International Journal of Innovative Research in Engineering
Volume 6, Issue 2(March-April 2025), PP: 90-93.
https://www.doi.org/10.59256/ijire.20250602010
www.theijire.com
ISSN No: 2582-8746

Sign Language Recognition Using Convolutional Neural Networks and Vgg19

Ramesh G[1], Hanish Ram B[2], Ajaykumar M[3]
[1,2,3]Department of Information Technology, K. L. N. College of Engineering, Madurai, Tamilnadu, India.

How to cite this paper:
Ramesh G[1], Hanish Ram B[2], Ajaykumar M[3] 'Sign Language Recognition Using Convolutional Neural Networks and Vgg19", IJIRE-V6I02-90-93.

Abstract: *Sign language is an expressive visual language that uses the hand and face and is an important tool for deaf individuals to communicate. Standard means of recognition objectives cannot recognize complex gestures effectively, in turn limiting accessibility and transactions of communication effectively. This project seeks to close communication barriers by translating hand gestures into speech, text, and even serve as an output to virtual assistants. To date, prior research in signing recognition has focused primarily on identifiable hand gestures, often selecting a subset of Indian Sign Language (ISL) for classification. In this work, a deep learning method is proposed to perform static sign recognition through Convolutional Neural Networks (CNN) with the VGG19 architecture. The system is trained through a dataset of ISL gestures for predicting and classifying signs with high confidence. The predicted sign is converted to a text output, to then a speech output, stored as an audio file. In conclusion, the system provides accessibility through detection, enabling a user user experience for both signers/non-signers to communicate and add encryption with accessibility. The model produces a far more particular/accurate performance in sign language recognition, and follows through with promising assistive technology solutions.*

Key Word: *Sign Language Recognition, VGG19, Convolutional Neural Network, Deep Learning, Indian Sign Language, Gesture Recognition.*

I.INTRODUCTION

Virtual Assistants and Cortana are today indispensable parts of daily life, but restricted to voice-only command inputs and not provided to the Deaf-Mute community. The objective of this project is to fill the gap between these two by combining Hand Gesture Recognition and Virtual Assistants so that deaf and speech-disabled individuals can interact with computer hardware comfortably. Sign language is a natural sign language utilized in the hearing-impaired through hand movement and facial expressions. Despite research on SLR with deep learning having surfaced across the world, Indian Sign Language recognition lacks development. SLR has poor accuracy for standard machine learning as features have to be manually extracted, whereas deep models can learn suitable features enhancing recognition performance. Through the creation of a real-time ISL recognition system that translates hand movement into voice and text outputs.

II.LITERATURE SURVEY

[1] Sign Language and Linguistic Universals – W. Sandler and D. Lillo-Martin Sandler and Lillo-Martin (2006) discuss how sign languages function as natural languages with complex grammatical structures, reinforcing their status as fully developed linguistic systems. Published by Cambridge University Press, this work significantly contributes to the understanding of sign languages within the broader field of linguistics.

[2] SF-Net: Structured Feature Network for Continuous Sign Language Recognition – Z. Yang, Z. Shi, X. Shen, and Y.-W. Tai Yang et al. (2019) proposed SF-Net, a Structured Feature Network designed for continuous sign language recognition. The model enhances feature extraction and representation, improving recognition accuracy in sign language sequences. Utilizing deep learning techniques, the approach effectively captures spatial and temporal dependencies in sign gestures. This study, published as an arXiv preprint, contributes to advancing automatic sign language recognition systems.

[3] Continuous Sign Language Recognition: Towards Large Vocabulary Statistical Recognition Systems Handling Multiple Signers – O. Koller, J. Forster, and H. Ney Koller et al. (2015) explored continuous sign language recognition, focusing on developing large- vocabulary statistical recognition systems capable of handling multiple signers. Their approach leverages computer vision and machine learning techniques to enhance recognition accuracy. The study addresses challenges in signer variability and large-scale vocabulary modeling. Published in Computer Vision and Image Understanding, this research significantly contributes to automatic sign language recognition advancements.

[4] How Many People Use ASL in the United States? Why Estimates Need Updating – R.E. Mitchell, T. A. Young, B. Bachelda, and M. A. Karchmer Mitchell et al. (2006) examined the number of American Sign Language (ASL) users in the United States, highlighting the need for updated estimates. Their study discusses methodological challenges in counting ASL users and the implications for linguistic and policy research. The findings emphasize the growing ASL-using population and the importance of accurate demographic data. Published in Sign Language Studies, this research provides valuable insights into ASL usage and its societal impact.

