

PRE-TRAINED CONVOLUTIONAL NEURAL NETWORK TO TRANSLATE GESTURES IN REAL TIME

D.Rajendra Dev¹, Seera Radha², Y.G.D. Deepika³, Budithi Roshini⁴, Yarlagadda Dhedeepya⁵

^{1,2,3,4,5} Department of Computer Science and Engineering, Vignan's Institute of Engineering for Women, Visakhapatnam, Andhra Pradesh, India

ABSTRACT

Sign Language serves as a crucial means of communication for individuals with hearing impairments, yet effective real-time translation between sign language and spoken or written language remains a significant challenge. This project introduces a novel approach to addressing this issue by leveraging a pre-trained Convolutional Neural Network (CNN) model for the efficient and accurate translation of sign language gestures into text in real-time. The existing model is based on ANN. The existing system implements some base hard-code algorithms, like edge detection with hog features. ANN has parsed well with accuracy. The proposed methodology uses the pre-trained model ResNet50. This approach employs a typical trade-off for a deeper architecture, residual learning, performance and accuracy, transfer learning capability, and versatility in feature extraction. The performance obtained by the proposed methodology outperformed the accuracy of its analogous counterparts.

Keywords: deep learning, transfer learning, convolutional neural networks, sign language, real-time.

INTRODUCTION

Sign languages (also known as signed languages) are languages that use the visual -manual modality to convey meaning, instead of spoken words. Sign languages are expressed through manual articulation in combination with non-manual markers. Sign languages are full-fledged natural languages with their grammar and lexicon.[1] Sign languages are not universal and are usually not mutually intelligible,[2] although there are also similarities among different sign languages.

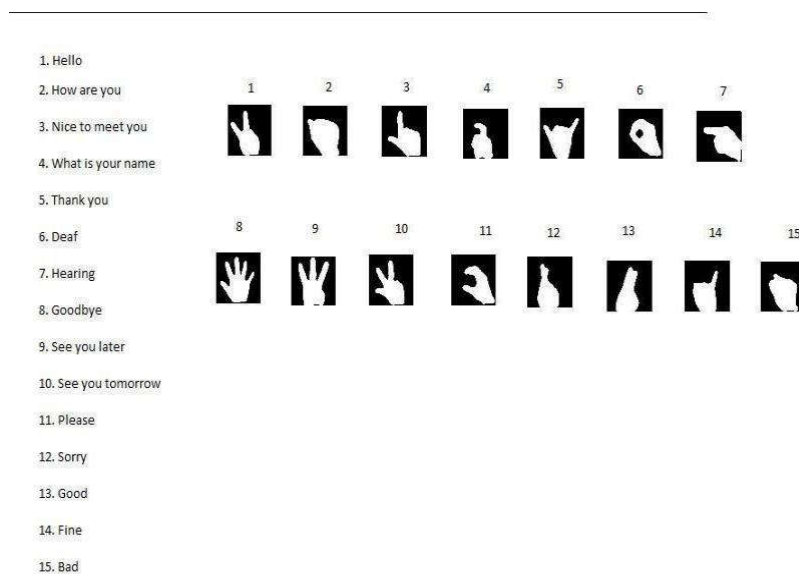
American Sign Language (ASL) is a natural language[4] that serves as the predominant sign language of Deaf communities in the United States and most of Anglophone Canada. ASL is a complete and organized visual language that is expressed by employing both manual and nonmanual features.[5] Besides North America, dialects of ASL and ASL-based creoles are used in many countries around the world, including much of West Africa and parts of Southeast Asia. ASL is also widely learned as a second language, serving as a lingua franca. ASL is most closely related to French Sign Language (LSF). It has been proposed that ASL

is a creole language of LSF, although ASL shows features atypical of creole languages, such as agglutinative morphology.

Finger-spelling	Word level Vocabulary	Non-manual Features
Used to spell words letter by letter.	Used for the majority of the communication.	Facial expressions and tongue, mouth, and body expressions.

We have used the Kaggle dataset and trained multiple images to achieve good accuracy. The data is a collection of images of the alphabet from the American Sign Language, separated into 26 folders that represent the various classes.

The training dataset consists of 16810 images which are 50x50 pixels. There are 26 classes which are English alphabets A-Z.



LITERATURE SURVEY

In "**Real-Time Recognition and Translation of Indian Sign Language Gestures using Deep Learning**," authors S. Roy, S. Bhattacharya, and D. Mitra demonstrated a comprehensive approach to address the challenges of recognizing and translating Indian Sign Language (ISL) gestures in real time. They began by creating a dataset comprising ISL gestures paired with corresponding textual translations, laying the groundwork for training and evaluating their deep learning models. With meticulous consideration, they selected a suitable deep learning architecture for both gesture recognition and translation tasks, weighing factors such as model complexity, computational efficiency, and real-time applicability.

