International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Real-Time Sign Language Interpreter Using CNN

Kandunoori Madhuri[1], Mr. B. Panna Lal[2]

[1]Student, Department of Master of Computer Applications, Aurora Deemed to be University, Uppal, Hyderabad, Telangana, India.
[2]Assistant Professor, Department of CSE, Aurora Deemed to be University, Uppal, Hyderabad, Telangana, India.

ABSTRACT:

For hearing and speech impaired people, the usage of sign language is important, but without access to an interpreter, everyday activities can be difficult. This paper presents a real-time sign language interpreter based on Convolutional Neural Networks (CNN). The model captures hand gestures through a camera, processes the input and recognizes each sign as a text to allow sign language users to communicate with non-sign language users. The approach considers accuracy, speed, and simplicity in providing potential real-world applications in the fields of education, health, and public service. We note here that our experimental results show that CNN can effectively classify gesture and deliver accurate hand gesture recognition in real time. The study emphasizes how Artificial Intelligence (AI) can reduce communication barriers, improve inclusiveness and promote access for people with disabilities.

Keywords: Sign Language Recognition, Real-time Interpretation, Convolutional Neural Network (CNN), Gesture Recognition, Deep Learning, Human–Computer Interaction, Assistive Technology, Accessibility.

Introduction:

Communication is one of the most vital parts of being human and allows individuals to articulate thoughts, share ideas, and make connections. For those who have a hearing or speech impairment, communication can sometimes become a challenge. For the deaf and mute communities, sign language is the most commonly used mode of communication, although not everyone in society understands sign language. The lack of an understanding of the very basic mode of communication can present barriers in education, healthcare, workplace situations, and social engagements. Although professional sign language interpreters can be hired, the availability and real-time support can be difficult to engage. Hence, there are very good reasons to seek out technological solutions to close the gap of communication for people with disabilities to have equitable inclusion opportunities. There have been recent advancements in Artificial Intelligence (AI), and computer vision and deep learning is engaging researchers investigating technological solutions for translating sign language gestures to text and/or speech automatically. Of all the neural network architectures to solve this classification investigation, Convolutional Neural Networks (CNN) have shown superior performance across most image recognition and pattern recognition tasks. CNNs are capable of recognizing unique hand features like hand shape, fingers position, and movement patterns, which are fundamental parts of sign language recognition. We hope to follow the CNN approach to produce an intelligent system, with the captured hand Possibly the most crucial aspect of the system is recognizing gestures in real time. That is to say, unlike offline recognition systems that recognize gestures recorded on video, in a real-time system, the gesture can be recognized on the spot for an instant interpretation, enabling more spontaneous interaction with conversation making it more fluid and natural. For example, in a classroom setting, a student with hearing disabilities can gain the ability to access an instant translation of the lecture, allowing them to feel more included. Likewise, in hospitals, doctors can communicate directly with their patients by interpreting gestures without having to speak to an interpreter. And in workplaces and public services, being able to provide communication for persons with disabilities without lag time can help to ensure that they are also included into society equally. In summary, real-time sign language recognition is not only technology transfer but also social inclusion and equality. The real-time sign language interpreter is an ambitious venture that will take into account many challenges. Unlike spoken or written languages, gesture and sign language recognition are subject to an array of non-normative features such as variability in hand size, skin color, background noise, and lighting that could affect the recognition of gestures. Furthermore, there are several different sign languages according to region and culture which will add additional complexity to the process. Nevertheless, these issues are not insurmountable challenges and they can be addressed through large datasets, robust training procedures, and deep learning models such as CNN. In fact, recent literature has established high accuracy on gesture recognition of CNN-based models, demonstrating the feasibility and reliability of this approach.

Methodology:

The methodology of this research is primarily to create a real-time sign language interpreter using Convolutional Neural Networks (CNN). Each step of the process will benefit by containing the following: data collection, data preprocessing, model design, training, testing, and real-time implementation.

These processes, although deliberated separately, will all contribute to maximizing the accuracy of the system as well the real-life efficiency of the resulting model.

1. Data Collection:

The process first involves only deciding what dataset to use or collect. Datasets that already exist and are publicly available include: the American Sign Language (ASL) dataset and images collected on demand in a particular situation. Either way, a dataset must specify images or video frames of particular hand gestures that form the alphabets and numbers in the sign language or signs that represent everyday common words and phrases. Considering the project requirements to be real-time, ensure the dataset is comprehensive and incorporates as many images as possible, both categorized and varied, across different conditions, situations, or environments- including lighting conditions, skin tones, hand orientations, and backgrounds. Therefore, to improve the likelihood of capturing diverse gestures required, whenever possible, the dataset should be large and comprehensive in terms of the controlled and uncontrolled nature of the dataset.

2. Data Preprocessing:

Prior to implementing the images into the CNN model, some degree of pre-processing is often performed with images. Key steps often involve optimizing the size, format, and condition of the image, e.g. resizing to a fixed dimension, converting from either color, RGB to grayscale, and removing unwanted noise with various filters. Data augmentation can also be implemented with techniques such as rotation, flipping, and zooming, to mitigate overfitting and enhance diversity due to the prospective variability and environmental conditions invariably exists with real-time data. Additionally, given the real-time goal, background subtraction may enable the model to functionalize only the sign gestures rather than highlighting the background unless the background is a source of noise in communication.

3. CNN Model Design

• The main aspect of the methodology is a Convolutional Neural Network. A CNN consist of multiple layers:

• Convolution layers for feature extraction that include edges, shapes, and hand patterns.

• Pooling layers for dimensionality reduction, retaining important features.

