INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT (IJSREM)
VOLUME: 08 ISSUE: 04 | APRIL - 2024 SJIF RATING: 8.448 ISSN: 2582-3930

Sign Language Recognition Using Convolutional Neural Network

Niranjana R, Shri Raksha M P, Thirisha S

Department of Computer Science and Engineering

Arunachala College of Engineering for Women

Tamilnadu,India.

Caroline Mishba J

Assistant Professor

Department of Computer Science and Engineering

Arunachala College of Engineering for Women

Tamilnadu,India.

Abstract— **This project introduces a real-time sign language detection system powered by Convolutional Neural Networks (CNNs), designed to aid individuals with hearing impairments in communication. Leveraging OpenCV for video capture and hand detection, along with custom modules for hand tracking and image classification, the system seamlessly integrates deep learning methodologies. A pre-trained CNN model, trained on a comprehensive dataset of sign language gestures, forms the core of the system, ensuring accurate classification in real-time. By capturing video frames from a webcam, detecting hands, and processing them through the CNN model, the system provides immediate feedback by overlaying predicted gestures onto the video stream. Through its effective implementation, this project underscores the potential of CNNs in facilitating accessibility and inclusivity, while also paving the way for future enhancements and applications in assistive technology and human-computer interaction.**

Keywords— Real-time, Sign Language Detection, Convolutional Neural Networks (CNNs), Communication, Hearing Impairments, Deep Learning.

I. INTRODUCTION

Deafness and hearing loss are not confined to specific regions or demographics but represent a global challenge affecting individuals worldwide. With over 1.5 billion people, approximately 20% of the global population, grappling with varying degrees of hearing impairment, and a staggering 430 million experiencing disabling hearing loss, the magnitude of this issue is undeniable. Despite the prevalence, individuals with hearing impairments often encounter barriers to effective communication, hindering their ability to fully engage in social, educational, and professional spheres. Recognizing the imperative to address these barriers and promote inclusivity, this project endeavors to develop a real-time sign language detection system powered by Convolutional Neural Networks (CNNs).By harnessing the capabilities of advanced technologies such as CNNs, this innovative solution seeks to bridge the communication gap and enhance accessibility for individuals with hearing impairments.

Through real-time gesture recognition and interpretation, the system aims to empower users to express themselves fluently in sign language, fostering meaningful connections and enabling participation in diverse contexts. Moreover, by advocating for greater awareness and understanding of deafness and hearing loss, this project aligns with broader initiatives aimed at promoting equality, inclusion, and social cohesion for individuals of all abilities, reaffirming the fundamental right to effective communication for everyone.

II. METHODOLOGY

A. *Project setup and Environment configuration*

First, make sure a webcam or other appropriate camera equipment that can record live video is available. Install Python and all required libraries, including cvzone and OpenCV. Furthermore, confirm that the Hand Tracking Module is accessible and deployed successfully.

Create distinct folders for every category of gesture in order to store the acquired photographs in an organised folder hierarchy. This makes organising and labelling the gathered data easier.

B. *Hand Tracking and Detection*

To track and identify hands in the camera feed, use the HandTrackingModule from the cvzone library. To make things easier and guarantee precise tracking of the user's hand movements, set the hand detector to detect a maximum of one hand (maxHands=1).

C. Real-Time Image Capture

Use OpenCV's VideoCapture module to continuously record frames from the camera feed.Run a real-time loop to process every frame and recognise the hand movements. Extract the area of interest (ROI) containing the hand motion as soon as a hand is detected in the frame so that it can be processed further.

D. Image Preprocessing

To guarantee consistency between samples, normalise the hand gesture image that was collected. To standardise the measurements for training machine learning models, resize the hand gesture image to a predetermined size (such as 300x300 pixels).

If colour information is not needed for gesture identification, then convert the hand gesture image to grayscale.

E. Classification and Labelling of Gestures

Provide visual feedback by displaying both the original frame and the preprocessed hand gesture image. Provide a feature that allows you to take and store pictures whenever the user presses a particular key (like the 's' key).

In order to avoid overwriting and maintain dataset integrity, give each captured image a unique filename that is determined by the current timestamp. Sort the photos that were taken into distinct folders based on the various sign language expressions (such as "please," "thank you," "hello," etc.).