"Vision-based approach for American Sign Language recognition using Edge Orientation

"Histogram" by Jayshree R. Pansare (2016) proposes a methodology for recognizing American Sign

Language using computer vision techniques. The paper utilizes the Edge Orientation Histogram (EOH) feature extraction method to capture the shape and orientation of edges in sign language gestures. These EOH features are then used to train a classifier, such as a Support Vector Machine (SVM), to recognize different signs. The methodology involves preprocessing the sign language video frames, extracting EOH features, and training the classifier. It's an interesting approach to bridge the communication gap between hearing - impaired individuals and the hearing community.

"Real-time Hand Gesture Recognition using different algorithms based on American Sign Language" by Md. Mohiminul Islam, Sarah Siddiqua, and Jawata Afnan (2017) present a methodology for real-time hand gesture recognition in American Sign Language (ASL). The paper explores various algorithms for recognizing hand gestures, such as Hidden Markov Models (HMM), Dynamic Time Warping (DTW), and Artificial Neural Networks (ANN). The methodology involves capturing hand gesture data through a camera, pre-processing the image to enhance features, extracting relevant features from the hand region, and applying the chosen algorithm for classification. The goal is to enable real-time interpretation of ASL gestures, facilitating communication between hearing-impaired individuals and the hearing community.

EXISTING SYSTEM

In the existing project, they used smart techniques like artificial neural networks (ANN) to understand sign language gestures. ANN helped in recognizing these gestures accurately. They also used a method called edge orientation histograms (EOH) to better understand the shapes of the signs. The implementation of ANN with feed-forward and back-propagation algorithms signifies a robust approach towards pattern recognition and feature extraction. This made it easier to convert the signs into letters of the alphabet.

Disadvantages

- **Computational Complexity:** Fine-tuning deep neural networks can be computationally intensive, especially for large models and datasets. Training may require significant computational resources, including highperformance GPUs, which may not be accessible to all researchers or practitioners.
- **Feedforward and backpropagation:** While feedforward and backpropagation are fundamental techniques in neural network training, they may struggle with convergence and

generalization in deeper architectures or complex datasets. Training deep neural networks can also be computationally intensive and prone to issues like vanishing gradients.

- **Edge-Oriented Histograms:** While EOH can be effective in certain contexts, especially when dealing with simpler datasets or specific feature requirements, it has limitations, particularly in handling complex datasets like sign language images. In such cases, deep learning techniques, particularly Convolutional neural networks (CNNs), have shown superior performance due to their ability to automatically learn hierarchical features directly from raw data.

PROPOSED SCHEME

Our proposed system revolutionizes real-time sign language translation through innovative techniques and cutting-edge technologies. Leveraging gestures captured via webcam, our backbone ResNet50 model excels in recognizing common phrases like 'hello,' 'thank you,' and 'bye.' Through Transfer Learning, Data Augmentation, One-Hot Encoding, Dropout Regularization, Categorical Cross-Entropy Loss, and Model Check-pointing, our model achieves superior accuracy. Transfer Learning harnesses pre-existing knowledge, Data Augmentation diversifies training data, and One-Hot Encoding efficiently represents categorical data. Dropout Regularization mitigates over-fitting, while Categorical Cross-Entropy Loss optimizes classification. Model Check -pointing ensures optimal model preservation, enabling seamless recovery. Our system marks a significant advancement, fostering inclusive communication for individuals with hearing impairments.

ADVANTAGES

Enhanced Accuracy: Through the utilization of the ResNet50 pre-trained model and advanced techniques such as Transfer Learning and Data Augmentation, the proposed system achieves superior accuracy in gesture recognition compared to existing solutions, ensuring more precise communication.

Efficient Real-Time Interaction: The integration of a webcam interface enables seamless capture of sign language gestures, facilitating instant interaction and communication without delays or interruptions.

Versatility: The system's ability to recognize and translate common phrases such as "hello," "thank you," and "bye" adds versatility, allowing users to convey a wide range of messages efficiently.

