NavaJyoti, International Journal of Multi-Disciplinary Research
Volume 9, Issue 1, June 2025

Sign Language Recognition using Deep Learning Models

Ayesha Banu[1], S Irene Getzi[2]

[1]Student, Department of Computer Science(PG), Jyoti Nivas College Autonomous, Bangalore, India
[2] Associate Professor, Department of Computer Science(PG), Jyoti Nivas College Autonomous, Bangalore, India

Abstract:

Sign language recognition is a crucial application of deep learning, improving communication accessibility for individuals with hearing or speech impairments. It plays a vital role in breaking down communication barriers, enabling seamless interaction between sign language users and non-signers across various settings such as educational institutions, workplaces, healthcare, and public services. Traditional sign language recognition systems often struggle with challenges such as variations in hand positioning, background noise, and occlusions, while others require expensive hardware, limiting accessibility. This study focuses on classifying American Sign Language (ASL) hand gestures using a deep learning approach. A dataset containing 35 ASL hand gesture images is preprocessed and used to train a convolutional neural network (CNN) based on the VGG16 architecture. Data augmentation techniques are applied to improve the model's robustness against variations in lighting, hand orientation, and occlusions. To improve generalization and reduce reliance on extensive labeled data, transfer learning is employed, with a pre-trained VGG16 model serving as a feature extractor. Fully connected layers are added for classification, and the model is fine-tuned using categorical cross-entropy loss and the Adam optimizer. Early stopping and model checkpointing are implemented to prevent overfitting and optimize performance. The model achieves high accuracy in ASL gesture recognition, demonstrating the effectiveness of deep learning in overcoming sign language recognition challenges and enhancing communication accessibility.

Keywords:
Sign Language Recognition, Deep Learning, Convolutional Neural Networks, Transfer Learning, VGG16.

Introduction

Communication is an essential aspect of human interaction, yet individuals with hearing or speech impairments often face significant challenges in engaging with those who do not understand sign language. Sign language recognition (SLR) systems aim to bridge this communication gap by translating hand gestures into meaningful text or speech, facilitating seamless interaction between signers and non-signers.

Traditional SLR systems rely on handcrafted feature extraction techniques, which struggle with variations in hand positioning, lighting conditions, background noise, and occlusions. Additionally, some methods also use proprietary machines, which are highly expensive. With the introduction of deep learning, especially CNNs, SLR has significantly improved because it is now possible to automate feature extraction from images and videos with high accuracy.

The goal of this research is to create a system that recognizes ASL hand gestures by employing deep learning techniques. A CNN model based on VGG16 architecture is trained with a preprocessed dataset of 35 American Sign Language (ASL) hand gesture images. To increase the model's performance and generalization, the pre-trained VGG16 model is used as a feature extractor with fully connected layers added for classification. To make the model more robust to changes in lighting, occlusion, and hand orientation, data augmentation strategies are applied. The model is trained with categorical cross entropy loss function and Adam optimizer. Early stopping and model checkpointing are implemented to avoid overfitting.

By using deep learning methods, this study shows how well CNN-based approaches work to make ASL recognition more accurate, overcoming the limitations of traditional methods, and enhancing accessibility for individuals with hearing or speech impairments. The findings contribute to the development of more efficient, scalable, and accessible sign language recognition systems.

2.Literature Review

Sign language recognition (SLR) has been an area of active research for many years. Researchers have developed different ways to help signers and non-signers communicate better. Traditional methods relied on handcrafted feature extraction and rule-based algorithms, while modern deep learning methods have significantly improved recognition accuracy and robustness.

- Traditional Approaches to Sign Language Recognition

Early SLR systems used sensor-based techniques and computer vision. Sensor-based methods used gloves or motion trackers to capture hand movements. But these needed special hardware, which made them costly and hard to get. Computer vision methods used image processing techniques, such as edge detection and skin color segmentation, but they struggled with variations in lighting, background noise, and occlusions

- Deep Learning in SLR

Deep learning, particularly Convolutional Neural Networks (CNNs), has significantly improved SLR accuracy. Mittal & Jain (2023) demonstrated a CNN-based model for ASL recognition using datasets like MNIST and Grassknoted ASL, achieving high accuracy. Transfer learning with pre-trained models such as ResNet, and InceptionV3 further enhances performance by leveraging knowledge from large datasets, making models more adaptable to real-world conditions.

- Research Gaps and Contributions

Despite advancements, challenges remain in real-world SLR applications due to environmental variations. This study employs VGG16-based transfer learning and data augmentation to develop a more efficient and scalable SLR model, improving communication accessibility for individuals with hearing or speech impairments.

3. Methodology

This study focuses on classifying American Sign Language (ASL) hand gestures using a deep learning-based approach. The methodology consists of **data collection and preprocessing, model selection, training, and evaluation** to develop an efficient sign language recognition system.

