2024 IEEE International Conference on Information Technology, Electronics and Intelligent Communication Systems (ICITEICS)
Karnataka, India. Jun 28-29, 2024

2024 IEEE International Conference on Information Technology, Electronics and Intelligent Communication Systems (ICITEICS) | 979-8-3503-8269-3/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICITEICS61368.2024.10625127

Sign Language Recognition System Using Convolutional Neural Network

Ms Pavan Kumar
Department Computer Science and Engineering,
Sathyabama Institute of Science and Technology,
Chennai-600119, India
Pavankumarms2002@gmail.com

Pasam Chanakya
Department of Computer Science and Engineering,
Sathyabama Institute of Science and Technology,
Chennai-600119, India
pasamchanakya@gmail.com

Saranya Velusamy
Assistant Professor
Department of Computer Science and Engineering,
Sathyabama Institute of Science and Technology,
Chennai-600119, India
saranya.cse@sathaybama.ac.in

***Abstract*— The various uses of hand gesture detection systems and their capacity to facilitate effective machine-human interaction have garnered significant interest in recent years. An overview of current hand gesture recognition systems is provided in this study. The challenges of the gesture system are highlighted with key issues of the hand gesture recognition system. Additionally, a review of the recent motions and gesture detection technology is provided. A synopsis of the research findings for databases, hand gesture techniques, and a comparison of the key stages of gesture recognition are also provided. Finally, the benefits and cons of the systems under discussion are clarified. Though there is a large social group that could benefit from it, the concept of using technology to recognize sign language is not widely used. There are numerous technologies at this disposal that could be useful in creating a connection between this social group and the outside world. Knowing sign language is one of the keystones to empowering users to interact with the general public. Computers are able to comprehend sign language through the application of machine learning and picture classification, which people can then interpret. This study uses convolutional neural networks to identify sign language motions. The RGB camera was used to capture static sign language gestures for the image collection.**

Keywords: Recognition of Sign Language, Inception V3, Deep Learning, and Image Processing

I. INTRODUCTION

Sign languages, also referred to as signed languages, employ visual-manual techniques for communication, utilizing a blend of manual and non-manual expressions. These languages are complete and autonomous natural languages, possessing distinct vocabularies and syntactic structures. It is crucial to note that sign languages are not mutually understandable or interchangeable; however, certain noteworthy similarities exist across various sign languages.

Advocates of sign language and spoken language view both as natural languages, indicating that their development has been gradual and abstract over time without deliberate design. It is essential to differentiate sign language, recognized as a natural language, from body language, which is a nonverbal form of communication.

In communities with a deaf population, sign languages have naturally evolved, playing a crucial role in local Deaf cultures and serving as an effective means of communication. While primarily utilized by the deaf and hard of hearing, signing is also adopted by hearing individuals who are mute, those facing challenges in comprehending spoken language (augmentative and alternative communication), and those who simply choose to communicate through signing.

Sign languages are distinct from one another in terms of their structure and regional variances. These are whole languages that are able to convey intricate concepts and feelings, not just gestures or pantomime. Sign languages can use hand shapes, movements, facial emotions, and body language to transmit meaning thanks to the visualmanual modality.

The awareness and comprehension of sign languages have increased dramatically over time. In order to increase sign language accessibility and inclusion in public spaces, education, and social interactions, efforts have been made. In a variety of settings, sign language interpreters are essential in promoting communication between the deaf and the hearing.

To sum up, sign languages are complex and dynamic modes of communication that have evolved within global deaf communities. They have their own vocabulary and grammar, and they develop together with the cultures in which they are ingrained. It is imperative that sign languages be acknowledged as unique and natural languages in order to promote tolerance and comprehension in our multicultural society.