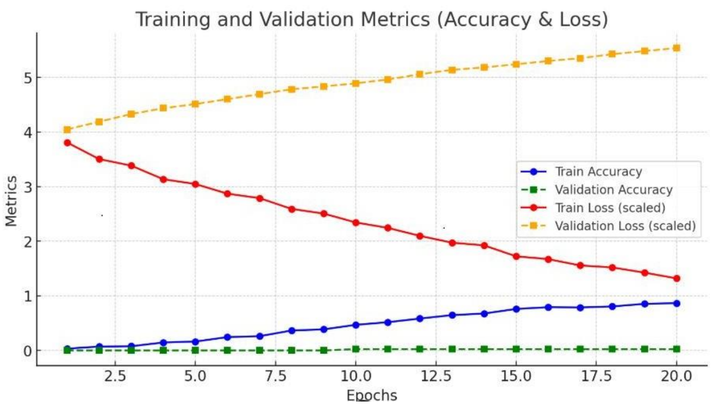
III.METHODOLOGY

3.1 Dataset

The dataset used in this project consists of a large collection of Indian Sign Language (ISL) static hand gesture images captured under various conditions. The dataset includes multiple samples of each sign to account for variations in hand shapes, skin tones, lighting conditions, and backgrounds. Each image is labeled with its corresponding sign to facilitate supervised learning. To enhance model robustness, data augmentation techniques such as rotation, flipping, and brightness adjustments are applied. The dataset is structured to ensure high-quality training and validation, making it suitable for deep learning-based sign language recognition.

3.2 Data Acquisition

The dataset is collected from publicly available ISL databases and self-captured images using a high-resolution camera. The images are acquired under diverse lighting conditions, camera angles, and user hand postures to improve generalization. The collected images are converted into a standardized format and stored for training and testing. The pre-processing phase deals with resolving missing or corrupted data, filtering out low-quality images, and ensuring uniform dimensions. After cleaning, the dataset is split into training (80%) and testing (20%) sets to train and validate the deep learning model effectively.



This analysis highlights the impact of training on sign language recognition using deep learning. The graph illustrates the variations in training accuracy, validation accuracy, training loss, and validation loss over the model.

3.3 Data Preprocessing

The data preprocessing stage is essential for improving the recognition accuracy of the sign language model. Several steps are followed to prepare the dataset:

- **Resizing:** All images are resized to a fixed resolution to ensure consistency across the dataset.
- **Grayscale Conversion:** Images are converted to grayscale to reduce computational complexity while retaining essential features.
- **Noise Removal:** Various filtering techniques are applied to eliminate background noise and unwanted artifacts.
- **Data Augmentation:** Techniques such as flipping, zooming, and rotation are applied to improve the model's generalization ability.
- **Normalization:** Pixel values are scaled between [0,1] to enhance convergence during training.

The effectiveness of preprocessing is validated through image quality assessment metrics and model accuracy improvements. The refined dataset ensures optimal feature extraction and recognition.

3.4 Vgg19 Model Architecture

The deep learning model used for sign language recognition is VGG19, a Convolutional Neural Network (CNN) architecture known for its deep hierarchical feature extraction capabilities. The model consists of 19 layers, including convolutional, pooling, and fully connected layers.

- **Convolutional Layers:** Extract spatial features such as edges, contours, and shapes from hand gesture images.
- **Pooling Layers:** Reduce dimensionality while retaining key features for efficient processing.
- **Fully Connected Layers:** Perform classification to determine the corresponding sign for each input gesture.
- **Softmax Activation:** Generates probability distributions to identify the most likely sign.

The model is trained using Categorical Cross-Entropy loss function and optimized with Adam optimizer to minimize classification errors and improve recognition accuracy.

3.5 Sign Language Translation System

The trained VGG19 model classifies hand gestures and converts them into textual outputs, which are further transformed into speech using Text-to-Speech (TTS) technology. The generated speech output is stored as audio files (.mp3 format) for real-time communication. This feature enables individuals with hearing impairments to interact seamlessly with non-signers through voice- controlled virtual assistants.

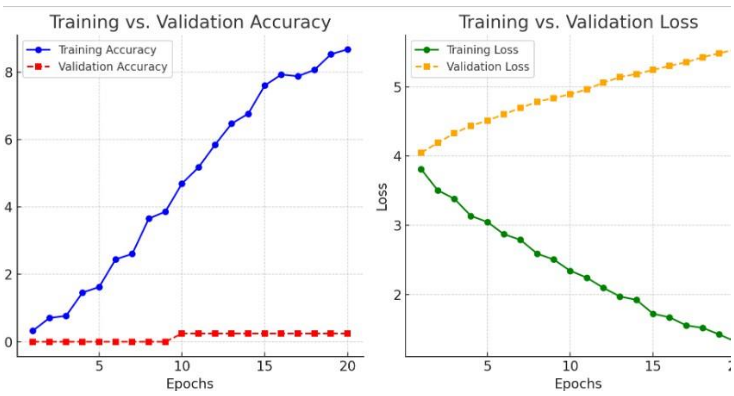
IV.SIGN LANGUAGE RECOGNITION

The proposed system aims to enable real-time Indian Sign Language (ISL) recognition using deep learning-based computer vision. The VGG19 model is optimized to process hand gestures efficiently and translate them into textual and speech outputs. The system integrates live video feed processing, ensuring fast and accurate classification of gestures. By utilizing real-time frame extraction, pre-processing, and classification, the model enhances accessibility for individuals with hearing impairments. The model is further fine-tuned using transfer learning and data augmentation, ensuring robustness across diverse lighting conditions and hand orientations.

