International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 3 March 2023, pp: 1005-1010 www.ijaem.net ISSN: 2395-5252

Sign Language Translation Using Deep Learning

Abhijeet Chandak, Mihir Nagalkar Kunal Deore, Ranjeet Hinge, Tanuja Patankar, Laxmi Kale
Department of Information Technology, Pimpri Chinchwad College of Engineering, Pune

Date of Submission: 09-03-2023 Date of Acceptance: 18-03-2023

ABSTRACT
Sign language translation is a key solution for bridging the communication barrier between those with hearing abilities and those without. By allowing for equal participation in both social and professional settings, it eliminates the communication barriers faced by individuals who experience hearing loss or difficulties. Advanced technologies such as video recognition, machine learning, and computer vision have made sign language translation more accurate and efficient. However, there are still some challenges to be addressed, including recognition of regional variations in sign language and improvement of real-time translation speed. Despite these challenges, the research and advancement of sign language interpretation systems remains vital in enhancing the daily lives of individuals with hearing difficulty. The importance of sign language translation cannot be overstated, as it provides equal access to information and opportunities for these individuals.
Keywords: CNN, RNN, LSTM, Sign Language

I. INTRODUCTION

People who are deaf-dumb commonly employ sign language as a means of interaction. A sign language is nothing more than a collection of varied hand gestures created by varying hand shapes, movements, and orientations, as well as face expressions. 34 million of the 466 million persons with hearing loss in the globe are children. People who are considered "deaf" have very limited or no hearing. They communicate using sign language.

Around the world, many sign languages are used by people. They are extremely few in number when compared to spoken languages. The sign language used in India is known as “Indian Sign Language (ISL)”. There are hardly many schools in poor nations that serve to deaf children. Adults with hearing loss have an extremely high unemployment rate in developing nations. According to data from Ethnologue, the literacy rate and number of children attending school are quite low among India's deaf community, which makes up roughly 1% of the country's total population. It continues by saying that offering transcription in sign languages, expanding the number of interpreters available, and officially recognizing sign languages all considerably enhance accessibility.

This study demonstrates researching unique, time-saving, user-friendly online learning platforms for deaf persons who can utilize them to communicate and learn

II. BACKGROUND

2.1 CNN

A convolutional neural network (CNN) is a type of deep learning algorithm designed for image and video recognition tasks. It is based on the idea of convolution, which is a mathematical operation used to extract features from images. Convolutional, pooling, and fully linked layers are only a few of the many layers that make up the CNN architecture. For example, edges, corners, and patterns are recognized in a picture by the convolutional layers. These features are then passed through the pooling layers, which reduce the dimensionality of the data and make it more manageable for the next layer.

In image and video identification tasks including object detection, facial recognition, and picture segmentation, CNNs have been shown to be quite successful. They are extensively employed in manydifferent industries, including robotics, computer vision, and medical imaging.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 3 March 2023, pp: 1005-1010 www.ijaem.net ISSN: 2395-5252

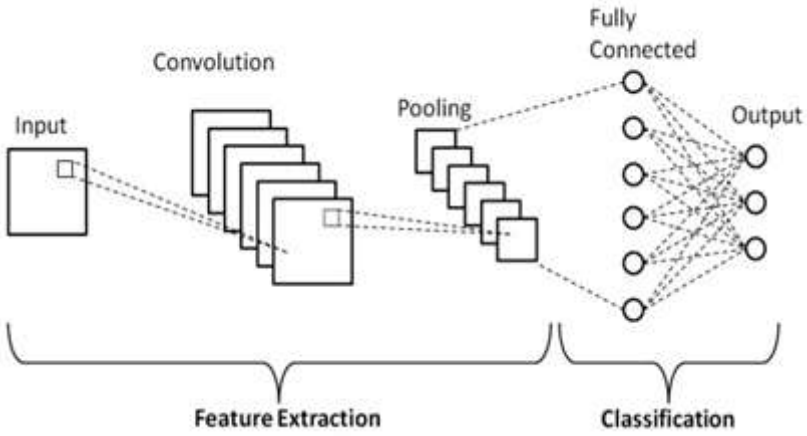

Fig.1. Block Diagram of CNN

CNN has several layers:

