WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

EchoGest: Echoing Gestures into Sound

Lec. V. G. Sonawane[1], Ms. Shreeya Bhosale [2], Ms. Anisha Sasane [3]

[1] Lecturer, Department of Artificial Intelligence & Machine Learning, AISSMS Polytechnic, Pune, Maharashtra, India
[2,3] Student , Department of Artificial Intelligence & Machine Learning, AISSMS Polytechnic, Pune, Maharashtra, India

ABSTRACT :

Sign language serves as a means of communication with those who are deaf. The inability to communicate verbally is a genuine handicap. Speech impairment is a handicap that hinders an individual's capacity to communicate via speech and auditory processing. Individuals impacted by this handicap use communication methods such as sign language. A communication barrier persists between non-signers and signers, despite the widespread use of sign language today. Significant gains have been made in motion and gesture identification due to recent developments in computer vision and deep learning methodologies. This effort primarily aims to develop a deep learning application that translates sign language into text, facilitating communication between signers and non-signers. We use a bespoke convolutional neural network (CNN) for sign recognition from a video frame.

Keywords: Gesture Recognition, Sound Mapping, Accessibility, Deep Learning, Real-Time Processing, Human-Computer Interaction.

INTRODUCTION:

People with impaired speech and hearing use sign language as a form of communication. Disabled People use these sign language gestures as a tool of non-verbal communication to express their emotions and thoughts to other common people. But these common people find it difficult to understand their expression; thus, trained sign language expertise is needed during medical and legal appointments and educational and training sessions. Over the past few years, there has been an increase in demand for these services. Other forms of services, such as video remote human interpretation using a high-speed Internet connection, have been introduced; thus, these services provide an easy-to-use sign language interpretation service that can be used and benefited from, yet they have major limitations. To address this, we use a custom CNN model to recognize gestures in sign language. We construct a convolutional neural network with 11 layers: four convolutional layers, three max-pooling layers, two dense layers, one flattening layer, and one dropout layer. We use the American Sign Language Dataset from MNIST to train the model to identify the gesture. The dataset contains the features of different augmented gestures. Introduced a custom CNN (Convolutional Neural Network) model to identify the sign from a video frame using OpenCV

LITERATURE SURVEY :

The model identifies American Sign Language using deep learning and computer vision [1]. The model gathers temporal and spatial attributes from video segments. Subsequently, to detect spatial attributes, we use Inception, a Convolutional Neural Network (CNN). Subsequently, to model temporal properties, we use a recurrent neural network (RNN). This research used an American Sign Language dataset.

A deep learning-based method for recognizing static sign language signals [2]. Humans can communicate proficiently using sign language, and significant research in computer vision is now under progress. The first research on Indian Sign Language (ISL) identification concentrated on the detection of significant, distinct hand signals; hence, only a limited number of ISL signs were selected for recognition. A cumulative collection of 35,000 signed photographs of over 100 static signs has been amassed from diverse users. The proposed system's efficacy is evaluated on around 50 CNN models.

Utilization of deep convolutional neural networks for the recognition of sign language [3]. This research utilizes a capture approach including continuous sign language video in selfie mode, enabling a hearing-impaired individual to freely use the SLR smartphone application. The dataset was developed with five unique people performing 200 signals from five different points of view under diverse background circumstances, addressing the scarcity of sign language datasets taken in smartphone selfies. In the video, each sign had around 60 frames. CNN training use three separate sample sizes, each including a varying number of persons and perspectives. The last two samples are used to evaluate the trained CNN.

Recognition of static sign language by deep learning [4]. The objective of the project was to develop a system capable of translating static sign language into its corresponding word counterparts, including letters, numerals, or fundamental static signals, to familiarize individuals with the fundamentals of sign language. The researchers established an assessment technique and performed many tests to validate the significance of the system's non-signer functionalities. The solution received high marks for usability and learning efficacy throughout the evaluation.

A Comprehensive Study on Deep Learning-Based Methods for Sign Language Recognition [5]. A comparative experimental evaluation is conducted on computer vision-based sign language recognition systems. Recent deep neural network methodologies in this domain are used to conduct a comprehensive assessment of various publically accessible datasets. This project aims to enhance sign language recognition by mapping non-segmented video streams to glosses. This paper presents two novel sequence training criteria derived from the fields of voice and scene text recognition.

