IJARSCT

ISSN (Online) 2581-9429

IJARSCT

Impact Factor: 7.53

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Volume 4, Issue 3, October 2024

Deep Learning-Driven Sign Language Recognition: A Multimodal Approach for Gesture-to-Text Translation Using CNN-RNN Architectures

Krishna Jitendra Jaiswal[1], Amaan Khan[2], Hitanshu Budhdev[3], Mr. Sonali Gandhi[4]

Students, Department of Computer Engineering[1,2,3]

Assistant Professor, Department of Computer Engineering[4]

Thakur College of Engineering and Technology, Mumbai, India

1032221316@tcetmumbai.in, 1032221318@tcetmumbai.in,

1032210401@tcetmumbai.in, sonali.gandhi@tcetmumbai.in

Abstract*: Sign language serves as a vital communication tool for individuals with hearing and speech impairments. However, a significant barrier exists when others do not possess a strong understanding of sign language often requiring interpreters to facilitate communication. To reduce the reliance on human interpreters, this research aims to develop an intelligent system capable of recognizing and translating sign language gestures into meaningful grammatically correct sentences. The proposed system processes both images and videos to interpret gestures, converting them into text that can be translated into any desired language. The system leverages state-of-the-art deep learning techniques, such as Convolutional Neural Networks (CNNs) for identifying key features in images and Recurrent Neural Networks (RNNs) for understanding temporal sequences in video data. By employing these advanced neural networks, the model is able to comprehend hand movements, facial expressions and other non-verbal cues to construct coherent sentences. Additionally, the system integrates natural language processing (NLP) to refine the output, ensuring the resulting sentences are grammatically correct. Our approach addresses common challenges such as differentiating between subtle hand gestures and reducing the impact of environmental noise in images. This solution holds the potential to significantly enhance communication for the hearing and speech impairedffering an efficient interpreter-free method of translating sign language into widely spoken languages.*

Keywords: Sign language recognition, deep learning, CNN, RNN, gesture detection, video processing, image processing, natural language processing, interpreter-free communication, neural networks

I. INTRODUCTION

Sign language is a visually-based form of communication that relies on hand movements and facial expressions to convey meaning. Small variations in hand gestures can significantly alter the message being communicated. It is equally important to mitigate the influence of background noise in images or videos to ensure accurate gesture recognition. Addressing these challenges, we propose a robust system for recognizing sign language. Although there are existing models that translate American Sign Language into text, many of these systems do not leverage deep learning techniques. Furthermore, they struggle to capture the intricate nuances found in different sign languages. To overcome these limitations we employ deep neural networks capable of identifying gestures and converting them into English sentences. This approach promises to enhance both the accuracy and versatility of sign language recognition.

DOI: 10.48175/IJARSCT-19946

II. BACKGROUND

Various methods have been explored to address the challenge of sign language recognition. However many of these approaches rely heavily on image processing using MATLAB followed by the application of traditional machine learning techniques. While some progress has been made through artificial neural networks, the majority of solutions lack sophistication. Image and video processing particularly with the rise of social media content has become a critical area of focus in both machine learning and deep learning. The techniques developed in these fields have the potential to significantly enhance the lives of individuals with hearing and speech impairments by providing tools that convert their sign language gestures into widely understood languages along with the conveyed emotions and tone.

Given the necessity of these systems ongoing research into creating more efficient and accurate algorithms for sign language interpretation will continue to expand. Deep learning offers far more potential than has been fully explored leaving ample room for the development of new models in this space. Such systems with slight modifications can be adapted to other domains including gesture-controlled interfaces and monitoring abnormal behaviors. Furthermore, they can be fine-tuned to recognize different sign languages beyond the ones commonly studied.

Looking forward this work could evolve to produce speech output in addition to text. Similar generative models could be applied to map text or speech into actions either by assembling pre-recorded videos or generating a skeletal framework. These models could easily be integrated into various platforms thereby boosting productivity across a range of applications.

III. LITERATURE SURVEY

Several methods have been explored to address the challenge of sign language recognition with many relying on image processing techniques implemented in MATLAB followed by the application of machine learning algorithms. While some work has utilized artificial neural networks these approaches often face limitations. Simulations for such tasks can vary in duration, ranging from hours to weeks depending on the complexity and scope.