1. **Input Layer:** The input layer receives the image or video data and passes it to the next layer.
2. **Convolutional Layer**: A series of filters are applied to the input picture by the convolutional layer in order to recognize characteristics like edges, corners, and patterns. The subsequent layer receives these traits after that.
3. **Pooling Layer:** The pooling layer creates the feature maps for the convolutional layer, reducing their dimensionality and making them more manageable for the subsequent layer. Typically, a max-pooling technique is used to accomplish this, which chooses the highest value from a limited area of the feature map.
4. **Fully Connected Layer:** In the fully connected layer, the network generates a probability that the input image belongs to a certain class for classification. Every neuron in the layer is connected to every other neuron in the layer underneath it. This layer is often made up of many neurons.
5. **Output Layer**: The fully connected layer sends the final class probabilities to the output layer, which then outputs the CNN's final prediction.

This is a basic architecture of CNN, and it may vary depending on the specific task and the dataset used. Some CNN architectures may include multiple convolutional and pooling layers or additional layers such as dropout layers or batch normalization layers for regularization.

2.2 RNN

A sort of neural network called a recurrent neural network (RNN) is made to handle sequential input, such as time series data or natural language text. RNNs are able to maintain a "memory" of past inputs, allowing them to better understand and make predictions about the current input based on the context of past inputs. This is accomplished through the use of feedback connections, which allow information to flow through the network across multiple time steps. RNNs are commonly used in a wide range of applications, Speech recognition, time series forecasting, and natural language processing.

RNNs have a unique structure compared to traditional feedforward neural networks, in that they have a "recurrent" connection that allows information to flow from one step of the network to the next. This recurrent connection means that the output of the network at a given time step is not only dependent on the current input, but also on the previous outputs. This allows RNNs to maintain a kind of internal memory, allowing them to process sequences of inputs in a way that is not possible with traditional feedforward networks.

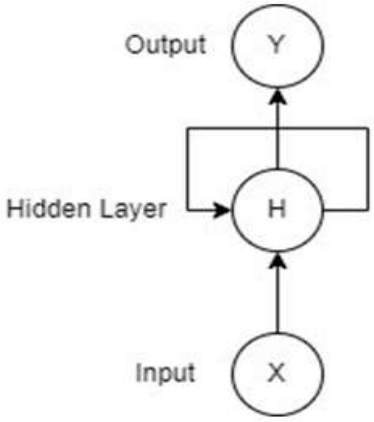

Fig.2. Architecture Diagram of CNN

There are several different RNN types,

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 3 March 2023, pp: 1005-1010 www.ijaem.net ISSN: 2395-5252

including simple RNNs, LSTMs, and gated recurrent unit (GRU) networks. Simple RNNs are the most basic form of RNNs and have a simple recurrent connection, which makes them prone to the vanishing gradient problem. LSTM and GRU networks address this problem by introducing gating mechanisms that allow them to selectively remember or forget information from previous time steps, making them better suited for processing long-term dependencies.

In summary, RNN is a kind of neural network made to handle sequential data., it utilizes the recurrent connection that allows information to flow from one step of the network to the next, giving the network a form of memory and the ability to process sequences of inputs that is not possible with traditional feedforward networks.

2.3 LSTM

Recurrent neural networks (RNNs) with Long Short-Term Memory (LSTM) are created expressly to handle theproblem of long-term dependencies in sequential input. Compared to conventional RNNs, LSTM networks mayretain data from earlier time steps for a significantly longer amount of time. Gating techniques, which enable thenetwork toselectively retainor forgetinformation, are used todothis.

The I/P gate, forget gate, and O/P gate are the three separate gates that each memory cell in the LSTM network design possesses. Information flow into the memory cell is governed by the I/P gate, information flow out of the memory cell is governed by the forget gate, and information flow from the memory cell to the network's output is governed by the O/P gate.