3.1 Dataset and Preprocessing:

The dataset used for training and testing consists of 35 distinct ASL hand gesture classes (A-Z and 0-8), with multiple images per class to ensure effective training. Several preprocessing steps are applied to prepare the data for optimal model performance. These steps include **resizing**, where the images are standardized to a consistent size to ensure uniformity. **Normalization** is then applied to scale the pixel values to a specific range (typically 0 to 1), improving the learning efficiency and stability of the model. Additionally, **data augmentation** techniques, such as random rotation, flipping, and brightness adjustments, are used to increase the robustness of the model. These augmentations simulate variations in hand orientation, lighting conditions, and potential occlusions, enabling the model to generalize better to real-world variations in ASL hand gestures.

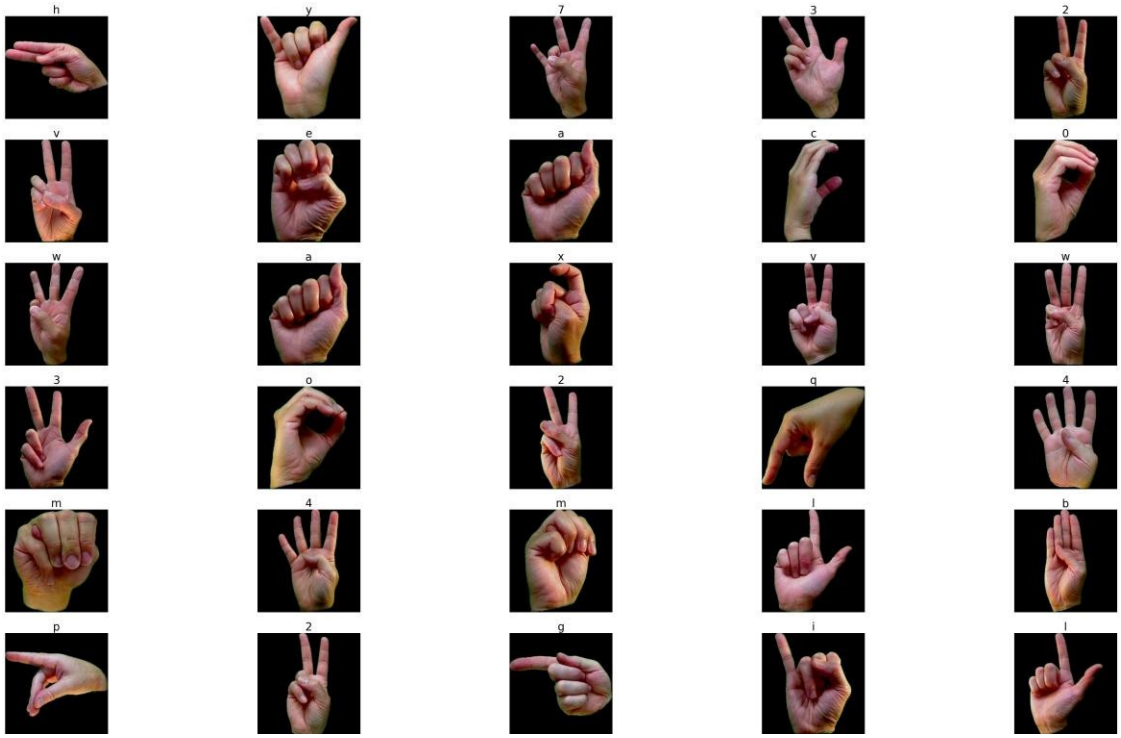


Fig.1 Dataset

3.2 Model Selection and Training:

A Convolutional Neural Network (CNN) based on the VGG16 architecture is employed for feature extraction. Transfer learning is utilized, leveraging a pre-trained VGG16 model to benefit from the knowledge acquired from large datasets. The classification layers of the model are fine-tuned and trained using the **Categorical Cross-Entropy** loss function, which is suitable for multi-class classification tasks. The **Adam** optimizer is used for training, offering an efficient and effective approach to updating the model's weights. The model's performance is evaluated using **accuracy** and **loss** metrics on both the training and validation sets. To mitigate the risk of overfitting, **early stopping** is applied, halting the training when the

validation loss stops improving, and **model checkpointing** is used to save the best-performing model during training based on validation performance.

3.3 Evaluation and Testing:

The trained model is evaluated on a separate test set to measure accuracy and generalization. Performance is compared against baseline models to validate the effectiveness of the approach.