A deep learning-based method for recognizing static sign language signals [6]. The research discusses the use of deep learning via convolutional neural networks for the accurate identification of static signals in sign language recognition. This research amassed 35,000 photographs of signs, with each image depicting 100 static signs, contributed by diverse users. The proposed system's efficacy is evaluated on around 50 CNN models. Sign language is a complex and intricate system that relies on computer vision to interpret messages produced by hand motions with face expressions. It is a natural language used by those with hearing impairment to communicate. Sign language employs diverse hand gestures to convey letters, words, or sentences. We provide a pragmatic method for identifying ISL numbers, letters, and phrases in commonplace contexts. The suggested CNN design first employs convolutional layers, followed by ReLU and max-pooling layers.

Recognition of British Sign Language by Transfer Learning to American Sign Language using Late Fusion of Computer Vision and Leap Motion [7]. Researchers conducted many tests in both British and American Sign Language, emphasizing solitary sensory and multimodal methodologies. The results indicate that a multimodal approach surpasses the two individual sensors in training and classifying unknown inputs. This work included a preliminary transfer learning experiment from a substantial BSL dataset to a medium-sized ASL dataset, whereby the multimodality model was identified as the most effective for ASL classification after the transfer of weights from the BSL model. This research benchmarked and assessed all network topologies that were trained, compared, and ultimately fused to achieve multimodality for the first time. The precise classification of sign language, especially with unobserved data, enables autonomous completion of the process, offering a computerized method for interpreting non-spoken language in scenarios where interpretation is required yet inaccessible.

The mArSL Database and Pilot Study: Advancing Hybrid Multimodal Recognition of Manual and Non-Manual Arabic Sign Language [8]. A novel multi-modality ArSL dataset that integrates many modalities. The dataset consists of 6,748 video samples captured using Kinect V2 sensors, with fifty signs shown by four signers. This dataset will enable researchers to formulate and evaluate their approaches to advance the field. Furthermore, we used cutting-edge deep learning algorithms to examine the amalgamation of spatial and temporal attributes of various modalities, both manual and non-manual, for sign language recognition.

Recognition System for Thai Finger-Spelling Sign Language Using Deep Learning and Multi-Stroke Techniques [9]. A vision-based methodology was used to achieve semantic segmentation via dilated convolution for hand segmentation, optical flow separation for hand strokes, and feature learning and classification through a convolutional neural network (CNN). The five CNN architectures that dictate the forms were then compared. The initial format employed 64 filters, each measuring 3x3, across 7 layers; the subsequent format utilized 128 filters, also 3x3 in size, with 7 layers; the third format incrementally increased the number of filters while maintaining 7 layers, all featuring a uniform 3x3 filter size; the fourth format mirrored this structure; the final format was a structured configuration.

Utilizing k-Nearest Neighbors with Dynamic Time Warping and Convolutional Neural Network algorithms in wearable technology for sign language recognition [10]. The research includes a wearable electronics-based device for sign language recognition that utilizes two separate classification algorithms. The wearable electronics captured finger, wrist, and arm/forearm movements with a sensory glove and inertial measurement units. k-Nearest Neighbors using Dynamic Time Warping (a non-parametric methodology) and Convolutional Neural Networks used as classifiers (a parametric method) were implemented. Ten sign-words from Italian Sign Language were analyzed, including cose, grazie, and maestra, alongside globally recognized phrases such as google, internet, jogging, pizza, television, twitter, and ciao. Seven individuals, including five males and two females, aged 29 to 54 years, each replicated the signals 100 times (SD).

A discrete time sliding mode controller (DSMC) is proposed for higher solicitation not withstanding defer time (HOPDT) frames [11]. As portion of structure states and botch, a sliding mode surface is selected and the tuning parameters of the sliding mode controller are resolved using overpowering post circumstance scheme. The control object for "ball in a barrel" is to handle the velocity of a fan blowing air into a chamber to keep a ball suspended in the barrel at a certain predestined position. The DSMC is attempted to coordinate the ball's position subsequently. But skillfully clear, this is a troublesome control issue due to the non-direct ramifications for the ball and the confounding material scienceT regulating its lead. The DSMC is attempted to coordinateT theT ball's position subsequently.