Information that should be erased from the memory cell by the forget gate, transmitted on to the next time step by the O/P gate, and information that should be stored there by the I/P gate are all determined by these gates. To operate these gates, a variety of weights that are learned throughout training are employed

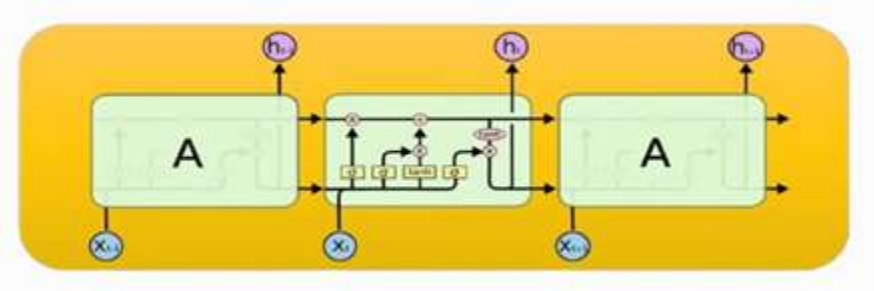

Fig.3. Block Diagram of LSTM

The LSTM network architecture, which consists of a series of memory cells with three different gates an I/P gate, a forget gate, and an output gate enables the network to selectively remember or forget information. In conclusion, LSTM is a type of RNN that is intended to address the problem of long-term dependencies in sequential data. A set of weights that are learnt during training are used to operate these gates.

2.4 SVM

The supervised machine learning method known as Support Vector Machines(SVMs) is utilised for classification and regression problems. Finding the best border (or"hyperplane")between several classes in the data is the basicgoal of SVMs. The margin the separation between the border and the nearest data points for each class ismaximized bythis boundary choice.

SVMs are particularly useful when the data has many features (also known as high-dimensional data) or when the data is not linearly separable. In these circumstances, SVMs may transfer the data into a higher-dimensional space where a linear boundary can be located using a method known as the "kernel trick".

Both linear and non-linear classification issues may be solved with SVMs. Because multiple Kernel functions may be selected for the decision function, they are efficient in high-dimensional spaces, memory-efficient, and flexible.

SVMs are often utilized in fields including bioinformatics, natural language processing, and image classification.

III. LITERATURE SURVEY

Some authors have described various methods for building assign language translation their approaches are described below:

In a study titled "Research of a Sign Language Translation System Based on Deep Learning," Siming He proposed a deep learning method for the identification of common sign language and hand location. The study utilized a

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 3 March 2023, pp: 1005-1010 www.ijaem.net ISSN: 2395-5252

region-based Convolutional Neural Network (R-CNN) to recognize the hand portion of images or sign language videos. The results showed a remarkable accuracy rate of 99%.[1].

In this work "Automated Speech to Sign language Conversion using Google API and NLP" by "Ritika Bharti, Sarthak Yadav, Sourav Gupta, and Rajitha Bakthula" described a method using Google API and NLP. The system first uses the Google API to transform voice into text. After tokenizing the text, videos of the words that are available are then retrieved, concatenated, and shown as a single combined video. In offline mode, the model's accuracy is 74%, whereas in online mode, it is 90%[2].

In this paper "Deep learning in vision-based static hand gesture detection", Oyebade K. Oyedotun and Adnan Khashman proposed a deep learning-based method for hand gesture recognition. They used the gesture recognition dataset created by Thomas Moeslund. When learning the complicated hand gesture categorization issue using a CNN and stacked denoising autoencoder, error rates are reduced. The accuracy for a small dataset of 24 photos reached up to 95%[3].

