ISSN: 2690-0688

Research Article

Journal of Humanities & Social Sciences

A Machine Learning Real-Time Solution to Unify 300 Disparate Sign Languages and Create a Universal Sign Language Translator

Mannat Vikramaditya Jain*

*Corresponding Author
Mannat Vikramaditya Jain

Submitted: 2023, Dec 22 ; **Accepted**: 2024, Jan 26: **Published**: 2024, Jul 19

Citation: Jain, M. V. (2024). A Machine Learning Real-Time Solution to Unify 300 Disparate Sign Languages and Create a Universal Sign Language Translator. *J Huma Soci Scie, 7*(7), 01-04.

Abstract

Hearing loss is one of modern society's most understated and overlooked clinical conditions, with several hundred million people (per WHO) around the world requiring rehabilitation to address "disabling hearing loss." Another large segment of sign language user is 'hearing nonverbal children.' They are nonverbal due to conditions such as down syndrome, autism, cerebral palsy, trauma, and brain disorders or speech disorders.

1. Introduction

Existing Solutions Are Not Scalable, Nor Cost-Efficient: Sign languages (e.g., the American Sign Language) are spoken by less than 2% of the hearing disabled. Also, its comprehension amongst the broader society, i.e., within the 'hearing population' is low. Furthermore:

i. No common standards: There are more than 300 sign languages around the world, each with its own grammar and vocabulary. It's not possible to translate one sign language to another easily. (E.g., ASL to Chinese Sign Lang.)

ii. Prosthetic devices (i.e., hearing aids) are expensive, especially as regards low-income countries, and they do not provide a solution for disabling hearing loss.

iii. Consequences: Social isolation and depression amongst those with disabling hearing loss. Absence of effective communication has real-life consequences, e.g., during interactions with first responders.

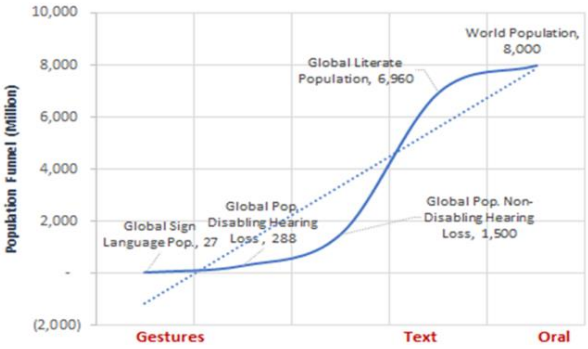

Figure 1: Population Funnel vs Method of Communication.

Source: Graph created using data sourced from the National Institute of Deafness and Other Communication Disorders (www.nidcd.nih.gov); US Census Bureau (www.census.gov) and www.nad.org. The Proposed Solution: A machine learning platform that unifies complex, non-standardized gestures of 300 discrete sign languages into any spoken language on the planet, thus enlarging two-way communication exponentially.

The process involves training a machine learning model on a dataset of new symbols that can unify the 300 global sign languages. These are used to train a machine, which then recognizes these hand and body pose estimation and gestures, converts them to text, and then to the spoken word in any language. Then, an animation engine reverses the process, and converts spoken words to text or gestures to allow for two-way communication.

2. Engineering Goals

First, to create a Universal Sign Language that simplifies complicated phrases into quick shortcuts which can be easily understood across geographies and cultures. The second was to translate this Universal Sign Language into text, and then into the spoken word in real time in any desired language.

Multi-Phase Standardized System

In Phase 1: To train the computer to recognize a standard set of sign gestures and then convert them to closed captioning text in any written language.

In Phase 2, apply a text-to-voice synthesizer and convert such text to any spoken language, thus amplifying the communication funnel more than 25x.

Self-Learning Machine Learning Solution

For 'hearing nonverbal children,' i.e., children who are nonverbal due to conditions such as down syndrome, autism, cerebral palsy, trauma, and brain disorders or speech disorders, the computer will be trained to interpret a wider inventory of non-standard gestures (i.e., standardized gestures with a significantly wider tolerance of variation), and arrive at the same result, i.e., gestures to text to the spoken word.

Use open-source and open access libraries in python to create a neural network that allows a camera to recognize any sign language in the world, and to translate that into any language of the user's choice.

An animation engine can be integrated into the program to allow for two-way communication between speakers of two different sign languages:

3. Methods

Using machine learning, it is possible to create a system that can recognize a database of signs (organized around deictic, motor, symbolic, iconic, and metaphoric), and interpret the signs in real-time. The system generates spoken or written text that represents the meaning of the signs being used, allowing people who use sign language to communicate with those who do not.

There are an optimal number of training epochs and runs-per-gesture for my translator system. (An epoch refers to a complete iteration through a dataset during training of a machine learning model.) The program functioned best at the dual equilibriums of 40 runs/gesture, and 2000 epochs/cycle, showing that this is the optimal for such projects. Adding more data only results in longer runtimes, while making no statistically significant enhancement in the levels of accuracy.

Figure 2: Independent vs. Dependent Variables

Logic and Principles of Organization: The Long-short Term Memory. An LSTM model is optimal as it is a type of a RNN (Recurrent Neural Network) because of the comparatively little training data they need. LSTMs solve this problem by introducing a memory cell, which can store information over a prolonged period. The cell has gates that control the flow of information into and out of the cell, allowing the network to selectively retain or forget information as needed. This is important for recognizing a vast library of gestures whose frequency of repetition is low.

Principles of Spatial Organization: The principal effort is to train the computer to recognize spatial coordinates (x, y, z) of sign gestures that, while intending to be standardized, almost always involve recognition variations arising on account of users possessing a different way of "writing".

4. Data Analysis

To reduce these variations, the program is trained to recognize each gesture of this new, unifying language with increasing numbers of turns until it reaches an acceptable level of accuracy.

Figure 3: Relationship between Training Runs and Accuracy

Using 40 runs/case (A "Run" is defined as the single execution of the training process) arrived at a 97% degree of accuracy. These are sample images of sequential runs. At 40 runs/gesture, the desired level of accuracy was observed. Further runs only provided marginal gains (see graph above) once the graph began to plateau.

The program can be trained with significantly higher runs/gesture to achieve higher levels of gains. An increase in the number of such gestures is a measure of the "vocabulary" of the program, and a measure of its 'intelligence,' and hence, the consequent precision and utility.

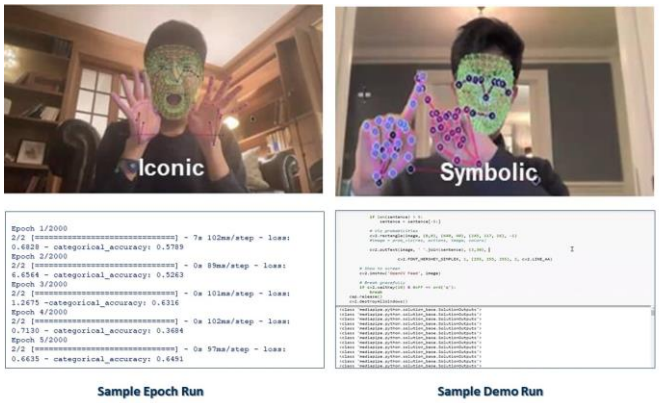

Figure 4: Epoch training runs for a new vocabulary organized around deictic, motor, symbolic, iconic, and metaphoric symbols. Sample symbols which derive meaning from universal iconography create the new dictionary.

This process is done iteratively, with the computer being trained on epochs of data. The quality and quantity of the data is crucial for the performance of the model, and it's also important to have a diverse and representative dataset. Both of these conditions will be met in the manner in which it is proposed to train the computer with 300 sign languages, each with their 100 initial gestures.

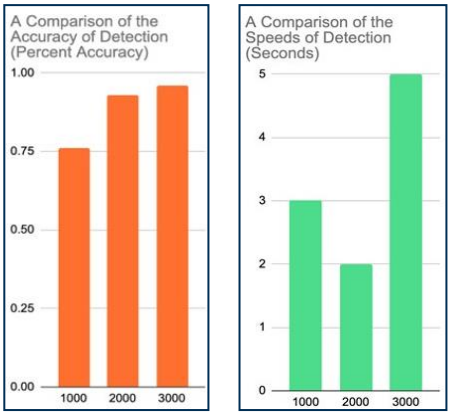

Figure 5: Graphical representation of accuracies of various models run with different numbers of epoch. The graph in green (right) compares the latencies of the different models.