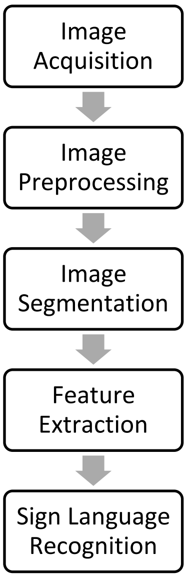

Fig. 1. Flowchart of Methods Used

III. ALOGORITHM USED

Convolutional neural networks (CNNs) are a kind of artificial neural network that analyses data using supervised learning and the perceptron learning rule.CNN is used for a variety of cognitive tasks, including image processing and natural language processing. similar to other

Convolutional neural networks are among the different types of artificial neural networks that have multiple hidden layers, an output layer, and an input layer.Convolutional layers use mathematical models to transfer information from one layer to the next. This mimics a few of the functions of the visual cortex in humans. A basic illustration of a deep learning algorithm is CNN.

1. Input Layers: This is the layer where our model receives its input. The total number of features is equal to the number of neurons in this layer.
2. Hidden Layer: The input from the input layer is fed into the hidden layer through this layer.Several hidden layers may exist, contingent on the amount of the data and our model. The number of neurons in each hidden layer varies, but they are usually more than the number of characteristics. The network is nonlinear because the output from each layer is calculated by multiplying the output of the previous layer by the learnable weights of that layer, adding learnable biases, and then applying the activation function.
3. Output Layer: A logistic function, such as a sigmoidal or softmax layer, receives the output from the hidden layer and uses it to calculate the probability score for each class.

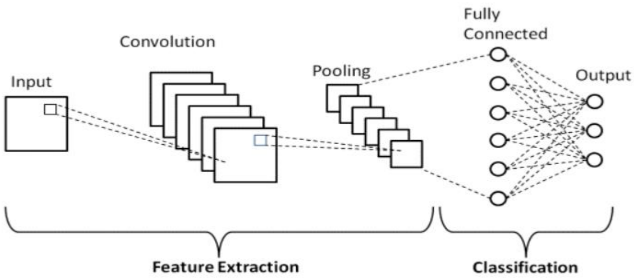

Fig. 2. Layers in Neural Network

IV. CONCLUSION

The project to collect images of sign language gestures is a significant step towards the creation of assistive devices for people with hearing loss. The project intends to identify and understand sign language motions in real-time by utilising computer vision techniques and machine learning models, thereby improving communication and connection for people who use sign language as their primary means of expression.

Significant progress has been made in capturing and preprocessing hand gesture photographs for the purpose of creating an extensive sign language dataset thanks to the creation of a strong system architecture that comprises modules for hand identification, image processing, user interaction, and data administration. Strong tools for image processing and hand tracking have been made possible by the use of the OpenCV and cvzone libraries, and the incorporation of classification methods.

User experience, ethical issues, and data integrity are given top priority in the project's methodology, which guarantees that the dataset gathered is not only representative but also considerate of participants' privacy and consent. The initiative maintains ethical standards and builds community confidence by following ethical rules and acquiring required licences.

In the future, this project's dataset will be extremely valuable for training machine learning models to reliably detect and understand motions in sign language.

Additionally, the dataset can be a useful tool for academics, researchers, and developers that want to advance inclusive technology and sign language recognition.In summary, the endeavour to gather images of sign language gestures paves the way for further developments in assistive technology and accessible solutions, ultimately fostering a more inclusive technology and sign language recognition.

V. OUTPUT

The output of the sign language gesture image collection project encompasses the culmination of efforts in capturing, processing, and organizing hand gesture images. Through the webcam feed, various sign language gestures are captured in real-time and stored as digital images. These images undergo preprocessing steps such as normalization and resizing to ensure uniformity and consistency across the dataset. Organized within structured folders based on gesture categories, the captured images serve as valuable raw data for subsequent tasks. Alongside the image data, detailed documentation and logs chronicle the project's progression, documenting key aspects such as data collection procedures, observations, challenges encountered, and any errors identified. This comprehensive documentation not only provides insights into the project's methodology but also serves as a repository of valuable insights and learnings. Together, the captured images and documentation lay the groundwork for further analysis, model training, and development of sign language recognition systems, contributing towards advancements in assistive technologies for individuals with hearing impairments.

