with a webcam. The system aims to address challenges in SLR, including the need for cost-effective data acquisition methods and the development of accurate and efficient processing techniques. Additionally, an extra feature of real-time text-to-audio conversion in Indian Sign Language Figure 1, will be integrated to enhance accessibility and usability for individuals with hearing impairments.

3. Data Acquisition

A real-time system for detecting Indian Sign Language (Figure 3) gestures is currently under development. To gather data, images are captured using a webcam with Python and YOLOv8. YOLOv8 offers a rich array of functions tailored for real-time computer vision tasks, providing a solid foundation for applications in machine perception. The development journey begins with the utilization of a webcam orchestrated by Python and guided by YOLOv8's specialized functionalities, purposefully tailored for real-time computer vision tasks. Through this amalgamation, images are captured with precision, capturing the nuanced intricacies of ISL gestures, laying the foundational groundwork for subsequent stages. The dataset is meticulously organized into distinct folders, each designated to house images portraying a specific ISL word. This systematic arrangement not only facilitates ease of access but also ensures a structured approach throughout the data processing pipeline. An integral aspect of this pipeline involves the meticulous labeling of each captured image, a task streamlined by the utilization of the Label Ing package. Figure 2.

Fig 2. Labeled Image Folders

This labeling process is paramount, as it furnishes each image with precise annotations delineating the intricate hand gestures it portrays, thereby providing indispensable ground truth data crucial for subsequent model training endeavors. Following the labeling phase, XML files are automatically generated for each annotated image, encapsulating critical meta data pivotal for subsequent stages of model development. The dataset then undergoes a pivotal phase of partitioning, where it is divided into distinct training and validation subsets, maintaining a balanced 80:20 ratio. This strategic allocation ensures that the model is trained on a diverse corpus of data, while simultaneously reserving a subset for validation purposes, essential for gauging the model's performance accurately. Further streamlining the model training process [10], TF records are curated from the XML files, consolidating the image data and annotations into a structured format conducive to seamless integration with TensorFlow. This confluence of meticulously curated data and sophisticated technological frameworks, underpinned by Python, YOLOv8, and TensorFlow, serves as the cornerstone for the development of a robust and proficient model adept at accurately detecting and recognizing ISL gestures. Such a model holds profound implications in bridging communication barriers for the hearing-impaired community, empowering individuals with greater accessibility and inclusivity in the digital landscape.

Fig 3. Indian Sign Language Words

4. System Methodology

4.1 Segmentation

The process of removing objects [1] or signs from the broader context of a captured image which includes context [12] subtracting, skin-color recognition, and edge detection. To recognize gestures, the hand's motion and location must be detected and segmented.

4.2 Features Extraction

In order to classify or recognize signs, predefined aspects including form, contour, geometrical features [1](position, angle, distance, etc.), colour features, histograms, and others are taken from the pre-processed images. Form, contour, geometrical features (position, angle, distance, etc.), colour features, histograms, and other preset elements are extracted from the pre-processed images in order to categorize or identify signs. Feature extraction is a stage in the dimensionality reduction process that separates and organizes a large volume of raw data into smaller, easier-to-manage groups. This would make processing simpler.