One study [1] focuses on recognizing double-handed Indian Sign Language by capturing images processing them in MATLAB and converting the data into speech and text. However, this method is highly sensitive to changes in lighting conditions. Another approach [2] combines image processing, computer vision, and neural networks to identify hand characteristics from video recordings via a webcam. This system processes continuous frames using a series of image processing techniques and interprets hand signs with a Haar Cascade Classifier. The recognized text is then converted to speech using a synthesizer. A third approach [3] involves three distinct phases: training, testing, and recognition. During the training phase, a multiclass support vector machine (MSVM) is used to classify data with Hu invariant moments and structural shape descriptors creating a combined feature vector. In the testing phase, preprocessing is applied to extract features from input images.

Deep neural networks (DNNs) have transformed how machine learning problems are tackled, including sign language recognition. This project aims to leverage DNNs to achieve high accuracy in gesture recognition, significantly improving on previous methods by incorporating cutting-edge learning algorithms.

IV. RISK IDENTIFICATION

Several risks could impact the effectiveness of the sign language recognition system:

- Gestures might be misinterpreted if the system processes input too quickly without waiting for pauses that indicate the end of a sentence.
- Variations in facial expressions can change the meaning of a sentence making it essential to accurately capture these nuances
- Missing video frames due to poor data transfer or recording issues could lead to incomplete gesture recognition.
- Inadequate training could result in poor model performance.
- Incorrect or mislabeled training data may cause the model to learn inaccurately.
- Insufficient lighting during video capture could obscure hand gestures, reducing recognition accuracy

DOI: 10.48175/IJARSCT-19946

V. DESIGN

Some of the most important uses of deep learning and machine learning are incorporated into this project's design. Numerous ideas are employed including image and video processing, various neural network topologies, etc. The main considerations in the design and development of this application were simplicity and efficacy.

A. System Design

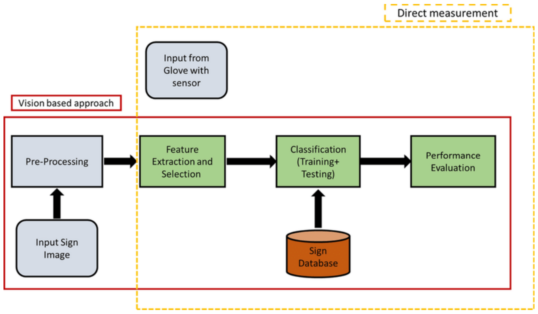


Fig.1.Architecture

Figure 1 gives a brief overview of the architecture that the Sign Language Recognition application follows.

Input module

A camera is utilized to capture the entire conversation. While individual letters and numbers can be treated as static frames, full words and sentences are captured in video format.

Image and Video Processing

Python and OpenCV are employed to handle the image and video processing tasks. Videos are broken down into multiple frames making it easier to detect and analyze the hand gestures.

Convolutional Neural Network

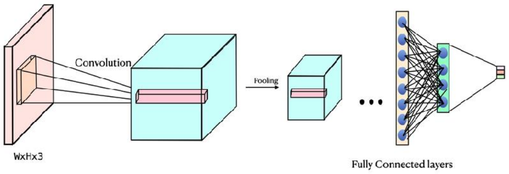
These CNNs a category of deep feed forward networks that utilize weight sharing are widely used in both image and video processing. In this project, CNNs are applied to the extracted frames to identify hand shapes and movements.

Recurrent Neural Network

RNNs are employed when the system needs to remember past inputs. In this case, RNNs are used to analyze a sequence of hand gestures by considering both the current frame and previous frames in the video.

Natural Language Processing

NLP techniques are incorporated to transform the output from the RNN into grammatically correct English sentences ensuring the generated text makes coherent sense.



VI. IMPLEMENTATION

A. Technology Introduction

Python

Python is a high-level, general-purpose programming language known for its simplicity and readability. Its syntax allows developers to express concepts in fewer lines of code, making it widely used across various domains.

OpenCV

OpenCV (Open Source Computer Vision Library) is a collection of functions aimed primarily at real-time computer vision tasks. In this project, OpenCV is used for recognizing hand gestures from image and video data.

CUDA

CUDA is a parallel computing platform and API developed by Nvidia. It enables developers to harness the power of a CUDA-enabled GPU for accelerating general-purpose computations, including the processing of large datasets.

PyTorch

PyTorch is an open-source machine learning framework in Python that is widely used for building, training, and testing deep learning models, including neural networks.

B. Algorithm

Convolutional Neural Network (CNN)

CNNs employ filters and max pooling to scan through input images, extracting specific features that help identify hand gestures within frames.

Recurrent Neural Network (RNN)

While CNNs process individual frames, RNNs are used to recognize gestures over time. They retain memory of previous inputs, making them better suited for handling gesture sequences.

Optimizer

The Adam optimizer, a combination of AdaGrad and RMSProp, is known for its adaptive learning capability and is considered one of the most effective optimization algorithms in machine learning.

C. Implementation of the Modules

Input

The CNN processes individual frames, while the RNN analyzes video sequences. Frames are loaded as tensors, enabling the model to be trained accordingly.

Image and Video Processing

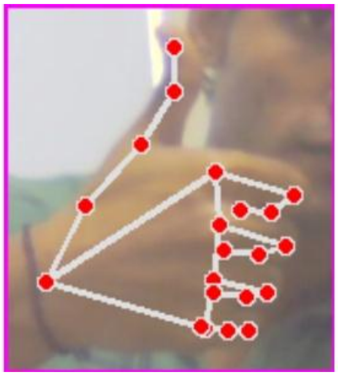
Using OpenCV, frames are extracted from videos converted to grayscale and then scaled and cropped. These are then transformed into tensors for processing.

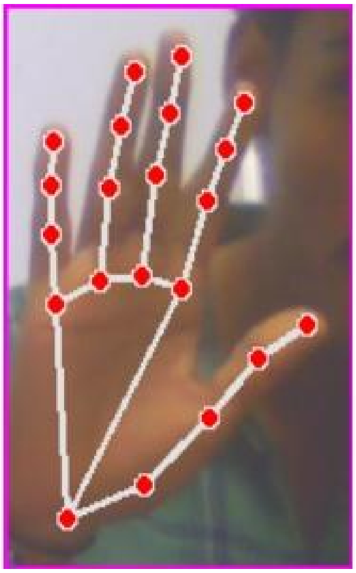
Neural Network

A unified model is developed that allows for the optional inclusion of the RNN while training the CNN on individual frames.

VII. RESULTS

The CNN-RNN model currently achieves an accuracy of 40% with minimal training. This performance can be significantly improved by training the model on a larger dataset and increasing the number of epochs. With more extensive training, the model's ability to recognize gestures and generate more accurate results is expected to improve.





VIII. CONCLUSION

Translating hand gestures into text presents significant challenges due to the diverse range of signs and their varying meanings. Achieving high accuracy in a sign language recognition system is crucial as it can greatly benefit many users. While existing applications provide some functionality there remains an opportunity to create more efficient and advanced systems through the adoption of newer technologies. This project explores one such innovative approach. The utilization of Python and its built-in libraries has streamlined the implementation process while the incorporation of a CUDA-enabled GPU has accelerated the training of neural networks enhancing overall performance and efficiency.

IX. FUTURE WORK

Since such applications will always be necessary, research and development of new algorithms to improve the efficiency of interpretation of sign language will be endless. Deep learning has a lot more to offer than what has already been explored. With such possibilities there will always be scope to develop newer models that can be used in sign language recognition. The models used in this can also be used in other applications such as gesture-controlled systems, tracking unusual activity, etc. with slight tweaks. With slight modification, it can also be used for recognition of other sign languages.

Moreover, the current work can be further extended to output speech instead of text. A similar generative model can be used to map speech or text to actions by stitching together pre-recorded videos or actually generate an interface with a skeletal structure. This model can be seamlessly integrated with various applications on different platforms to support high productivity.

REFERENCES

[1] Human hand gesture recognition using a convolution neural network, Hsien-I Lin, Ming-Hsiang Hsu, and Wei-Kai Chen, 10.1109/CoASE.2014.6899454, August 2014

[2] "Sign Language to Text and Speech Conversion," International Journal of Advance Research, Ideas and Innovations in Technology, www.IJARIIT.com, Bikash K. Yadav, Dheeraj Jadhav, Hasan Bohra, Rahul Jain.

[3] Real-time American sign language recognition using convolutional neural networks was published in 2016 by Garcia, B., and Viesca, S. A. in Convolutional Neural Networks for Visual Recognition, 2, 225–232.