The fine-tuned PID controller have proposed for the air levitation system [12]. Advanced genetic algorithm is used for tuning parameters of PID controllers. For demonstration of efficiency and applicability of the proposed PID controller, simulation-based experimentations have been conducted. The proposed PID design method has been linked with other three optimisation techniques. Ant colony optimisation, particle swarm optimisation and fuzzy logic have been used for performances comparison of advanced genetic algorithm-based PID controllers. In experimental results, we have got very smallest value of IAE, ISE and ITAE using proposed method. It indicates that the proposed PID design method offers better performances than other three optimisation-based PID design methods and other existing methods.

METHODOLOGY

The figure above illustrates the initial segmentation of the image from a webcam video input. The frames are dropped from the video with a region of interest so as to avoid background conflicts. A custom CNN model with 11 layers is used. The gesture image segmented from the video frame is then converted to a grayscale image. The input image from the webcam is converted to grayscale since the model is trained with the features of the grayscale images; i.e., the MNIST dataset is a pre-processed dataset of RGB images that are converted to grayscale. The converted image is then scaled in respect to the size of the images with which the model was trained. The image is fed into the pre-trained custom CNN model post scaling and transformation. We obtain the gesture prediction from the CNN model and then classify it based on the categorical label. The classified gesture is displayed as text.

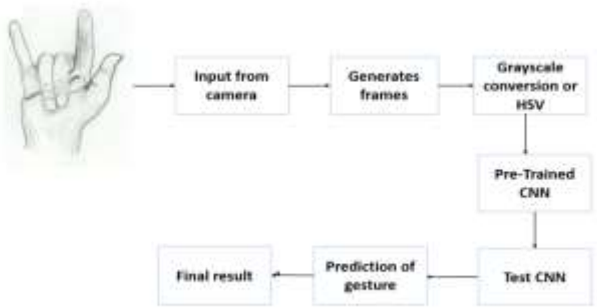

Figure 1: System Architecture

Module 1: Data Collection: We created different images based on hand sign image to different users.

Module 2: Data Training: Using hand sign images we collect artificial as well as real time and train with any classification.

Module 3: Testing with deep learning: We use hand sign image weights according to any deep learning classifier for artificial input data.

Module 4: Analysis: We demonstrate the accuracy of proposed system and evaluate with other existing systems.

The gesture prediction from the CNN model is obtained and post that, it is classified based on the categorical label. The classified gesture is displayed as text.

The gesture prediction from the CNN model is obtained and post that, it is classified based on the categorical label. The classified gesture is displayed as text.

Step 1: - Pre-processing To prepare the data for further processing at this stage, we did some preprocessing on the images of our proposed dataset. We first converted the gray-scale image from our dataset and then resized it to res 64x64 pixels by keeping the aspect ratio locked.

Step 2: - Each manuscript was designed to draw structural/geometric features with the image's pixel-based data, such as letter structure, letter width, character height, text image ratio, horizontal, and number. These features were then embedded with pixel-based data from the image, such as the image's vertical lines, the number and location of the loops and arcs, etc., in order to obtain accurate classification results.

Step 3: - CNN's architecture is somewhat different from the model of a traditional neural network. Input values are transformed in the conventional neural network into a sequence of hidden layers by traversing. Each layer consists of a series of neurons, where each layer is entirely linked to all the neurons in the previous layer. The explanation behind the better performance of CNNs is that the intrinsic properties of images are captured by these networks. This critical CNN feature gave us the courage to use it in our proposed dataset analysis.

Step 4: - Download the dataset from open-source websites like Kaggle, MNIST database etc. or create your own database. Database is divided into two parts, 70% for training and 30% for testing purposes.

Step 5: - Extract various features by using CNN model in a convoluted neural network, there are four layered principles that we can understand:

- Convolution
- Pooling
- Activation Function
- Fully Connected Layer

Step 6: - Finally classification has done in dense layer and show the output for current input.

RESULTS

The graph below illustrates the classification of the system. The graphs illustrate the system's allocation of total inputs among several scenarios. The suggested methodology utilizes a mixture of CNNs that yields exceptional results universally. The input photographs were supplied for training, and assessments were conducted for performance evaluation using several classification models.

