IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

American Sign Language To Text Conversion Using Cnn Model

[1]Girija V, [2]Akshay Kumar Singh, [3]Nayab Sahil, [4]Shibu Singh, [5]Tulika Paul

[1]Assistant Professor,

Computer Science and Engineering,

Cambridge Institute of Technology,

Bengaluru, India

Abstract— Sign language is an essential communication language that helps individuals with hearing loss interact with others. Convolutional Neural Network is a useful tool for image processing tasks, including sign language recognition. This paper proposes a novel CNN-based approach to sign language to text translation. The CNN model is intended to effectively extract temporal and spatial properties from sign language containing video sequences. Convolutional layers are utilized to extract hierarchical features, while pooling layers are employed to decrease spatial dimensions without sacrificing important information. The model in this work is trained on a large dataset of sign language images, allowing strong representation learning for accurate translation. The CNN model has performed well in translating American Sign Language into text, according to test results. Using datasets of American Sign Language, the model surpasses previous methods and reaches high accuracy. In general, the suggested CNN-based approach for translating sign language to text provides a pathway between people who use sign language and others who are not familiar with it. By providing real-time translation capabilities, this tool can improve accessibility and inclusivity for the community in a range of situations, such as everyday communication, healthcare, and education. This research promotes more equality and integration for those with hearing loss and enhances assistive technologies.

Keywords—Sign language recognition, Image processing, Deaf communication, Gesture-to-text conversion.

I. INTRODUCTION

Deaf people use sign language, a visual-gestural language, to communicate. It is a rich and expressive a language with unique syntax and grammar. Deaf individuals often face communication barriers, especially with those who are not familiar with sign language. Bridging this communication gap is essential for their integration and participation in society. Technology can play a significant role in addressing these challenges. Systems for recognizing and converting sign language motions into text or speech can be developed, making it easier for non-signers to understand. Recognizing the pressing need to surmount these barriers, this paper introduces a pioneering project mainly focused on the building of a real-time Sign language to text translation system. CNNs are a powerful technology for image classification. They enable the automatic extraction of features for classification by applying a number of layers to input images. In the first layer, convolutional layers scan the image with learnable filters to detect local features like edges and textures. Pooling layers then downsample these features, retaining essential information. After being flattened, the feature matrices are fed to fully connected layers for analysis of global relationships and trends. Non-linearity is

introduced by activation functions, and the network's output is converted into probability scores for various classes by a softmax layer. During training, the model's parameters are adjusted to minimize a loss function. Figure 1 shows the CNN model.

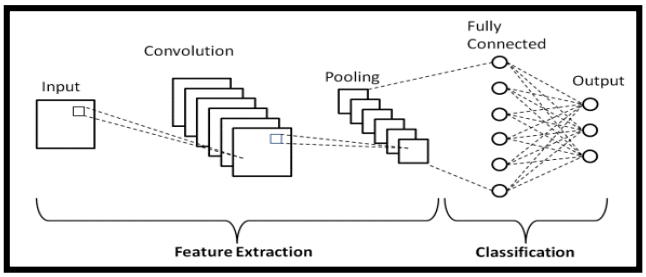

Fig 1: CNN Model

By harnessing CNNs, This approach aims to close the gap between the users of Sign language and non-users, thereby fostering greater inclusivity and engagement within society. Central to this project's scope is the meticulous design of a CNN architecture tailored to process sign language gesture images or video frames. In the subsequent sections, this paper delineates the methodology underpinning the CNN-based sign language translation system, elucidating its architectural intricacies, training methodologies, and performance evaluation metrics. Artificial Neural Network is a connection of neurons, replicating the structure of human brain. Each connection of neuron transfers information to another neuron. Inputs are fed into first layer of neurons which processes it and transfers to another layer of neurons called as hidden layers. After processing of information through multiple layers of hidden layers, information is passed to final output layer.

Additionally, we explore the broader societal implications of this system and its capacity to engender transformative change in communication accessibility for the deaf community. Through this research endeavor, we endeavor to advance the frontiers of inclusive technology and champion the cause of linguistic equity and accessibility for all.

II. RELATED WORK

After a comprehensive analysis of several academic papers and studies concerning finger spelling recognition and sign language, we set out to combine this abundance of knowledge into a useful application.

P. Vijayalakshmi [2] suggested converting sign language to speech. Flex sensors are used in a sensor-dependent gesture interpretation system that has been designed to detect hand movements. By recording the shift in resistor values, the flex sensor in the suggested system is implicitly utilized to calculate the angular tilt to which the finger is bent.