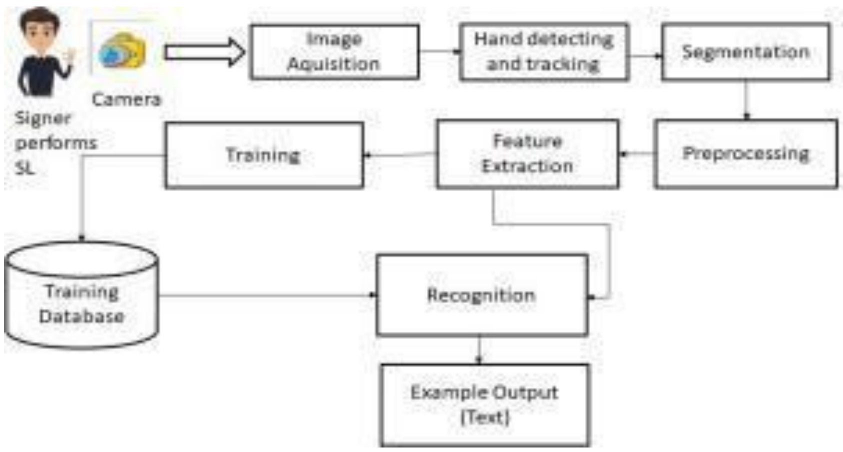

Fig 4. Proposed Methodology

4.3 Preprocessing

Every image frame is pre-processed to remove noise using a range of filters, such as Gaussian smoothing, dilation, and erosion. The conversion of a colour image to grayscale reduces the image's size. Greyscale conversion of an image is a popular technique for preprocessing are as follows:

4.4 Morphological Transform

A [12] structuring feature on an input image is used by morphological processes to determines the value of each [9] pixel in the output image by comparing it with its neighbours in the input image. Morphological changes can be divided into two categories: erosion and dilatation.

i. Dilation: The [4] value of the output pixel is equal to the maximum value of all nearby pixels. In a binary image, a pixel is set to 1 if all of its neighbours also have the value 1.By filling in small spaces, morphological dilatation makes artifacts more visible.

ii Erosion: The value among all close pixels. In a binary image, a pixel is set to 0 if every neighbouring pixel has the same value. Significant things are left behind after morphological erosion removes tiny artifacts.

4.5 Blurring

Blurring can be achieved by [7] applying a low-pass filter to an image. In computer vision, the term "low-pass filter" describes removing noise from a picture while preserving the remainder of the image. .

4.6 *Thresholding*

An image's pixels are altered in the threshold isoform of image segmentation to facilitate picture interpretation. A colour or grayscale image can be thresholder to become a binary image, which is only black and white. Typically, thresholding is used to highlight areas of interest in a photograph while disregarding the parts that don't bother us.

Table 2. Computational Time for each Phase

S.No	Phase	Time taken(in ms)
1	Data Transfer over WLAN	46.2
2	Skin Colour Segmentation and Morphological Operations	12.3
3	Face Detection and Elimination	99.9
4	Object Stabilization	14.8
5	Feature Extraction	11.2
6	Hand Pose Classification	1.7
	Average Time per Frame	186.1

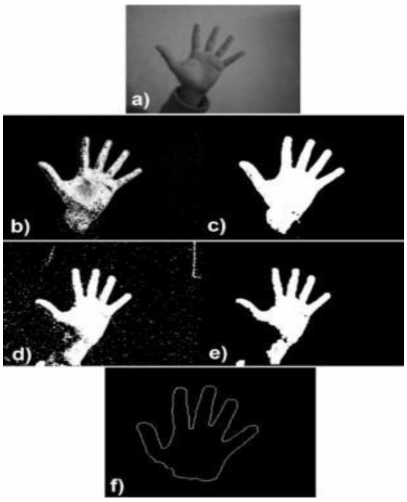

Fig. 5. (a) original image with hand (b)image of hand after skin color detection (c) after morphological operations and binarization (d) image of hand after background extraction (e) after binarization and morphological operations; (f)hand's contour made of image can "d" image "e" concatenation

5. Result and Discussions

The implemented system demonstrates real-time detection of Indian Sign Language Words. Utilizing the TensorFlow object detection API, the system leverages a pre-trained model sourced from the TensorFlow model zoo, specifically SSD Mobile Net v2 320x320. Through transfer learning, the model has been trained on a custom dataset comprising. The Indian Sign Language Recognition system, employing state-of-the-art techniques, attained an accuracy range of 93-96% [4]. Despite its commendable accuracy, it lacked real-time functionality. This limitation is addressed in the present study. Despite working with a relatively small dataset, our system achieved an average confidence rate of 85.45%.