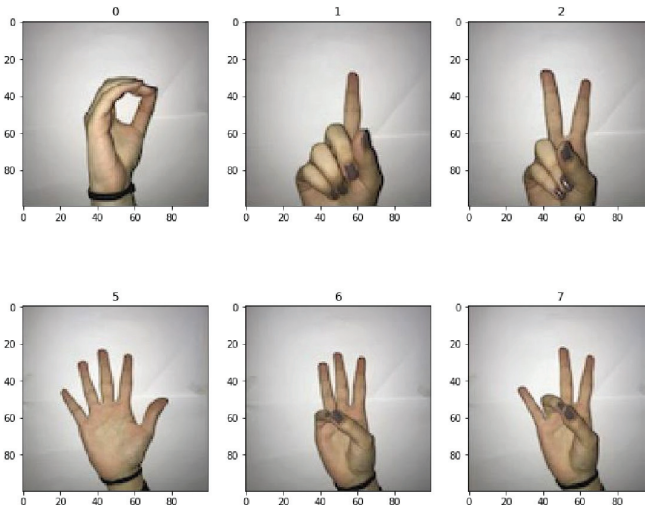

Fig. 1. Example

II. LITERATURE REVIEW

This research [1] The literature review investigates the use of convolutional neural networks (CNNs) for realtime

recognition of American Sign Language (ASL). A strong model was produced by transfer learning, which involved fine-tuning with Surrey and Massey University ASL datasets after a GoogLeNet that had already been trained on the ILSVRC2012 dataset. The system functions satisfactorily for letters a-k and correctly recognizes letters a-e. The CNN-based video processing system does not include the dynamic letters j and z. It functions via a web application. This study represents a major advancement in improving public-deaf community communication.

The integration of RGB and Time-of-Flight (ToF) cameras for real-time 3D hand gesture interaction is examined in this paper's literature review [2]. In order to increase accuracy in a variety of scenarios, the study presents a novel hand identification method that combines color and depth information. Advanced interaction modes are made possible by precise 3D hand location tracking through the use of ToF camera depth data. The study goes on to compare algorithm performance on RGB photos against depth data for hand gesture detection. The result is a robust real-time 3D hand gesture interaction system that remains unaffected by external factors or other people.

This study [3] An automated approach for identifying finger-spelled words in British Sign Language (BSL) from video is investigated in the literature review. To train their model, the authors used a dataset of 1,000 grainy webcam films with 100 words each. Their suggested approach combines strong visual features for precise hand shape recognition and just considers hand shape, eliminating the requirement for motion cues. With a 98.9% word recognition accuracy, this device is a major step forward in improving communication between the hearing and the deaf communities.

This research [4] The application of the Kinect depthmapping camera for American Sign Language (ASL) recognition is examined in the literature review, following Zafrulla et al.'s investigation. For the study, 1000 ASL phrases were gathered, and the words were recognized using a hidden Markov model (HMM). For the ASL alphabet, the accuracy was 96.7%, and for the ASL numbers, it was 91.5%. A major advancement in improving communication between the deaf population and the broader public is this system.

A real-time finger spelling recognition system for American Sign Language (ASL) is examined in the literature study [5], which presents Pugeault and Bowden's method. The interactive interface makes use of the OpenNI+NITE framework for hand recognition and tracking, as well as a Microsoft Kinect device to capture appearance and depth images. Random forests are used to characterize and classify hand shapes that represent alphabet letters. The study highlights that a combination of appearance and depth photos produces the best results when comparing the efficacy of classification. The technology functions in real-time and is included into an interface that enables signers to make decisions in the event of unclear detections. It is also connected to an English dictionary to facilitate effective writing. This ground-breaking method is a significant step in improving public-deaf community communication.

Kuznetsova et al. (1) presented a novel method for realtime sign language detection using a consumer depth camera, which is explored in the literature study [6]. Their technique uses a multi-layered random forest (MLRF) to accurately recognize static motions by classifying feature vectors that are obtained from depth pictures. One major benefit is that MLRF requires less memory and training time than a basic random forest while maintaining a similar level of accuracy. The American Sign Language (ASL) signals in a publicly accessible dataset, artificial intelligence (AI) generated data, and a recently gathered dataset created using the Intel Creative Gesture Camera are all used by the authors to assess their system. This study represents a critical turning point in the advancement of public-deaf community communication.

The literature review looks at a unique method that Dong et al. [7] suggest for using the Microsoft Kinect depth camera to recognize the American Sign Language (ASL) alphabet. Using a per-pixel classification technique based on depth contrast features, the approach generates a segmented hand configuration. After that, a hierarchical mode-seeking technique with kinematic restrictions is shown to localize hand joint locations. A Random Forest (RF) classifier is used in the last phase to identify ASL signals based on joint angles. Using a publicly accessible dataset from Surrey University, the authors validate their technique and show that it is more effective than prior benchmarks, detecting 24 static ASL alphabet signs with over 90% accuracy.