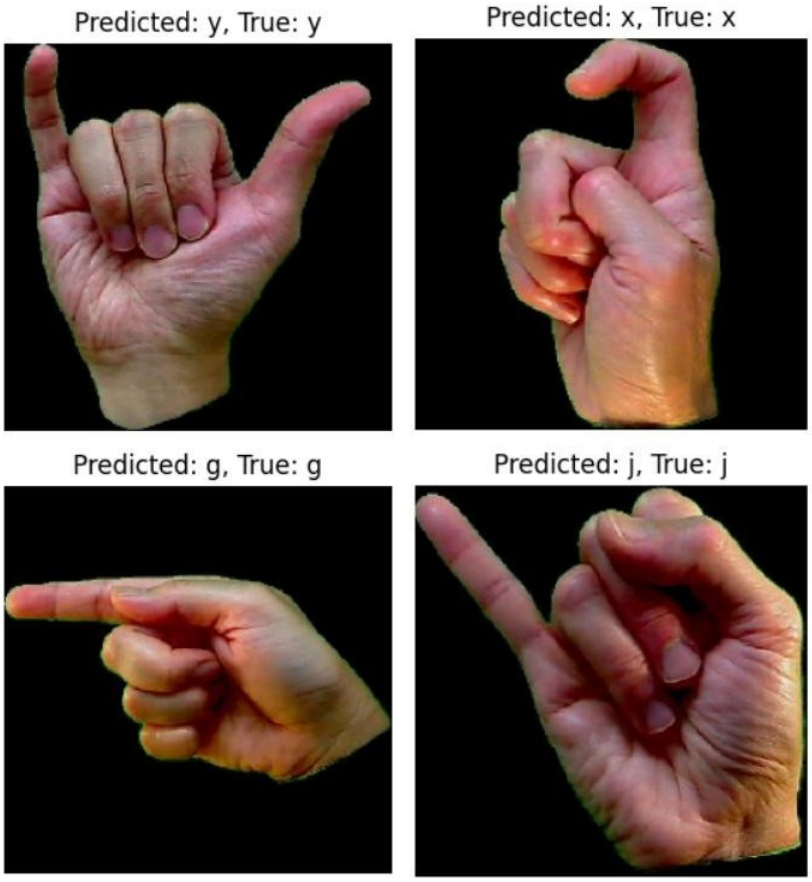


Fig.2. Prediction result

4. Results

The model achieved 97% accuracy on the test set. It performed well in recognizing most ASL gestures. The results show that using VGG16 and transfer learning improved recognition accuracy.

5. Conclusion

This study highlights the effectiveness of deep learning, particularly VGG16-based transfer learning, in recognizing American Sign Language (ASL) hand gestures. By leveraging a pre-trained CNN model and applying data augmentation, the system achieves high accuracy while remaining robust against variations in lighting, hand orientation, and occlusions. Techniques like early stopping and model checkpointing further enhance performance by preventing overfitting.

Future research can focus on integrating real-time detection, expanding the dataset, and exploring hybrid models for improved recognition. This study contributes to enhancing **communication accessibility** for individuals with hearing or speech impairments, reinforcing the importance of AI-driven solutions in assistive technologies.

The proposed sign language recognition system achieves high accuracy using VGG16-based transfer learning method along with data augmentation techniques. However, with future improvements, these systems will be far more applicable in real life. One key area is the real-time gesture recognition that allows the user to interact seamlessly using live video streams. Extending the datasets to include more gestures, deeper backgrounds, and multiple sign languages could further influence the model's generalization. In addition, hybrid deep learning models, such as CNN-LSTM or CNN-Transformers, could enhance sequential gesture modeling. Optimization for mobile and edge devices will make it easier to access without requiring high-end equipment. Multimodal inputs that include facial expressions and hand movement tracking integrated in the application could boost the recognition rate. These enhancements would make the system more scalable, efficient, and beneficial for bridging communication gaps for hearing and speech-impaired persons.

References

1. Mittal, S., & Jain, S. (2023). Recognition of Static American Sign Language Hand Gestures Using Convolutional Neural Networks.
2. Murali, R. S. L., Ramayya, L. D., & Santosh, V. A. (2022). Sign language recognition system using convolutional neural network and computer vision. *Int. J. Eng. Innov. Adv. Technol*, *4*, 138-141.
3. Kothadiya, Deep, Chintan Bhatt, Krenil Sapariya, Kevin Patel, Ana-Belén Gil-González, and Juan M. Corchado. "Deepsign: Sign language detection and recognition using deep learning." *Electronics* 11, no. 11 (2022): 1780.
4. Verma, Prashant, and Khushboo Badli. "Real-Time Sign Language Detection using TensorFlow, OpenCV and Python." International Journal for Research in Applied Science & Engineering Technology (IJRASET)10, no. V (2022): May2022.
5. Adeyanju, Ibrahim Adepoju, Oluwaseyi Olawale Bello, and Mutiu Adesina Adegboye. "Machine learning methods for sign language recognition: A critical review and analysis." Intelligent Systems with Applications12 (2021): 200056.
6. Khan, R.Z. and Ibraheem, N.A. Hand gesture recognition: a literature review. International journal of artificial Intelligence & Applications, 3(4), p.161.