Fig 6. Real-Time Sign Language Detection

Fig 7. Hand Gesture 'Phone' and Sleep recognized by the app.

6. Conclusion and Future Works

Sign languages serve as vital visual means for communication, relying on hand, body, and facial expressions. They offer individuals with disabilities a channel to express themselves and connect with others emotionally. However, the limited prevalence of sign language proficiency poses a barrier to effective communication. To address this challenge, automated Sign Language Recognition (SLR) systems have emerged, bridging the gap by translating sign language gestures into spoken language. This paper introduces a pioneering system developed using the TensorFlow object detection API, specifically trained on the Indian Sign Language alphabet dataset, enabling real-time detection of sign language. The process of data acquisition involves capturing images through a webcam using Python and Yolov8, thereby reducing costs associated with specialized equipment. Despite being trained on a relatively small dataset, the system achieves an impressive average 13000 images in total, with 100 images allocated for ISL Words. During the final stage of training, at 10,000 steps, the cumulative loss reached 0.25. Among the specific losses, localization loss accounted for 0.18, classification loss for 0.13, and regularization loss for 0.10. Additionally, the graph illustrates that the lowest loss of 0.17 was observed at step 9900. confidence rate of 85.45%. Looking ahead, there is immense potential for further enhancement of the system. Enlarging the dataset holds promise for improving the system's ability to recognize a broader range of gestures accurately. Moreover, exploring alternative TensorFlow models could lead to refinements in performance and efficiency. Additionally, the adaptability of the system to accommodate various sign languages by adjusting the dataset represents a significant opportunity for future development. In conclusion, our system represents a significant step forward in inclusive communication, incorporating innovative features such as text-to-audio conversion and real-time Indian Sign Language (ISL) detection. As we look to the future, our ongoing efforts will focus on dataset expansion, model refinement, and adaptation to cater to the diverse needs of different sign languages and communities.

References

1. Kapur, R.: The Types of Communication. MIJ. 6, (2020).
2. Suharjito, Anderson, R., Wiryana, F., Ariesta, M.C., Kusuma, G.P.: Sign Language Recognition Application Systems for Deaf-Mute People: A Review Based on Input Process-Output. Procedia Comput. Sci. 116, 441–448 (2017). https://doi.org/10.1016/J.PROCS.2017.10.028
3. Konstantinidis, D., Dimitropoulos, K., Daras, P.: Sign Language Recognition Based on Hand and Body Skeletal Data. 3DTV-Conference (2018). https://doi.org/10.1109/3DTV.2018.8478467
4. Dutta, K.K., Bellary, S.A.S.: Machine Learning Techniques for Indian Sign Language Recognition. Int. Conf. Curr. Trends Comput. Electr. Electron. Commun. CTCEEC2017, 333–336 (2018). https://doi.org/10.1109/CTCEEC.2017.8454988
5. Bragg, D., Koller, O., Bellard, M., Berke, L., Boudreault, P., Braffort, A., Caselli, N., Huenerfauth, M., Kacorri, H., Verhoef, T., Vogler, C., Morris, M.R.: Sign Language Recognition, Generation, and Translation: An Interdisciplinary Perspective. 21st Int. ACM SIGACCESS Conf. Comput. Access. (2019). https://doi.org/10.1145/3308561
6. Rosero-Montalvo, P.D., Godoy-Trujillo, P., Flores-Bosmediano, E., Carrascal-Garcia, J., Otero-Potosi, S., Benitez-Pereira, H., Peluffo-Ordonez, D.H.: Sign Language Recognition Based on Intelligent Glove Using Machine Learning Techniques. 2018 IEEE 3rd Ecuador Tech. Chapters Meet. ETCM (2018). https://doi.org/10.1109/ETCM.2018.8580268
7. Zheng, L., Liang, B., Jiang, A.: Recent Advances of Deep Learning for Sign Language Recognition. DICTA 2017 - Int. Conf. Digit. Image Comput. Tech. Appl. (2017). https://doi.org/10.1109/DICTA.2017.8227483
8. Rautaray, S.S.: A Real-Time Hand Tracking System for Interactive Applications. Int. J. Comput. Appl. 18, 975–8887 (2011).
9. Zhang, Z., Huang, F.: Hand Tracking Algorithm Based on Super-Pixels Feature. Proc. - 2013 Int. Conf. Inf. Sci. Cloud Comput. Companion, ISCC-C 2013, 629–634 (2014). https://doi.org/10.1109/ISCC-C.2013.77
10. Lim, K.M., Tan, A.W.C., Tan, S.C.: A Feature Covariance Matrix with Serial Particle Filter for Isolated Sign Language Recognition. Expert Syst. Appl. 54, 208–218 (2018). https://doi.org/10.1016/J.ESWA.2016.01.047
11. Lim, K.M., Tan, A.W.C., Tan, S.C.: Block-Based Histogram of Optical Flow for Isolated Sign Language Recognition. J. Vis. Commun. Image Represent. 40, 538–545 (2016). https://doi.org/10.1016/J.JVCIR.2016.07.020
12. Gaus, Y.F.A., Wong, F.: Hidden Markov Model-Based Gesture Recognition with Overlapping Hand-Head/Hand-Hand Estimated Using Kalman Filter. Proc. - 3rd Int. Conf. Intell. Syst. Model. Simulation, ISMS 2012, 262–267 (2012). https://doi.org/10.1109/ISMS.2012.67