This literature review investigates a Karhunen-Loeve

Transform (KLT)-based hand gesture detection system that Singha and Das [8] suggested. Skin filtering, palm cropping, edge detection, feature extraction, and classification are the five essential procedures that make up the system. First, skin filtering is used for hand identification, and then palm cropping is used to separate the palm area. Palm outline extraction is then done using Canny Edge Detection. After extracting hand characteristics using the K-L Transform approach, an appropriate classifier is applied for gesture identification. Tested with ten distinct hand motions, the system gets an impressive 96% identification rate. In terms of promoting communication between the deaf community and the wider public, this research is a noteworthy accomplishment [8].

The literature review investigates a gesture detection setup that aims to highlight and identify the most unclear static single-handed sign language actions, as demonstrated by Sharma et al. [9]. The authors use two methods for segmenting hand contours: thresholding gray level intensities and using RGB and YCbCr color spaces for skin color detection. The gesture contours that are produced by each segmentation technique are described by a distinct rotation and size invariant contour tracing descriptor. Gray level segmented contour traces classified by multiclass Support Vector Machine (SVM) achieve up to 80.8% accuracy on the most ambiguous American Sign Language (ASL) alphabet gestures, with an overall accuracy of 90.1%, according to performance evaluations of k-Nearest Neighbor (k-NN) and SVM classification techniques. This study represents a critical turning point in the advancement of communication.

The literature review, as given by Starner et al. [10], investigates real-time recognition of American Sign Language (ASL) using video input. Two hidden Markov model-based systems for continuous ASL sentencelevel recognition are presented in this paper. The first system achieves a 92%-word accuracy rate by using a desk-mounted camera. With a 40-word vocabulary, the second system, which uses a camera

embedded in a hat worn by the user, achieves a greater accuracy of 98% (97% with an unconstrained grammar). This study represents a significant breakthrough in improving public-deaf and public-deaf communication [10].

III. PROPOSED SYSTEM

The proposed model prioritizes a robust deep network architecture, leveraging Convolutional Neural Network (CNN) technology from deep learning in conjunction with OpenCV for the identification of numerical movements conveyed through hand gestures. In this approach, the hand motions are recorded using OpenCV, and the dataset is then trained using CNN. It is noteworthy that sign language, being non-universal, is a language understood by only a limited number of individuals.

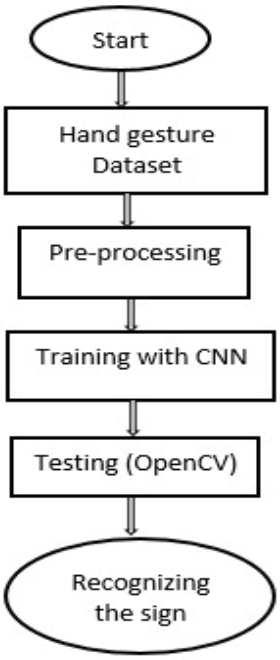

Fig. 2. Block diagram of proposed method

Benefits

- High compatibility with features
- Saving Time
- Minimal complexity

IV. DATA COLLECTION

The significance of data collection in this study cannot be overstated. We have meticulously curated our own American Sign Language (ASL) dataset, comprising 2000 images featuring ten stationary letter signs. The dataset encompasses static alphabets A, B, C, D, K, N, O, T, and Y, with two distinct datasets created by two different signers. Each signer performed each alphabetical motion 200 times under various lighting conditions, resulting in a comprehensive dataset. The dataset folder, containing images of alphabetic sign motions, has been subdivided into two additional folders: one for training and the other for testing. Out of the 2000 captured images, 1600 are allocated for training purposes, while the remaining images are designated for testing. To ensure consistency, a webcam is employed each time an image is captured against a consistent background upon receiving a command. The acquired images are saved as PNG files, and it is noteworthy that PNG format preserves image quality without loss during opening, closing, and saving operations. PNG's capability to handle complex, highly contrasted pictures is advantageous. The images are recorded by the webcam in the RGB colorspace..