Sarah Ebling et. al Association for Computational Linguistics in 2019, has designed to make the components technologies which are easily accessible to all the experts who are sign language experts but not computer science experts. The author has use SiGML used SiGML method which then converted into the signed animation using JASigning avatar system[4]

B. Gallo et. al IEEE in 2019, the author has presented many experiments that are used in the development of a statistical system of the translation system for the deaf people to translate the speech into sign language. The author has used the Automatic Speech Recognition (ASR) System and statistical machine translation to obtain the required results. The error rate in the result was found out to be 28.21% and the results were found by the author with the help of finite state transducer[5].

Table.1. Comparison of Research Papers

Parameter	Paper 1	Paper 2	Paper 3
Authors	Siming He	Ritika Bharti, Sarthak Yadav, Sourav Gupta, and Rajitha Bakthula	Oyebade Oyedotun, Adnan Khashman
Publication	2019 International Conference on Artificial Intelligence and Advanced Manufacturing (AIAM)	ICAEEC-2019, IIIT Allahabad India, 31st May - 1st June,2019	Springer London, journal ISSN: 0941-0643
Algorithm used	CNN, RNN, LSTM	Google API and NLP	Autoencoders
Use of Algo	R-CNN: to recognize the sign language video CNN & LSTM: to build the recognition algorithm	Google API: speech to text conversion NLP: text tokenization	When learning the complicated hand gesture categorization issue using a convolutional neural network and stacked denoising autoencoder, error rates are reduced

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 3 March 2023, pp: 1005-1010 www.ijaem.net ISSN: 2395-5252

Advantages	High Accuracy as compared to other technologies	Easy to implement and process	CNN & Autoencoder is used
Disadvantages	The scope of the data collected is constrained and not all sign language terms are included.	Low accuracy and low performance in offline mode.	Only suitable for small subsets of the ASL hand gestures.
Accuracy	99%	Offline: 74%	95%

IV. EXPERIMENTAL RESULTS

The research paper "Research of a Sign Language Translation System Based on Deep Learning" by Siming (2019), used the Convolutional Neural Network (CNN) algorithm to develop a sign language translation system. The system was trained using the Asl Alphabet Train dataset, which contained 87,000 images of alphabets and other signs. The effectiveness of the system was evaluated using the Asl Alphabet Test dataset, which consisted of 28 images. By utilizing the CNN algorithm and these datasets, the research aimed to demonstrate the potential for deep learning in sign language translation and improve communication between hearing and non-hearing individuals.

The accuracy graph and modelloss are shown below.

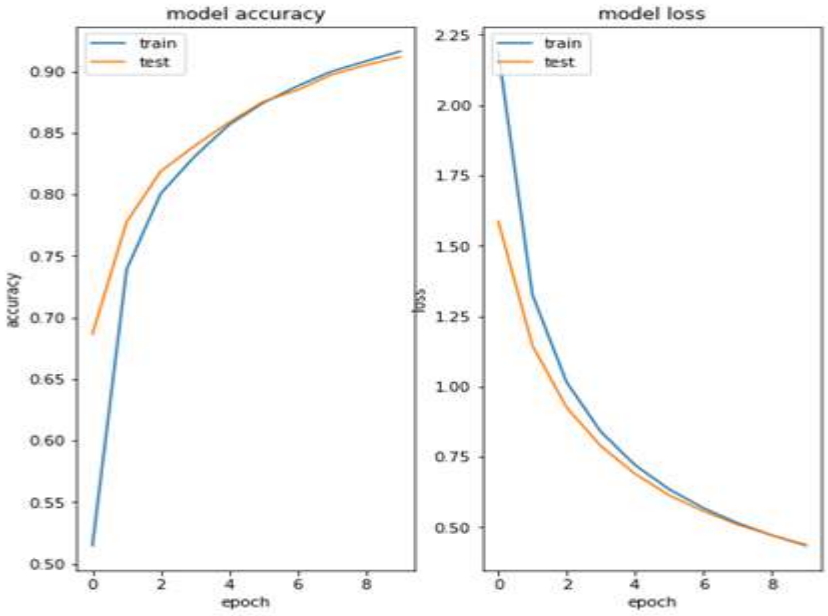

Fig 4: Model Accuracy & Model Loss

From above graph we can see that as number of epochs increases accuracy of model increases and loss decreases. The overall accuracy of model is 95%.