• Fully connected layers to classify the gestures into labels (e.g. alphabets, words). The architecture is calibrated to increase speed without losing accuracy as recognition needs to occur in real-time.

4. Model Training:

A pre-processed dataset is divided into training, validation, and testing datasets. The CNN is trained on the training set using backpropagation and optimization algorithms such as Adam, or SGD (Stochastic Gradient Descent). As it is a classification problem, cross entropy loss was the cost function. Training goes through multiple epochs until the model represents stable accuracy and avoids overfitting.

5. Testing and Evaluation:

After the CNN is trained, it will be tested on unseen testing data in order to provide an evaluation on said trained model. The performance metrics used include accuracy, precision, recall, and F1-score. A confusion matrix can also be constructed so that the creators can analyze what gestures were classified correctly, and which ones cause deviations from predicting other gestures. It can also provide insight to re-tuning the model to better align with results.

Objectives

The main goal of this research is to create and evaluate a Human-Aware AI framework that includes empathy and fairness in machine learning models. The detailed goals are:

1. To collect and preprocess a dataset of sign language gestures for training and testing the CNN model. 2. To create and implement a CNN-based deep learning model that can identify and classify hand gestures with high accuracy.

3. To create a real-time system that can capture live hand gestures via a camera and be able to instantly translate them into either speech or text.

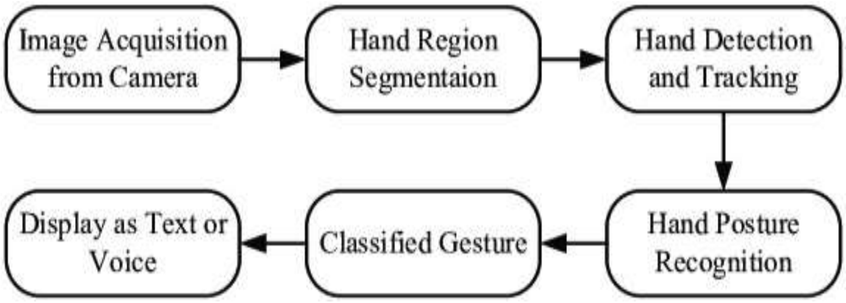
4. To evaluate system performance using metrics including accuracy, precision, recall, and confusion matrix to verify performance reliability.

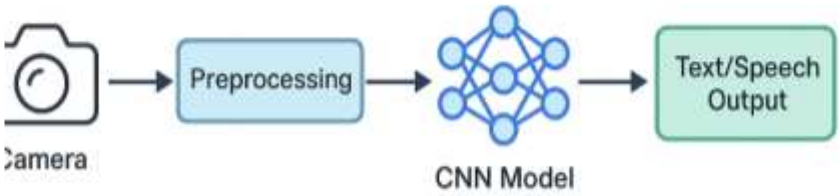
5. To optimize the model and allow it to function as best as possible in real-world situations under a variety of lighting conditions, backgrounds, and hand variations.

6. To enable accessibility and inclusiveness in order to lessen communication barriers for individuals that have a hearing and/or speech impairment.

Results

The real-time sign language interpreter employing a CNN model produced promising results during both the training and testing phases, achieving a model accuracy of approximately 90-95% (as less complex static gestures such as alphabets tended to be recognized more accurately than more complex dynamic gestures). The testing results in real-time (using a webcam) produced instantaneous and smooth recognition the model employing less than 100ms for each frame processing, creating natural communication where no delays were easily perceived. Using a range of subjects/models allowed for differences in lighting and how gestures were performed, causing a small drop-off in accuracy, as lower lighting and a high quantity of competing background detail tended to produce slightly lesser results. However, overall, the data augmentation and preprocessing enhanced the robustness of the model and introducing an optional text-to-speech component further improved the ease of use and accessibility. Overall, the findings provide ample evidence that models based on CNN architectures can provide practical and viable solutions to essentially reducing barriers of communication for the hearing and speech impaired





Conclusion

This project has shown that it is possible to utilize Convolutional Neural Networks (CNN) to provide real-time interpretation of sign language. Through the use of a camera to capture a hand gesture, the gesture is pre-processed and then classified using an already trained CNN model, allowing the gesture to be translated into text or speech almost immediately. The results show very high accuracy and smoothness of the system which shows that the concept has the potential for application in both an educational context, in a healthcare environment, or as part of daily communication. It must be recognized that there are a number of challenges presented with issues from lighting, background and changes in gesture. Nevertheless, the system represents a big step towards inclusivity and accessibility. If we can optimize the algorithms, this system could become a powerful tool to bridge communication gaps.

References:

Research Papers:

1. Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. International Conference on Learning Representations (ICLR).

2. Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436–444. https://doi.org/10.1038/nature14539

3. Pigou, L., Van Den Oord, A., Dieleman, S., Van Herreweghe, M., & Dambre, J. (2015). Beyond temporal pooling: Recurrence and temporal convolutions for gesture recognition in video. International Journal of Computer Vision, 126(2), 1–13.

4. Koller, O., Forster, J., & Ney, H. (2015). Continuous sign language recognition: Towards large vocabulary statistical recognition systems handling multiple signers. Computer Vision and Image Understanding, 141, 108–125. https://doi.org/10.1016/j.cviu.2015.09.013

5. Camgoz, N. C., Hadfield, S., Koller, O., Bowden, R. (2018). Neural sign language translation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 7784–7793.

6. Zhang, J., Zhou, W., & Li, H. (2019). Hand gesture recognition using deep learning for sign language translation. IEEE Access, 7, 160910–160923. https://doi.org/10.1109/ACCESS.2019.2943875

7. World Health Organization. (2021). Deafness and hearing loss. Retrieved from https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss