Title: Sign language interpretation using deep learning LSTM model

Introduction:

We are living in a world where global economy and technology advancements are taking place in a rapid pace. Each day a new invention occurs making how we lived yesterday a story of the past. These advancements can be seen all around us from the AI we have on our devices to the self-driving cars we use to the IoT that assists in our daily routine. We no longer need to interact with other fellow beings to get our work done. We can order food using an AI, convey our grievances with a push of a button, and even live in a virtual space with the introduction of metaverse.

However, not everyone finds these advancements easy. Certain communities with special needs might even find these advancements quite tedious and daunting as it is not friendly for them. For this research project our focus is on people communicating through sign language. Be it ordering food, chatting online, addressing grievances, or any task that can be solved with an AI is proving to be a hardship for people who communicate through sign language. For a civilization to advance forward all the members of the community should have access to the advancements taking place. The people affluent with sign language are lower than those who use sign language as a means of communication. The advanced systems we see today whether it is for entertainment, transport, information delivery, and so on also do not take into account sign language as a mode of communication.

Artificial Intelligence based sign language models do exist around the world but most of the interpreters are focused too specifically for a single purpose application. Along with that the resources, time, and cost needed for the existing technology assisting those people with special needs tend to be exorbitantly expensive for organizations or any entity to afford. It becomes more challenging when the system will have to be integrated with regular functioning systems as the task is expensive and increases complexity of operation. Although organizations and businesses try their best to keep advancing while maintaining an all-inclusive environment the aforementioned factors restrict the efficacy of being absolutely inclusive.

Problem Statement:

Build an efficient sign language interpreter using deep learning LSTM model which requires less cost, time, and technology while maintaining high accuracy.

Literature Review:

Salian et al (2016), developed an image based sign language translator which can be used through mobile devices. The sign's image is converted into its appropriate word using image processing and machine learning techniques. Image processing is used to extract the image's features, and machine learning techniques are used to classify the feature vector. Other than a smartphone, no extra tools are required. However this method has a lot of drawbacks. This method requires high level of power, won't perform well in low light conditions, and always need to be placed in a distance for the device to function.

Another major creation for assisting those communicating through sign language was the glove based interpreter. The various sign language movements are translated into their appropriate text using smart gloves made of several sensors and a low-power printed circuit board (Elmahgiubi et. al., 2015). Each letter in the American Sign Language has a unique combination that is communicated through Bluetooth to a smartphone or computer screen. The circuit board is a bespoke design that can only accommodate the 26 alphabets. However, these devices required separate gadget to be purchased, large computation power and time, and it requires a third party device to be carried around always. Moreover the gloves have to be re programmed for updates and it depends on chip based functioning which is undesirable for daily communications.

Extensive research has been done trying to integrate gloves with Neural network. One such research was conducted by Mehdi and Khan (2002). The training procedure for a neural network uses sensor gloves with 5 sensors on each finger to detect the tension between the knuckle and the first joint of the finger and 2 sensors to assess the tilt and rotation of the hand. The data is mapped to the appropriate character using the neural network that was developed. The 24 alphabets used in the English language as well as two punctuation marks are recognized by the system. The technology translates hand movements into their matching alphabet or digit in real-time. This however still seemed impractical as it required a third party device. This requires high cost and computation power.

Johnny and Nirmala (2022), developed a new sign language interpreter model. The suggested system intends to convert hand signals into words utilising the Australian Sign Language signs (High Quality) Data Set and data collected by Fifth Dimension Technologies (5DT) gloves. The performance of several machine learning methods used to categorise the data into words, including neural networks, decision trees, and k-nearest neighbours, has been compared. It is discovered that the kNN clustering method provides greater accuracy than the other two machine learning algorithms discussed above. However, the technology fails to work when integrated in visual sensor cases. The dependence of this model was highly on the physical gloves.

Madahana et. al. (2022), conducted a review of all the AI technology existing for the those communicating through sign languages. In order to offer an AI-based real-time translation solution for South African languages from speech-to-text to sign language, the research did a scoping review on the application of artificial intelligence (AI) for real-time speech-to-text to sign language translation. In order to find peer-reviewed papers supporting AI-based real-time speech-to-text to sign language translation as a treatment for the hearing impaired, electronic bibliographic databases including ScienceDirect, PubMed, Scopus, MEDLINE, and ProQuest were searched. Prior to the suggested real-time South African translator, this review was conducted. The analysis found a paucity of research on the use of AI and machine learning (ML) as potential remedies for hearing impairment. The clinical use and research of these technical advancements clearly lag. Most methods depend on either 3rd party devices or Machine learning models that function only when a large number of parameter criteria are met. This is a huge drawback.

