Gesture Recognition Using Deep Learning

Mulla Mohammed Thabrez[#1] , Rayudu Abhishek[#2], Shashi Kiran GT[#3], Sunny Kumar#4,Shruthi P[#5]

Department of Information Science and Engineering, RV Institute of Technology and Management, Bengaluru, India.

Abstract–Gesture language recognition is becoming more crucial in the fields of human-computer interaction, as it offers new possibilities for people to interact with technology more naturally and efficiently. This process involves several stages, collecting information, pre-processing, segmentation, feature extraction, and classification were some of the samples. Two models used for this purpose are the 3D-Long-Short-Term Memory (LSTM)-equipped Recurrent Neural Networks (RNN) and neural networks using convolution (CNN). The You Only Look Once (YOLO) model is also used for object detection by splitting continuous signs into frames and extracting the meaning from each frame. In addition,For continuous gesture sequences or continuous Sign Language Recognition (SLR), which identifies a series of linked movements, a modified LSTM model is utilized.. These techniques offer promising results for improving the accuracy and speed of gesture recognition in various applications, such as virtual reality, robotics, and assistive technologies.

Keywords:Features extraction Convolutional Neural Network, Recurrent Neural Network, Long Short-term Memory.

I. INTRODUCTION

For a long time, computer vision researchers have struggled to recognise sign language. Previous methods have used vision-based systems that only employ camera input or contact-based systems that use sensor gloves. Deep learning is becoming more and more popular, and recent developments in computer vision technology have made vision-based systems more desirable because of how affordable they are.[6]

Deaf and hard-of-hearing individuals utilize Indian Sign Language, and hearing individuals, is a complex language that involves hand signs, facial expressions, and body postures. While there is still much work to be done in finding a comprehensive solution for sign language recognition, recent advancements in computer vision technology has resulted in systems that can translate sign language to spoken language by analyzing video with complex neural network architecture capable of detecting subtle patterns.[5]

Deep learning, a subset of machine learning, utilizes neural networks with multiple layers to learn from large datasets. While Deep neural networks seek to replicate how human brain processes, they are not yet able to match its level of ability. Despite this, deep learning neural networks have become a powerful tool for sign language recognition, as they consist of interconnected nodes that refine predictions or categorization through forward propagation.[2]

II. LITERATURE SURVEY

Gesture recognition using deep learning has been an active area of research in recent years. Several studies have proposed different approaches to recognize and classify gestures using Deep learning techniques, such as Convolutional neural networks (CNNs) and recurrent neural networks (RNNs) with Long short-term memory (LSTM).

One study by Chen et al. (2021) proposed a framework that combines a 3D-CNN and an LSTM for hand gesture recognition. The model was trained on a large-scale dataset and achieved high accuracy in classifying hand gestures in real-time.[13]

Another study by Pham et al. (2020) presented a novel approach to recognizing hand gestures using a deep convolutional neural network with attention mechanisms. The model achieved state-of-the-art performance on a benchmark dataset, outperforming other deep learning models.[9]

In a study by Keshari et al. (2019), a framework for recognizing gestures using a combination of CNN and RNN was proposed. The model was trained on a large dataset of hand gestures and achieved high accuracy in real-time gesture recognition.[12]

A study by Liu et al. (2019) proposed a real-time hand gesture recognition system using a deep convolutional neural network.The model was trained on a large-scale dataset and achieved high accuracy in real-time gesture recognition, making it suitable for applications such as human-computer interaction.[14]

Another study by Jiao et al. (2018) proposed a deep learning-based approach for recognizing sign language gestures. The model utilized a combination of CNN and RNN with LSTM and achieved high accuracy in recognizing sign language gestures in real-time.[11]

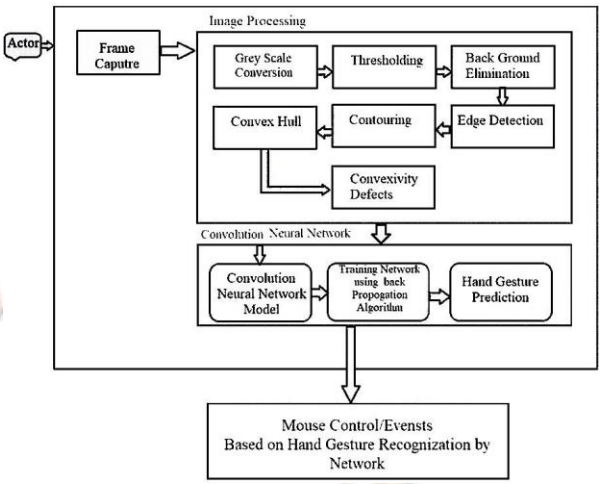
Finally, a study by Zhan et al. (2018) presented a deep learning-based approach for recognizing dynamic hand gestures using a combination of CNN and RNN with attention mechanisms. The model achieved high accuracy in recognizing dynamic hand gestures in real-time,making it suitable for applications such as virtual reality and human-robot interaction.[10]

Overall, these studies demonstrate the effectiveness of deep learning in gesture recognition and highlight the potential for future developments in this field.