V.SYSTEM DEPLOYMENT AND IMPLEMENTATION

The trained model is deployed in an interactive user-friendly application, enabling seamless sign language communication. The system is integrated with voice-controlled virtual assistants, allowing users to perform gestures that are converted into text and speech in real time. The application is designed to run on various platforms, including desktop, web, and mobile devices, making it widely accessible. Cloud-based implementation is considered for scalability, ensuring the model can process large datasets while maintaining performance efficiency.

VI.RESULT AND PERFORMANCE ANALYSIS



The trained model is deployed in an interactive user-friendly application, enabling seamless sign language communication. The system is integrated with voice-controlled virtual assistants, allowing users to perform gestures that are converted into text and speech in real time. The application is designed to run on various platforms, including desktop, web, and mobile devices, making it widely accessible. Cloud-based implementation is considered for scalability, ensuring the model can process large datasets while maintaining performance efficiency

VII.FUTURE ENHANCEMENT

The trained model is deployed in an interactive user-friendly application, enabling seamless sign language communication. The system is integrated with voice-controlled virtual assistants, allowing users to perform gestures that are converted into text and speech in real time. The application is designed to run on various platforms, including desktop, web, and mobile devices, making it widely accessible. Cloud-based implementation is considered for scalability, ensuring the model can process large datasets while maintaining performance efficiency.

VIII.CONCLUSION

The trained model is deployed in an interactive user-friendly application, enabling seamless sign language communication. The system is integrated with voice-controlled virtual assistants, allowing users to perform gestures that are converted into text and speech in real time. The application is designed to run on various platforms, including desktop, web, and mobile devices, making it widely accessible. Cloud-based implementation is considered for scalability, ensuring the model can process large datasets while maintaining performance efficiency.

References

1. *W. Sandler and D. Lillo-Martin, Sign language and linguistic universals. Cambridge University Press, 2006.*
2. *Z. Yang, Z. Shi, X. Shen, and Y.-W. Tai, "Sf-net: Structured feature network for continuous sign language recognition," arXiv preprint arXiv: 1908.01341, 2019.*
3. *O. Koller, J. Forster, and H. Ney, "Continuous sign language recognition: Towards large vocabulary statistical recognition systems handling multiple signers," Computer Vision and Image Understanding, vol. 141, pp. 108–125, 2015.*

4. R. E. Mitchell, T. A. Young, B. BACHELDA, and M. A. Karchmer, "How many people use asl in the united states? why estimates need updating," Sign Language Studies, vol. 6, no. 3, pp. 306–335, 2006.
5. D. Bragg, O. Koller, M. Bellard, L. Berke, P. Boudrealt, A. Braffort, N. Caselli, M. Huenerfauth, H. Kacorri, T. Verhoef et al., "Sign language recognition, generation, and translation: An interdisciplinary perspective," arXiv preprint arXiv:1908.08597, 2019.
6. G. T. Papadopoulos and P. Daras, "Human action recognition using 3d reconstruction data," IEEE Transactions on Circuits and Systems for Video Technology, vol. 28, no. 8, pp. 1807–1823, 2016.
7. C. Padden, "Verbs and role-shifting in american sign language," in Proceedings of the fourth national symposium on sign language research and teaching, vol.National Association of the Deaf Silver Spring, MD, 1986, p. 57.
8. K. Emmorey, "Space on hand: The exploitation of signing space to illustrate abstract thought." 2001.
9. H. Cooper, B. Holt, and R. Bowden, "Sign language recognition," in Visual Analysis of Humans. Springer, 2011, pp. 539–562.
10. F. Ronchetti, F. Quiroga, C. A. Estrebou, L. C. Lanzarini, and A. Rosete, "Lsa64: an argentinian sign language dataset," in XXII Congreso Argentino de Ciencias de la Computacion (CACIC 2016). ´ , 2016.
11. M. W. Kadous et al., "Machine recognition of auslan signs using powergloves: Towards large-lexicon recognition of sign language," in Proceedings of the Workshop on the Integration of Gesture in Language and Speech, vol. 165, 1996.
12. C. Wang, Z. Liu, and S.-C. Chan, "Superpixel-based hand gesture recognition with kinect depth camera," IEEE transactions on multimedia, vol. 17, no. 1, pp. 29–39, 2014.
13. G. D. Evangelidis, G. Singh, and R. Horaud, "Continuous gesture recognition from articulated poses," in European Conference on Computer Vision. Springer, 2014, pp. 595–607.
14. J. Zhang, W. Zhou, C. Xie, J. Pu, and H. Li, "Chinese sign language recognition with adaptive hmm," in 2016 IEEE International Conference on Multimedia and Expo (ICME). IEEE, 2016, pp. 1–6.
15. Jungpil shin yuto akiba, koki hirooka 1,(member, ieee)(graduate student member, ieee), najmul hassan (graduate student member, ieee), and yong seok hwang