V. FUTURE SCOPE

The future scope for sign recognition systems is vast and has the potential to improve communication andaccessibilityforindividualswhoaredeaforhardofhearinginavarietyofsettings.Somepotentialareasofgrowth include:

1. **Increased integration with smart devices:** Sign recognition systems could be integrated with smart devices such as smartphones, tablets, and smart watches, making it more accessible for individuals to use.

2. **Improved accuracy:** Advancements in machine learning and computer vision technology will lead to more accurate sign recognition systems, reducing errors and increasing the overall effectiveness of the system.
3. **Real-time translation:** Sign recognition systems could be used to provide real-time translation, allowing individualswho are deaf or hard of hearing to communicate with individuals who do not know sign language in real-time.
4. **Assistive technology:** Sign recognition systems could be integrated into assistive technology, such as hearing aids, and cochlear implants, to improve accessibility for individuals who are deaf or hard of hearing.
5. **Increase of Sign language database:** The database of signs of various sign languages will increase which will improve the recognition system, and it will help in creating sign language recognition for new sign languages.

Overall, the future of sign recognition systems is promising and has the potential to greatly improve the lives of individuals who are deaf or hard of hearing

VI. CONCLUSION

In this study, various methods for sign language translation were analyzed, including Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), and Support Vector Machines (SVM). The results showed that among these methods, our research has found that Convolutional Neural Networks (CNNs) provide highly accurate results in sign language translation, with an accuracy rate of 92% when trained on data.

However, when tested in real-world scenarios, the accuracy rate may decrease, highlighting the need for further refinement and improvement of these systems This high accuracy rate demonstrates the potential of CNN in delivering accurate sign language translations.

As a result, future work in this field will likely focus on further improving the application of CNN in sign language translation systems. Additionally, it is important to note that while the results indicate that CNN is the best algorithm for sign language interpretation, it is still important to consider other methods and potentially integrate them for more comprehensive and effective sign language translation systems.

REFERENCES

[1]. Siming, "Research of a Sign Language Translation System Based on Deep Learning", 2019 International Conference on Artificial Intelligence and Advanced Manufacturing (AIAM).

[2]. Ritika Bharti, Sarthak Yadav, Sourav Gupta, and Rajitha Bakthula, "Automated Speech to Sign language Conversion using Google API and NLP", ICAEEC-2019, IIIT Allahabad India, 31st May - 1st June,2019.

[3]. Oyebade K. Oyedotun, Adnan Khashman, "Deep learning in vision-based static hand gesture recognition", Springer London, journal ISSN: 0941-0643.

[4]. B. Gallo, R. San-Segundo, J.M. Lucas, R. Barra, L.F. D'Haro, F. Fernández "Speech into Sign Language Statistical Translation System for Deaf People" in IEEE LATIN AMERICA TRANSACTIONS, VOL. 7, NO. 3, JULY 2019.

[5]. Andrej Karpathy; George Toderici; Sanketh Shetty; Thomas Leung; Rahul Sukthankar; Li Fei-Fei "Large scale video classification with convolutional neural network" DOI: 10.1109/CVPR.2014.223.

[6]. K. Schindler, L. Van Gool, "Action snippets: How many frames does human action recognition require?" in IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2008, IEEE, 2018, pp. 1–8.

[7]. Ilias Papastratis, Kosmas Dimitropoulos, Dimitrios Konstantinidis, Petros Daras," Continuous sign language Recognition Through cross model Alignment of video and text embedding in a joint-latent Space" DOI: 10.1109/ACCESS.2020.2993650.

[8]. "Indian Sign Language Recognition using Convolutional Neural Network" https://doi.org/10.1051/itmconf/20214003004

[9]. "Research of a Sign Language Translation System Based on Deep Learning" DOI: 10.1109/AIAM48774.2019.00083