[4] Time Series Neural Networks for Real-Time Sign Language Translation, 17th IEEE International Conference on Machine Learning and Applications (ICMLA), Orlando, FL, 2018, pp. 243-248, Doi: 10.1109/ICMLA.2018.00043, S. S. Kumar, T. Wangyal, V. Saboo, and R. Srinath.

[5] "Arabic Sign Language Recognition using Lightweight CNN-based Architecture," by AlKhuraym, Batool Yahya, et al. International Journal of Advanced Applications of Computer Science (2022)
[6] "Hidden Markov Model for Gesture Recognition," Yang, Jie, and Y. Xu (1994), doi: 10.21236/ada282845
[7] "A translator for American sign language to text and speech," V. N. T. Truong, C. Yang, and Q. Tran, 2016 IEEE 5th Global Conference on Consumer Electronics, 2016, pp. 1-2, doi: 10.1109/GCCE.2016.7800427
[8] 3D MobileNet-v2 and Knowledge Distillation for Sign Language Recognition, X. Han, F. Lu, and G. Tian, ICETIS 2022: 7th International Conference on Electronic Technology and Information Science, 2022, pp. 1-6.
[9] "American sign language recognition and training method with recurrent neural network," Lee, C. K. M. et al. Syst. Appl. Expert. 167 (2021)
[10] (2020) "Text Classification Using Long Short-Term Memory with GloVe Features," in Sari, Winda, Rini, Dian Palupi, and Malik, Reza, Journal of Computer Science and Information Technology, 5. 85, 10.26555/jiteki. v5i2.15021
[11] "Arabic Sign Language Recognition through Deep Neural Networks Fine-Tuning," by Saleh, Yaser, and Ghassan F. Issa. 2020; Int. J. Online Biomed. Eng. 16: 71–83.
[12] S. Liu and W. Deng," Very deep convolutional neural network-based image classification using small training sample size," 2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR), 2015, pp. 730-734, -doi: 10.1109/ACPR.2015.7486599
[13] Sunitha K. A, Anitha Saraswathi.P, Aarthi.M, Jayapriya. K, Lingam Sunny, Deaf Mute Communication Interpreter- A Review, International Journal of Applied Engineering Research, Volume 11, pp 290-296, 2016.
[14] Mathavan Suresh Anand, Nagarajan Mohan Kumar, Angappan Kumaresan, An Efficient Framework for Indian Sign Language Recognition Using Wavelet Transform Circuits and Systems, Volume 7, pp 1874- 1883, 2016.
[15] Mandeep Kaur Ahuja, Amardeep Singh, Hand Gesture Recognition Using PCA, International Journal of Computer Science Engineering and Technology (IJCSET), Volume 5, Issue 7, pp. 267-27, July 2015.
[16] Nakul Nagpal,Dr. Arun Mitra.,Dr. Pankaj Agrawal, Design Issue and Proposed Implementation of Communication Aid for Deaf & Dumb People, International Journal on Recent and Innovation Trends in Computing and Communication, Volume: 3 Issue: 5,pp- 147 149.
[17] S. Shirbhate1, Mr Vedant D. Shinde2, Ms Sanam A. Metkari3, Ms Pooja U. Borkar4, Ms. Mayuri A. Khandge/Sign-Language- Recognition-System. 2020 IRJET Vol3 March 2020.
[18] Nandy, A.; Prasad, J.; Mondal, S.; Chakraborty, P.; Nandi, G. Recognition of Isolated Indian Sign Language Gesture in Real Time. Commun. Comput. Inf. Sci. 2010, 70, 102–107.
[19] Mekala, P.; Gao, Y.; Fan, J.; Davari, A. Real-time sign language recognition based on neural network architecture. In Proceedings of the IEEE 43rd Southeastern Symposium on System Theory, Auburn, AL, USA, 14–16 March 2011.
[20] Chen, J.K. Sign Language Recognition with Unsupervised Feature Learning; CS229 Project Final Report; Stanford University: Stanford, CA, USA, 2011.
[21] Sharma, M.; Pal, R.; Sahoo, A. Indian sign language recognition using neural networks and KNN classifiers. J. Eng. Appl. Sci. 2014, 9, 1255–1259.
[22] Agarwal, S.R.; Agrawal, S.B.; Latif, A.M. Article: Sentence Formation in NLP Engine on the Basis of Indian Sign Language using Hand

DOI: 10.48175/IJARSCT-19946