V. SYSTEM ARCHITECTURE

In our research, hand gesture prediction and feature extraction from frames are accomplished through the deployment of a Convolutional Neural Network (CNN) model, recognized for its efficacy in image recognition within feedforward neural networks. The CNN architecture involves multiple layers, each convolution layer comprising a pooling layer, an activation function, and optional batch normalization, along with several fully connected levels. The sequential flow of information through the network causes a reduction in the size of one of the photos due to max pooling. The final layer of the network predicts class probabilities.

For the classification task, we implement a 2D CNN model utilizing the TensorFlow library. The convolution layers employ a 3 by 3 filter to scan images, computing the dot product between filter weights and frame pixels. This step extracts significant features from the input image. Subsequent to each convolution layer, pooling layers reduce the activation map's dimensionality, combining features from preceding layers to enhance the network's representation and mitigate overfitting. The activation function used in this instance is Rectified Linear Unit (ReLU), and the input layer of the CNN consists of 32 feature maps with a size of three by three. The largest pool layer has dimensions 2×2. The layer is then flattened, incorporating a dropout set at fifty percent. The network's final layer is a fully connected output layer with 10 units, employing SoftMax as the activation function. The model is constructed using Adam as the optimizer and categorical cross-entropy as the loss function..

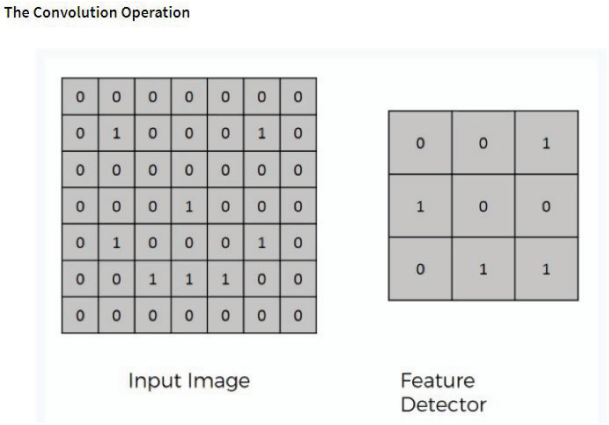

Fig. 3. Architecture

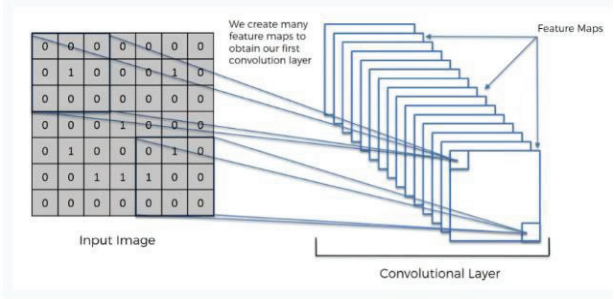

Fig. 4. Architecture CNN

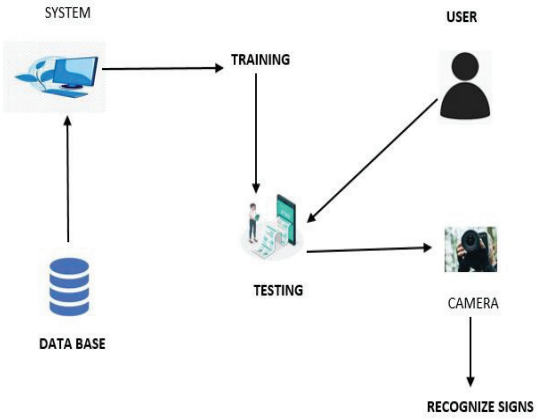

Fig. 5. System Architecture

VI. RESULTS

A. Model Performance:

The performed accuracy and metrics display the effectiveness of the CNN version in recognizing hand signs. The confusion matrix and sophisticationparticular metrics provide insights into the version's strengths and potential regions for improvement.

B. Challenges and Limitations:

The version may additionally face demanding situations in accurately spotting certain complicated hand symptoms, especially in situations with low lights or numerous hand orientations. Strategies consisting of facts augmentation and extra version complexity may be explored to address those barriers.