A "Real-time Dynamic Hand Gesture Recognition using Hidden Markov Models" was proposed by M.M. Gharasuie [1]. They have presented a system that uses a Hidden Markov Model to identify dynamic hand movements for English digits 0–9. The isolated and dynamic gesture recognition is done using the HMM, which has average recognition rates of 99.167% and 93.84%, respectively

Dabre Kanchan [4] presented a ML model that uses webcam photos to understand sign language. It focuses on translating sign language, particularly Indian Sign Language, at the word level to text and then voice. In the classification stage, each static gesture is trained using 500 positive samples, 500 negative samples, and 50 test picture samples. The movement is then interpreted using the Haar Cascade classification method. The last phase turns the text into speech. The outcomes show that the accuracy is 92.68 percent accurate.

Bantupalli Kshitij [3] suggested utilizing machine vision and deep learning to recognize American sign language. A micro-proposed technique makes advantage of the author's data-set, which consists of a restricted set of motions with the most often used words. Together with recurrent neural networks (RNNs), convolutional neural networks (CNNs) are used to classify individual tr gestures and image sequences. CNN was constructed using code from the Inception Model. The pool layer's accuracy was approximately 58 LDR percent, while the SoftMax layer's accuracy was nearly 90%.

Kadam Kunal [6] suggested hiring a US sign language interpreter. A glove made of flex sensors is developed. Flex sensors, an LCD, an accelerometer, and a keypad make up the system. In addition to bridging the communication gap, the project aims to create a self-learning system that enables individuals to learn American Sign Language. There are two modes: the teaching mode and the learning mode. When using the teaching mode, a database is built by executing various gestures and stored in the microcontroller's EEPROM.

Aditya Das [5] suggested utilizing deep learning to recognize sign language from static gesture photographs that have been specially processed. Cropping, scaling, and flipping are methods used in data augmentation to make sure the neural network is not restricted to a certain kind of image. The photos are divided into testing, validation, and training sets using a bespoke algorithm that takes as parameters the percentages of testing and validation. The percentages of the validation are more than 90%.

Adithya V[8] Proposed a method for Indian sign language using ANN. Image acquisition, hand segmentation and then classification based on supervised feed forward backpropogation algorithm was used by Adithya, Vinod and Usha Gopalkrishnan for hand feature extraction having average rate of 91.11%.

Nobuhiko MUKAI[7] Proposed a "Japanese Fingerspelling Recognition based on Classification Tree and Machine Learning", NICOGRAPH International, 2017. Japanese sign language incorporates fingerspelling that are rooted in the American alphabet system, with additional elements influenced by Japanese characters, gestures, numbers, and specific meanings. The researchers sought to overcome this particular problem given by the language form's complexity by incorporating cutting-edge machine learning techniques. The core of their approach involved the utilization of a classification tree, a decision tree-based structure commonly employed in pattern recognition tasks. This tree-based model was designed to discern and categorize the intricate patterns inherent in fingerspelling, allowing for the accurate identification of individual signs. The researchers reported a noteworthy achievement of 86% accuracy rate for their proposed recognition method. This success underscored the SVM-based model in deciphering the nuanced and diverse aspects of fingerspelling in JSP. The high accuracy implies the system's capability to reliably recognize and classify a significant portion of the complex linguistic expressions.

Cao Dong[10] Proposed a method for ASL. This study used Microsoft's Kinect sensor to obtain depth data. The Hand segmentation is done using per-pixel classification method. Random Forest (RF) gesture classifier was implemented to recognize ASL signs using the joint angles. The system considered 24 static alphabets and achieved accuracy of 92%.

M. Mohandes[9] Proposed a method using the leap motion controller. To recognize Arabic sign language, they employed a Leap motion controller. Using a jump motion sensor, 10 samples of the 28 letters were gathered for this system. Twelve of the 23 attributes that the LMC returned for every frame of data were deemed most pertinent for additional processing. They used Nave Bayes classifier and Multilayer Perceptron (MLP) to classify 28 letters in Arabic SL. A correct recognition rate of 99.1% achieved using Multilayer Perceptron and 98.3% using Nave Bayes classifier

Cippitelli, Daniele [12] DeepASL: Enabling Pervasive and Non-Intrusive Mobile sign recognition was proposed. In the range of sign language recognition, it is a noteworthy contribution. This study addresses the need for practicality in real-world applications by emphasizing the creation of a deep CNNs model designed for mobile devices.

Ball, Jonathan [11] suggested using CNNs for sign language translation and recognition. It presents a novel method for using CNNs to recognize and translate American Sign Language (ASL). The authors' thorough research, which made use of a sizable dataset of ASL motions, produced encouraging results for both recognition and translation tasks. CNNs are used here to highlight the importance of DL methods for CV computer vision and language processing.

Alex Graves[14] Proposed a Sign Language translation Using a CNNs. It presents a noteworthy contribution to the field of sign language recognition. The authors designed and implemented a CNNs based system specifically designed for recognizing Australian Sign Language (Auslan) gestures. By harnessing the powers of CNNs to process and test the visual cues inherent in sign language, they achieved remarkable accuracy.