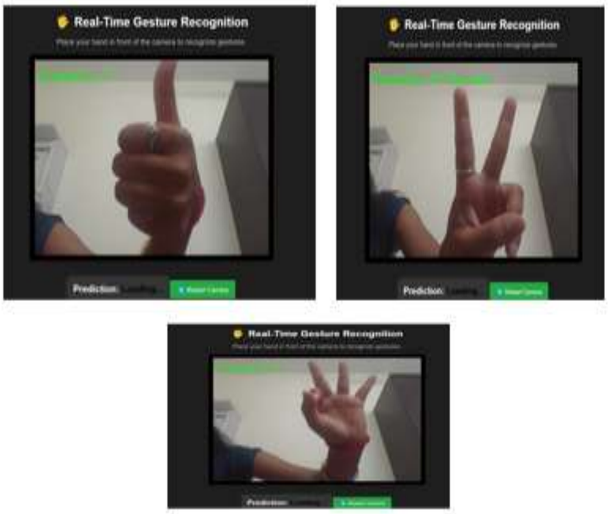

Figure 2 below illustrates a real-time assessment of the suggested model. The model demonstrates its capability to recognize sign language

CONCLUSION :

The proposed system will successfully predict the signs of sign and some common words under different lighting conditions and different speeds. Accurate masking of the images is being done by giving a range of values that could detect human hand dynamically. The proposed system uses CNN for the training and classification of images. For classification and training, more informative features from the images are finely extracted and being used. The proposed system can deal with various deep learning frameworks for predicting the activity of sign. By using a convolutional neural network system can able to provide higher accuracy.

Acknowledgments

We extend our heartfelt gratitude to all contributors, mentors, and organizations that supported this project. Special thanks to researchers and developers in the fields of AI and accessibility for their pioneering work, which served as the foundation for EchoGest. Additionally, we thank our families and friends for their unwavering support and encouragement throughout this journey.

REFERENCES :

[1] Rajan, Rajesh George, and M. Judith Leo. "American Sign Language Alphabets Recognition using Hand Crafted and Deep Learning Features." 2020 International Conference on Inventive Computation Technologies (ICICT). IEEE, 2020.

[2] Wadhawan, Ankita, and Parteek Kumar. "Deep learning-based sign language recognition system for static signs." Neural Computing and Applications 32.12 (2020): 7957-7968.

[3] Dhivyasri, S., et al. "An Efficient Approach for Interpretation of Indian Sign Language using Machine Learning." 2021 3rd International Conference on Signal Processing and Communication (ICPSC). IEEE, 2021.

[4] Tolentino, Lean Karlo S., et al. "Static sign language recognition using deep learning." International Journal of Machine Learning and Computing 9.6 (2019): 821-827.

[5] Adaloglou, Nikolaos M., et al. "A Comprehensive Study on Deep Learning-based Methods for Sign Language Recognition." IEEE Transactions on Multimedia (2021).

[6] Krishnan, Palani Thanaraj, and Parvathavarthini Balasubramanian. "Detection of alphabets for machine translation of sign language using deep neural net." 2019 International Conference on Data Science and Communication (IconDSC). IEEE, 2019..

[7] Bird, Jordan J., Anikó Ekárt, and Diego R. Faria. "British sign language recognition via late fusion of computer vision and leap motion with transfer learning to american sign language." Sensors 20.18 (2020): 5151.

[8] Luqman, Hamzah, and El-Sayed M. El-Alfy. "Towards Hybrid Multimodal Manual and Non-Manual Arabic Sign Language Recognition: mArSL Database and Pilot Study." Electronics 10.14 (2021): 1739.

[9] Pariwat, Thongpan, and Pusadee Seresangtakul. "Multi-Stroke Thai Finger-Spelling Sign Language Recognition System with Deep Learning." Symmetry 13.2 (2021): 262.

[10] Saggio, Giovanni, et al. "Sign language recognition using wearable electronics: implementing k-nearest neighbors with dynamic time warping and convolutional neural network algorithms." Sensors 20.14 (2020): 3879.

[11] Bharat Kumar Shamrao Patil, Laxman M. Waghmare, M. D. Uplane, "Implementation of SMC Control Action with Pi Sliding Surface for Non Linear Plant Along with Changing Set Point", IJITEE, ISSN: 2278-3075, Volume-8 Issue-8S3, June 2019.

[12] Dwarkoba P. Gaikwad, Bharat S. Patil, Laxman S. Patil," Advance genetic algorithm-based PID controller for air levitation system", International Journal of Modelling, Identification and Control, 2022 Vol.41 No.3, pp.243 – 255.