C. Future Improvements:

Future paintings can consciousness on refining the model structure, incorporating greater diverse datasets, and exploring advanced techniques together with transfer studying. Fine-tuning hyperparameters and optimizing the schooling technique can make a contribution to in addition upgrades in model performance.

D. Real-international Applications:

The a hit implementation of a hand signal popularity gadget the use of CNNs has considerable implications for actual-global packages, such as human-laptop interaction, accessibility, and verbal exchange gadgets for individuals with speech impairments.

E. Conclusion:

In summary, the CNN-based entirely hand sign reputation version exhibits excellent performance and accuracy. The study lays the groundwork for future developments in gesture-based communication systems, opening doors for better accessibility and humanmachine interaction.

VII. DISCUSSION

In this paper, we effectively utilized a Convolutional Neural Network (CNN) to predict hand gestures and extract features from frames. The CNN's architecture, incorporating convolution and pooling layers, proved adept at feature extraction and mitigating overfitting. By employing Rectified Linear Unit (ReLU) and dropout, the network demonstrated robustness, showcasing its potential for real-time gesture recognition applications.

VIII. CONCLUSION

In conclusion, our research employs Convolutional Neural Networks (CNN) for precise recognition of numeric hand gestures. Through dedicated training on a numerical sign dataset, our model achieves enhanced accuracy. The real-time implementation using OpenCV showcases the practical effectiveness of our CNN-based approach, contributing to improved communication in sign language.

IX. PROBLEMS

Sign languages exhibit significant variations in body language, facial expressions, and movements across different nations, accompanied by notable differences in sentence structure and grammar. Our study encountered challenges in learning and documenting gestures due to the precision and deliberation required in hand movements. Some intricate movements proved difficult to replicate, and the task of creating a dataset posed additional challenges, particularly in achieving consistent hand postures.

X. FUTURE WORK

In future applications, this methodology can extend its utility to recognize additional signs, including those featuring numerical values or signs outside the scope of the English language.

REFERENCES

[1] Garcia, B., & Viesca, S. A. (2016). Real-time American Sign Language recognition with convolutional neural networks. Convolutional Neural Networks for Visual Recognition, 2, 225-232.

[2] Van den Bergh, M., & Van Gool, L. (2011, January). Combining RGB and ToF cameras for realtime 3D hand gesture interaction. In 2011 IEEE workshop on applications of computer vision (WACV) (pp. 66-72). IEEE.

[3] Liwicki, S., & Everingham, M. (2009, June). Automatic recognition of finger spelled words in British sign language. In 2009 IEEE computer society conference on computer vision and pattern recognition workshops (pp. 50-57). IEEE.

[4] Zafrulla, Z., Brashear, H., Starner, T., Hamilton, H., & Presti, P. (2011, November). American Sign Language recognition with the Kinect. In Proceedings of the 13th international conference on multimodal interfaces (pp. 279-286).

[5] Pugeault, N., & Bowden, R. (2011, November). Spelling it out: Real-time ASL fingerspelling recognition. In 2011 IEEE International conference on computer vision workshops (ICCV workshops) (pp. 1114-1119). IEEE.

[6] Kuznetsova, A., Leal-Taixé, L., & Rosenhahn, B. (2013). Real-time sign language recognition using a consumer depth camera. In Proceedings of the IEEE International Conference on Computer Vision Workshops (pp. 83-90).

[7] Dong, C., Leu, M. C., & Yin, Z. (2015). American Sign Language alphabet recognition using Microsoft Kinect. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops (pp. 44-52).

[8] Singha, J., & Das, K. (2013). Hand gesture recognition based on Karhunen-Loeve transform. arXiv preprint arXiv: 1306.2599.

[9] Sharma, R., Nemani, Y., Kumar, S., Kane, L., & Khanna, P. (2013, July). Recognition of single handed sign language gestures using contour tracing descriptor. In Proceedings of the World Congress on Engineering (Vol. 2, pp. 3-5).

[10] Starner, T., Weaver, J., & Pentland, A. (1998). Real-time American Sign Language recognition using desk and wearable computer based video. IEEE Transactions on pattern analysis and machine intelligence, 20(12), 1371-1375.