13. Nikam, A.S., Ambekar, A.G.: Sign Language Recognition Using Image-Based Hand Gesture Recognition Techniques. Proc. 2016 Online Int. Conf. Green Eng. Technol. IC-GET2016 (2016). https://doi.org/10.1109/GET.2016.7916786
14. Mohandes, M., Aliyu, S., Deriche, M.: Arabic Sign Language Recognition Using the Leap Motion Controller. IEEE Int. Symp. Ind. Electron., 960–965 (2014). https://doi.org/10.1109/ISIE.2014.6864742
15. Enikeev, D.G., Mustafina, S.A.: Sign Language Recognition Through Leap Motion Controller and Input Prediction Algorithm. J. Phys. Conf. Ser. 1715, 012008 (2021). https://doi.org/10.1088/1742-6596/1715/1/012008
16. Cheok, M.J., Omar, Z., Jaward, M.H.: A Review of Hand Gesture and Sign Language Recognition Techniques. Int. J. Mach. Learn. Cybern. 10, 131–153 (2017). https://doi.org/10.1007/S13042-017-0705-5
17. Wadhawan, A., Kumar, P.: Sign Language Recognition Systems: A Decade Systematic Literature Review. Arch. Comput. Methods Eng. 28, 785–813 (2019). https://doi.org/10.1007/S11831-019-09384-2
18. Camgöz, N.C., Koller, O., Hadfield, S., Bowden, R.: Sign Language Transformers: Joint End-to-End Sign Language Recognition and Translation. Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., 10020–10030 (2020). https://doi.org/10.1109/CVPR42600.2020.01004
19. Cui, R., Liu, H., Zhang, C.: A Deep Neural Framework for Continuous Sign Language Recognition by Iterative Training. IEEE Trans. Multimed. 21, 1880–1891 (2019). https://doi.org/10.1109/TMM.2018.2889563
20. Bantupalli, K., Xie, Y.: American Sign Language Recognition Using Deep Learning and Computer Vision. Proc. - 2018 IEEE Int. Conf. Big Data, Big Data 2018, 4896–4899 (2019). https://doi.org/10.1109/BIGDATA.2018.8622141
21. About - OpenCV (2017). Poster of the Manual Alphabet in ISL | Indian Sign Language Research and Training Center, Government of India. Transfer Learning and Fine-Tuning | TensorFlow Core (2017).
22. Wu, S., Yang, J., Wang, X., Li, X.: IoU-Balanced Loss Functions for Single-Stage Object Detection (2020).