5. Results

This paper shows that not only is it possible to build a fast, free-to-use and functional sign-language translator to the spoken word, it is also possible to run it locally, on the user's device without costs to quality.

The applications of such translative technology are near-limitless, being able to run conversions between multiple different sign-languages at will, adding closed captions to ASL videos to allow immediate understanding by global viewers, and more.

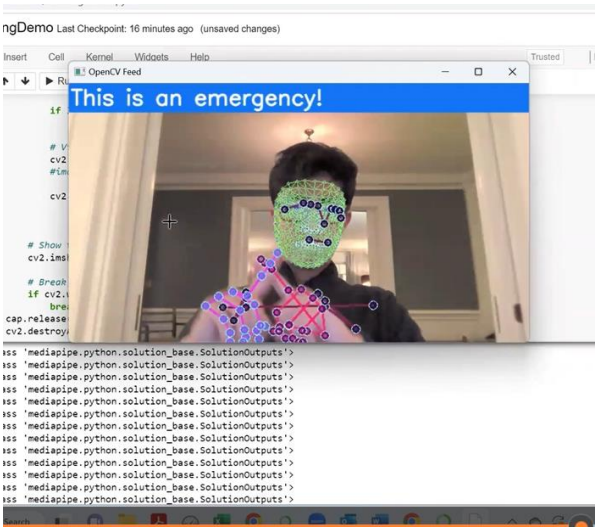

Figure 6: The results: Live gestural recognition of a database of signs.

The Result: Figure 6 is a demonstration of how a new shortcut can be used to convey a powerful message. By using a new gesture, and one which is not rooted in any existing sign language, the shortcut can be adopted by anyone around the world who has a hearing disability. I coded one small gesture to quickly say: "This is an emergency."

The result of the gesture recognition is displayed in the blue ribbon. (I'm showing closed captioning in English. But it could be in any language.) A voice synthesizer has also been added which can convert this text into any spoken language from the thousands spoken around the world. All of this is possible on a smartphone app.

Applications can range from the most basic (e.g., ordering a pizza) to something of critical importance, such as describing an emergency. (See Figure 5.) The main technology can support the creation of various affiliated technology-enabled platforms, such as:

1. In Educational settings: This technology can be used to make it easier for the 'hearing disabled' and the 'hearing non-verbal' children to be educated.
2. In Medical settings: This is probably one of the most important applications. Once a more normal conversation is possible between the hearing disabled and the doctor, a more accurate and nuanced diagnosis will be possible.
3. In the Workplace: This can exponentially expand the number of jobs and professions in which the hearing disabled can participate. This will have a direct consequence on their quality of life.

Scaling the system enables improved accuracy, faster times, and the ability to run on reduced hardware - something that could benefit the many who lack access to advanced smartphone systems. Additional improvements can also include adding in live avatars to translate from text-to-gestures [1-5].

Participants

The human subject used for the experimentation is the author of this paper himself.

Acknowledgements

I wish to thank Dr. Steven Gordon of Garden City High School in NY who guided my work and provided invaluable feedback throughout the process.

References

1. Mitchell, R. E., Young, T. A., Bachelda, B., & Karchmer, M. A. (2006). How many people use ASL in the United States? Why estimates need updating. *Sign Language Studies, 6*(3), 306-335.
2. Zheng, J., Zhao, Z., Chen, M., Chen, J., Wu, C., Chen, Y., ... & Tong, Y. (2020). An improved sign language translation model with explainable adaptations for processing long sign sentences. *Computational Intelligence and Neuroscience, 2020.*
3. Chai, X., Li, G., Lin, Y., Xu, Z., Tang, Y., Chen, X., Zhou, M. (2013). Sign Language Recognition and Translation with Kinect. *Language Recognition and Translation with Kinect.*
4. De Castro, G. Z., Guerra, R. R., & Guimarães, F. G. (2023). Automatic translation of sign language with multi-stream 3D CNN and generation of artificial depth maps. *Expert Systems with Applications, 215,* 119394.
5. "Deafness and Hearing Loss." World Health Organization, World Health Organization, 1 Apr. 2021, https://www.who.int/news-room/fact-sheets.