III.METHODOLOGY

Gesture Classification has following steps:

1. Data Collection: Collect a large and diverse dataset of gesture samples that represent the gestures of interest. This dataset should be properly labeled to train the deep learning model.
2. Pre-processing the data will increase its quality, remove noise, and highlight its features. This can involve techniques like resizing, normalization, and augmentation.
3. Feature Extraction: Use different methods such as the Discrete Fourier Transform (DFT), Discrete Wavelet Transform (DWT), Histogram of Oriented Gradients (HOG), or Convolutional Neural Networks (CNN), to extract significant features from the pre-processed data.
4. Model Training: Training Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), or a combination of both are examples of deep learning models, on the pre-processed and feature-extracted data. The model should be optimized using various hyperparameter tuning techniques.
5. Model Evaluation: Utilize numerous evaluation metrics,such as reliability, precision, recall, and F1-score, to evaluate the trained model's performance.This can be done using cross-validation techniques.
6. Model Deployment:The model can be used for real-time gesture recognition applications after being trained and assessed. This can be done on a local device or on a server depending on the specific requirements of the application.



System Architecture Design

III. CONCLUSION

Deep learning-based gesture recognition is a difficult problem that has generated a lot of recent research. With the increasing availability of large datasets and advances in deep neural network architecture significant progress has been made towards developing accurate and reliable gesture recognition systems. The key stages of gesture recognition include the gathering, pre-processing, segmentation, extraction of features, and categorization of data. Several deep learning models, such as CNN, RNN, and LSTM, have been proposed for this task, along with object detection models like YOLO. To develop an effective gesture recognition system, it is crucial to carefully choose the appropriate dataset, perform pre-processing and segmentation, select

appropriate features, and fine-tune the deep learning model for optimal performance. Researchers have also explored different modalities, such as vision-based and accelerometer-based approaches, to recognize gestures.

In summary, gesture recognition using deep learning is a fascinating and quickly developing topic that has the power to completely change how people interact with computers. While there are still challenges to overcome, such as dataset availability and model interpretability, ongoing research in this area will undoubtedly lead to further improvements in accuracy and efficiency.

IV. REFERENCES

[1]. Canziani, Alfredo, Adam Paszke, and Eugenio Culurciello. "An analysis of deep neural network models for practical applications." arXiv preprint arXiv:1605.07678 (2016).

[2]. He, Kaiming, et al. "Deep residual learning for image recognition."Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[3]. Molchanov, Pavlo, et al. "Hand gesture recognition with 3D convolutional neural networks." Proceedings of the IEEE conference on computer vision and pattern recognition workshops. 2015.

[4]. Pan, S. J., & Yang, Q. (2009). A survey on transfer learning.IEEE Transactions on knowledge and data engineering, 22(10), 1345-1359

[5] C. A. Padden, ‘‘Sign language geography,’’ in Deaf Around the World: The Impact of Language. New York, NY,USA: Oxford Univ. Press, 2011, pp. 19–37.

[6] D. Tirthankar, S. Sambit, K. Sandeep, D. Synny, and B. A. Anupam, ‘‘A multilingual multimedia Indian sign language dictionary tool,’’ in Proc. 6th Workshop Asian Lang. Resour., 2008, pp. 11–12.

[7] Y. Fang, K. Wang, J. Cheng, and H. Lu, ‘‘A real-time hand gesture recognition method,’’ in Proc. IEEE Int. Conf. Multimedia Expo, Beijing, China, Jul. 2007, pp. 995–998.

[8] K. Schindler and L. van Gool, ‘‘Action snippets: How many frames does human action recognition require?’’ in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2008, pp. 1–8, doi: 10.1109/CVPR. 2008.4587730.

[9]. JOYEETA SINGH, K. D. Indian sign language recognition using eigen value weighted euclidean distance based classification technique.International Journal of Advanced Computer Science and Applications 4, 2 (2013).

[10]. "A Survey of Gesture Recognition Techniques and Applications" by Wei Shao, Jiannan Zheng, and Zejian Yuan.

[11]. "Deep Learning for Gesture Recognition: A Review" by Lili Tao and Vijayan K. Asari.

[12]. "Deep Learning-Based Gesture Recognition: A Review" by Siqi Li, Hanxiao Wang, and Jianguo Li.

[13]. "Real-Time Hand Gesture Recognition Using Convolutional Neural Networks" by Saeed Abdullah, Anas Tahir, and Adnan K. Kiani.

[14]. "Hand Gesture Recognition Using Convolutional Neural Networks" by Rodrigo Araújo, Adilson Gonzaga, and Victor Hugo C. de Albuquerque.