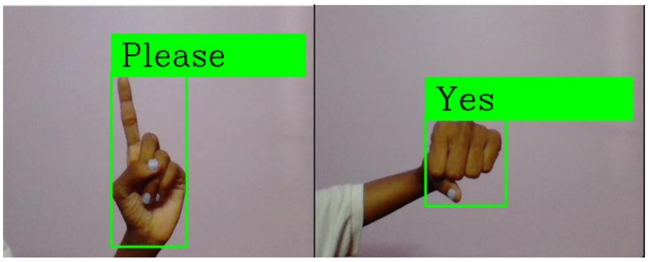

Fig.3 .output images

VI. FUTURE WORK

The gesture count can be adjusted to recognise a greater number of movements than the one utilised here, which can recognise any two distinct gestures. This model may be trained to recognise temporal features, or features that vary in time, by employing different methods like LSTM. It can detect gestures in a single frame.

It can be more accurate and adaptive to investigate ways to combine several modalities—like audio and depth data—with visual data, especially when dealing with difficult surroundings.

It may be possible to improve communication between sign language users and non-signers by delving deeper into cross-modal translation techniques, such as bidirectional translation between spoken and sign language. Moving forward, there are several promising avenues for enhancing the capabilities and impact of the sign language gesture image collection project. Firstly, expanding the dataset to include a wider range of sign language gestures, covering various alphabets, words, and phrases, would enrich the dataset's diversity and utility. Additionally, implementing data augmentation techniques to introduce variability in the dataset could enhance model performance and reduce overfitting.

Further improvements could be made through model fine-tuning and optimization, exploring advanced architectures and hyperparameter tuning strategies to boost accuracy and efficiency. Development of a real-time sign language recognition system, coupled with gesture translation and synthesis capabilities, would enable seamless communication for individuals with hearing impairments. Enhancements in user interface design, accessibility integration, and collaboration with the sign language community could further enrich the project's usability and relevance. Moreover, prioritizing ethical considerations, privacy protection, and scalability in future endeavors will ensure the project's sustainability and societal impact. By pursuing these avenues for future work, the project can continue to evolve and contribute towards creating inclusive technologies for individuals with diverse communication needs.

VII. REFERENCES

[1] 1. B. Jiang, X. Li, L. Yin, W. Yue and S. Wang, "Object Recognition in Remote Sensing Images Using Combined Deep Features," 2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), Chengdu, China, 2019, pp. 606-610.

[2] 2. A. Gupta, Y. Kumar and S. Malhotra, "Banking security system using hand gesture recognition," 2015 International Conference on Recent Developments in Control, Automation and Power Engineering (RDCAPE), Noida, 2015, pp. 243-246.

[3] 3. R. Y. Wang and J. Popovi, “Real-Time Hand-Tracking with a Color Glove”, ACM Transactions on Graphics, vol. 28(3), Jul 2009.

[4] 4. D. Minnen, “Fast Fingertip Detection for Initializing a Vision-Based Hand Tracker,” U.S. Patent, Application NO. 13/430,509, Oct. 25, 2012.

[5] 5. P. T. Keaton, S. Dominguez and A. H. Sayed, “Vision-Based Pointer Tracking and Object Classification Method and Apparatus,” U.S. Patent 7148913, Dec. 12, 2006.

[6] 6. J. Sun, T. Ji, S. Zhang, J. Yang and G. Ji, "Research on the Hand Gesture Recognition Based on Deep Learning," 2018 12th International Symposium on Antennas, Propagation and EM Theory (ISAPE), Hangzhou, China, 2018, pp. 1-4

[7] 7. S. Hussain, R. Saxena, X. Han, J. A. Khan and H. Shin, "Hand gesture recognition using deep learning," 2017 International SoC Design Conference (ISOCC), Seoul, 2017, pp. 48-49.

[8] 8. D. Y. Xu, "A Neural Network Approach for Hand Gesture Recognition in Virtual Reality Driving Training System of SPG," In Proceedings of the 18th International Conference on Pattern Recognition, vol. 3, pp. 519-522, 2006.

[9] 9. W. T. Freeman and C. D. Weissman, "Television Control by Hand Gestures," In Proceedings of International Workshop on Automatic Face and Gesture Recognition, pp. 179-183, 1995.

[10] 10. T. Starner, J. Weaver and A. Pentland, "Real-Time American Sign Language Recognition Using Desk and Wearable Computer Based Video," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20(12), pp. 1371-1375, 1998.