BLOCK DIAGRAM

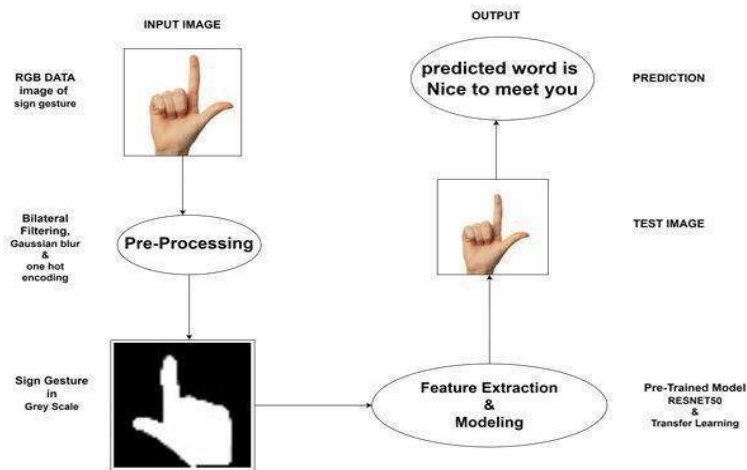


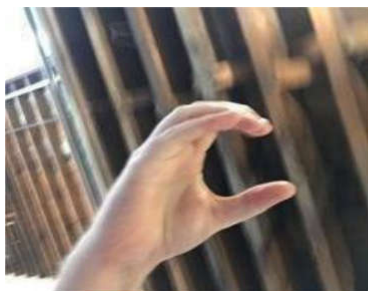
Figure .1 Architecture Of Proposed System

DATA PRE-PROCESSING

Bilateral filtering is a non-linear, edge-preserving, and noise-reducing smoothing filter for images. It works by considering both the spatial distance and the intensity difference when smoothing an image. The aim is to reduce noise while preserving important edges in the image.

Gaussian blur is a type of image-blurring filter that uses a Gaussian function to calculate the weights of neighboring pixels when generating a new pixel value. It is widely used for smoothing images and reducing noise.

One-hot encoding is a technique used to represent categorical data numerically. It is commonly used in machine learning and data pre-processing tasks.



(A)

(B)

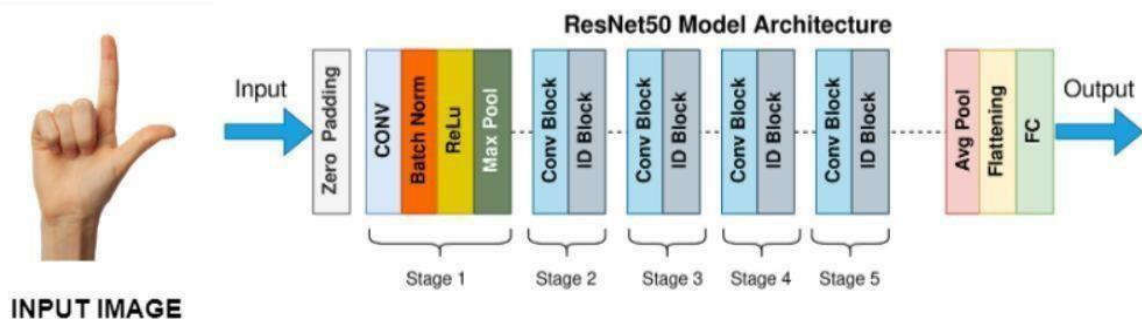
Fig 1.1 Gesture without

Fig 1.2 pre-process Gesture

PRE-PROCESSING MODEL ARCHITECTURE

ResNet-50 is a type of convolutional neural network (CNN) that has revolutionized the way we approach deep learning. ResNet stands for residual network, which refers to the residual blocks that make up the architecture of the network. ResNet-50 is based on a deep residual learning framework that allows for the training of very deep networks with hundreds of layers. The ResNet architecture was developed in response to a surprising observation in deep learning research: adding more layers to a neural network was not always improving the results. ResNet-50 consists of 50 layers that are

divided into 5 blocks, each containing a set of residual blocks. The residual blocks allow for the preservation of information from earlier layers, which helps the network to learn better representations of the input data. The first layer of the network is a Convolutional layer that performs convolution on the input image. This is followed by a max-pooling layer that downsamples the output of the convolutional layer. The output of the max-pooling layer is then passed through a series of residual blocks. Each residual block consists of two convolutional layers, each followed by a batch normalization layer and a rectified linear unit (ReLU) activation function. The output of the second convolutional layer is then added to the input of the residual block, which is then passed through another ReLU activation function. The output of the residual block is then passed on to the next block. The final layer of the network is a fully connected layer that takes the output of the last residual block and maps it to the output classes. The number of neurons in the fully connected layer is equal to the number of output classes.