Dataset:

The original MNIST image dataset of handwritten digits serves as a prominent benchmark for image-based machine learning techniques, but academics are working hard to update it and create drop-in replacements that are both more difficult for computer vision and unique for practical applications. This page presents the Sign Language MNIST, which uses the same CSV format with single rows for labels and pixel values. With 24 classes of letters, the American Sign Language letter database of hand gestures is a multi-class problem (excluding J and Z which require motion). Training video dataset of the two letters will be utilized specifically to train the letters J and Z.

The dataset format roughly resembles the traditional MNIST. There are no cases for 9=J or 25=Z due to gesture motions; each training and test case represents a label (0–25) as a one-to-one map for each alphabetic letter A–Z. The training data (27,455 cases) and test data (7172 cases) are roughly half the size of the standard MNIST but otherwise similar, both having a header row of labels that read "pixel1, pixel2,..." through "pixel784," each of which represents a single 28x28 pixel image with grayscale values ranging from 0-255. Multiple users repeating the move across various backgrounds were represented by the original hand gesture image data. The MNIST data for sign language was obtained by significantly increasing the limited number (1704) of colour photos that were included without being clipped around the hand region of interest. An picture pipeline based on ImageMagick was utilised to produce fresh data, which included cropping to hands-only, gray-scaling, resizing, and finally producing at least 50+ variants to increase the amount. The filters "Mitchell," "Robidoux," "Catrom," "Spline," and "Hermite" were used as part of the modification and expansion plan, coupled with random pixelation at 5%, +/- 15% brightness/contrast, and eventually 3 degrees of rotation. The small size of the photos allows these adjustments to effectively change the resolution and class distinction in fun, manageable ways.

Along with the aforementioned dataset, a trial will be made to create and train the model with live data collection.

Methodology:

For the live data collection the face, hand, and pose landmarks will be determined for key point extraction. This will be stored in separate data collection folders. Then collection of key point sequences will take place along with preprocessing of Data and label generation.

An LSTM deep learning model will be created. Here the model will be trained with the MNIST American sign language dataset. The model will also be trained with the aforementioned dataset. Priority will be given to the aforementioned dataset collected. The model will be used to make predictions post which the weights of the model will be saved. Finally an extensive evaluation will be done which will include a confusion matrix and comparative evaluation between other existing models as well.

A unique kind of recurrent neural network called the LSTM is able to recognise long-term relationships in data. This is made possible by the model's recurring module, which consists of four levels that interact with one another. The model created will be tested in practical use case scenarios and a participant basis feedback evaluation is also hoped to be achieved.

Contributions:

The contributions from my part will be the creation and collection of the aforementioned dataset. The collection, filtering, and identification of the optimal data from the MNIST dataset. The creation, training, and testing of the LSTM model. The practical application testing of the LSTM model which includes testing with participants and evaluating the model. The comparative evaluation and the thorough performance evaluation of the model including the confusion matrix generation will be done.

References

Elmahgiubi, M., Ennajar, M., Drawil, N. and Elbuni, M.S., 2015, June. Sign language translator and gesture recognition. In *2015 Global Summit on Computer & Information Technology (GSCIT)* (pp. 1-6). IEEE.

Johnny, S. and Nirmala, S.J., 2022. Sign Language Translator Using Machine Learning. *SN Computer Science*, *3*(1), pp.1-6.

Madahana, M., Khoza-Shangase, K., Moroe, N., Mayombo, D., Nyandoro, O. and Ekoru, J., 2022. A proposed artificial intelligence-based real-time speech-to-text to sign language translator for South African official languages for the COVID-19 era and beyond: In pursuit of solutions for the hearing impaired. *South African Journal of Communication Disorders*, *69*(2), p.915.

Mehdi, S.A. and Khan, Y.N., 2002, November. Sign language recognition using sensor gloves. In *Proceedings of the 9th International Conference on Neural Information Processing, 2002. ICONIP'02.* (Vol. 5, pp. 2204-2206). IEEE.

Salian, S., Dokare, I., Serai, D., Suresh, A. and Ganorkar, P., 2017, March. Proposed system for sign language recognition. In *2017 International Conference on Computation of Power, Energy Information and Commuincation (ICCPEIC)* (pp. 058-062). IEEE.