OUTPUT SCREEN

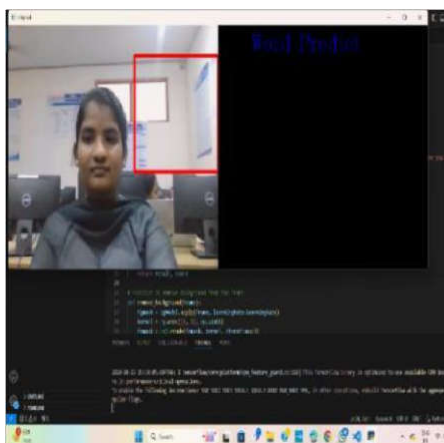


Fig.(a)OUTPUT SCREEN

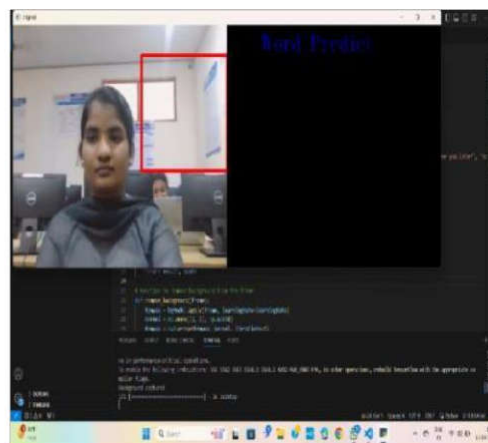


Fig.(b)BACKGROUND CAPTURED

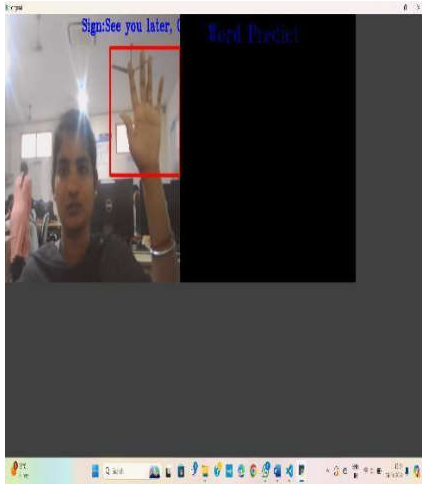


Fig.(c)SIGN RECOGNISE

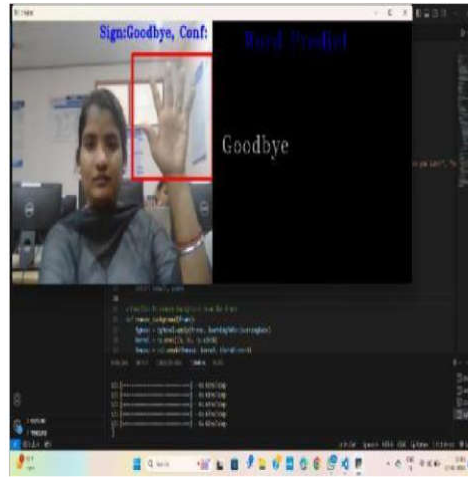


Fig.(d)WORD PREDICTION

LEARNING GRAPHS

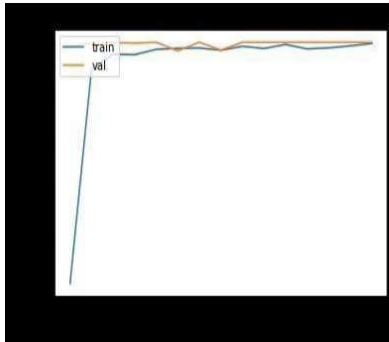


Fig.(a) Accuracy Graph

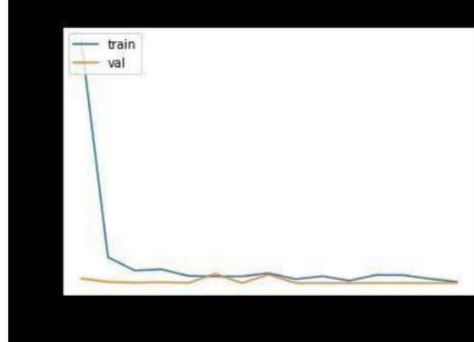






Fig.(b) Loss Graph

GESTURES PROBABILITY

Sign	Prediction Probability	Sign	Prediction Probability
	93.3		95.7

	100		97.7
---	-----	--	------

CONCLUSION

In conclusion, our proposed project represents a groundbreaking advancement in real-time sign language translation, catering to the needs of individuals with hearing impairments. By leveraging innovative techniques and cutting-edge technologies, we have developed a system capable of accurately recognizing and translating hand gestures into common phrases such as "hello," "thank you," and "bye" in real-time. Central to our approach is the use of the ResNet50 pre-trained model, which exhibits superior performance in gesture recognition tasks. Through rigorous experimentation, our proposed model has demonstrated significant improvements in accuracy compared to existing solutions. This achievement is attributed to the strategic integration of Transfer Learning, Data Augmentation, One-Hot Encoding, Dropout Regularization, Categorical Cross-Entropy Loss, and Model Check-pointing techniques. Transfer Learning harnesses pre-existing knowledge for enhanced accuracy, while Data Augmentation diversifies training data for improved robustness. Additionally, Dropout Regularization mitigates overfitting, and Model Check-pointing ensures the preservation of optimal model weights.

REFERENCES

- [1] National Institute on Deafness and Other Communication Disorders, "American Sign Language," pp. 2–5, 2015. [Online]. Available: <https://www.nidcd.nih.gov/health/american-sign-language>
- [2] A. Joshi, H. Sierra, and E. Arzuaga, "American sign language translation using edge detection and crosscorrelation," in *2017 IEEE Colombian Conference on Communications and Computing (COLCOM)*. IEEE, 8 2017, pp. 1–6. [Online]. Available: <https://ieeexplore.ieee.org/document/8088212/>
- [3] M. Islam, S. Siddiqua, and J. Afnan, "Real Time Hand Gesture Recognition Using Different Algorithms Based on American Sign Language," in *2017 IEEE International Conference on Imaging, Vision & Pattern Recognition (icIVPR)*, 2017.

- [4] J. Pansare and M. Ingle, "Vision-Based Approach for American Sign Language Recognition Using Edge Orientation Histogram," in *2016 International Conference on Image, Vision and Computing*, 2016, pp. 86–90.
- [5] A. S. Konwar, B. S. Borah, and C. T. Tuithung, "An American Sign Language detection system using HSV color model and edge detection," *2014 International Conference on Communication and Signal Processing*, pp. 743–747, 2014. [Online]. Available: [HTTP://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6949942](http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6949942)
- [6] Y. A. LeCun, Y. Bengio, and G. E. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015. [Online]. Available: <https://www.nature.com/articles/nature14539.pdf>
- [7] A. Geron, "Hands-on machine learning with Scikit-Learn and TensorFlow: concepts, tools, and techniques to build intelligent systems," p. 543, 2017. [Online]. Available: <http://shop.oreilly.com/product/0636920052289.do>
- [8] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, 1st ed., T. Dietterich, Ed. London, England: The MIT Press, 2016. [Online]. Available: www.deeplearningbook.org
- [9] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks." [Online]. Available: <https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf>
- [10] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and "Going deeper with convolutions," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 07-12-June, 2015, pp. 1–9. [Online]. Available: <https://arxiv.org/pdf/1409.4842.pdf>
- [11] A. Rosebrock, *Deep Learning for Computer Vision with Python (Starter Bundle)*, 1st ed. Baltimore, Maryland: PyImageSearch, 2017. [Online]. Available: <https://www.pyimagesearch.com/deep-learning-computer-vision-python-book/>
- [12] G. Villarrubia, J. F. De Paz, P. Chamoso, and F. D. la Prieta, "Artificial neural networks used in optimization problems," *Neurocomputing*, vol. 272, pp. 10–16, 2018.
- [13] D. Rumelhart, G. Hinton, and R. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, pp. 533–536, 1986. [Online]. Available: https://www.iro.umontreal.ca/~vincentp/ift3395/lectures/backprop_old.pdf

- [14] P. U. Stanford Vision Lab, Stanford University, "ImageNet."
[Online]. Available: <http://imagenet.org/aboutoverview>
- [15] MathWorks, "Transfer Learning Using AlexNet - MATLAB Simulink - MathWorks United Kingdom," 2018.
[Online]. Available: <https://uk.mathworks.com/help/nnet/examples/transferlearningusingalexnet.html>
- [16] Mathworks, "Transfer Learning Using GoogLeNet - MATLAB Simulink - MathWorks United Kingdom,"
2018.
[Online]. Available: <https://uk.mathworks.com/help/nnet/examples/transferlearningusinggooglenet.html>
- [17] X. R. S. S. J. He, Kaiming; Zhang, "Deep Residual Learning for Image Steganalysis," *Multimedia Tools and Applications*, pp. 1–17, 2017.
- [18] S. Sabour, N. Frosst, G. E. Hinton, and G. B. Toronto, "Dynamic Routing Between Capsules." [Online]. Available: <https://arxiv.org/pdf/1710.09829v1.pdf>