Thad Starner[13] Proposed a Sign Language translation with Microsoft Kinect. It was conducted by Starner et al. stands as a pioneering endeavor in the realm of sign language recognition. Employing the Microsoft Kinect sensor, this project predates the

deep learning era, yet its significance is undeniable. It serves as a foundational milestone that paved the way for subsequent research and innovations in sign language recognition.

E. Assogba[16] Proposed a DL for Sign Language translation. They provide a thorough synopsis of the most recent advancements in sign language interpretation. Through the utilization of DL methodologies, such as CNNs, the writers explore the most recent developments in the domain. They offer a comprehensive analysis of several models and datasets. The study puts light on the capabilities and trends in the range of sign language technology, making it a useful resource for scholars, practitioners, and enthusiasts. It emphasizes how deep learning has the ability to change communication for the deaf community by increasing accessibility to sign language.

Juyoung Shin[15] proposed a method using Wearable Myoelectric Sensors. It represents an innovative approach to Sign language translation. Although it places a greater emphasis on utilizing myoelectric sensors rather than conventional CNNs, it makes a noteworthy addition to the field. This work investigates a new approach to recognition of the language using wearable myoelectric sensors.

Hrishikesh Kulkarni[18] Proposed a Deep Learning-Based ASL Recognition System. It introduces an innovative deep learning system for the translation of ASL gestures. The authors leverage combination of CNNs and Long Short-Term Memory networks (LSTMs) to achieve precise and robust recognition results. Importantly, their work extends beyond theoretical application by a practical model built for real-time usage. This development holds good promise for the deaf community, as it paves the way for accessible and proper communication in real-world scenarios, underlining the transformative potential of deep learning in making ASL more accessible and inclusive.

Oscar Koller [17] Proposed a Neural Machine Translation for Sign Language. It presents an in-depth exploration of applying neural machine translation techniques to sign language, providing a thorough synopsis of the area. While its focus is not exclusively on CNNs, it provides valuable insights into the broader landscape of SL translation. By examining various methods and strategies, this work addresses the critical challenge of bridging the gap between spoken and sign languages. It highlights the significance of advanced machine learning techniques in making sign language more accessible, facilitating communication, and promoting inclusivity for the deaf community, within the broader context of neural machine translation.

Siawpeng Er, Jie Zhang,[20] Proposed a SL Recognition and Translation: A Multimodal Deep Learning Approach. It represents an innovative approach to sign language translation. This study combines visual and depth data obtained from RGB-D sensors, thus leveraging a multimodal approach. CNNs are used to process the visual data, allowing for a more comprehensive understanding of SL movements. This research is unique in that it focuses on integrating various modalities and using depth and visual information to increase recognition accuracy.

Chien-Wei Wu[19] Proposed a Sign Language Recognition Using 3D CNNs. They represents a significant advancement in the area of SL recognition. This work adopts a novel approach by incorporating 3D CNNs, which are capable of capturing spatiotemporal data, making them perfect for understanding gestures in sign language. This research introduces a more holistic and nuanced understanding of SL recognition.

III. PURPOSED METHOD

Dataset collection, Data-preprocessing, training/testing, and gesture categorization were the stages.

A. Dataset

B. In this model, data collection is carried out as images depicting various signs at different angles, covering the sign letters A to Z, utilizing the OpenCV library. A total of 180 raw images of the alphabet from A to Z are captured, meeting the requirements for ASL (American Sign Language) representation.

B. Data pre-processing

In this hand detection method, the mediapipe library which is used for image processing is initially considered for the purpose of detecting hands from webcam-captured images. The region of interest from the webcam image is the hand; it is cut from the picture and, after applying Gaussian blur, is converted to a grayscale image using the OpenCV package. The OpenCV library,

commonly referred to as the Open Computer Vision Library, makes it simple to apply the filter. Figure 2 displays the image data collected from the conversion of the grayscale image to a binary image using threshold and adaptive threshold techniques.

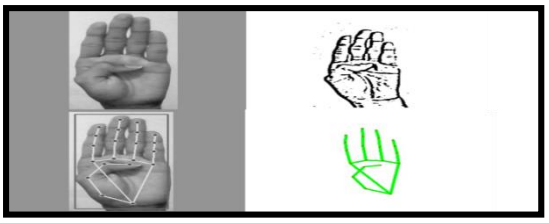

Fig 2: Data preprocessing

C. Classification Algorithm

The method predicts the user's final symbol using two levels of algorithms. Once the features matrices have been extracted from the frame captured with OpenCV, apply the Gaussian blur filter and threshold to obtain the processed image. After theprocessing of this image, it is sent into the CNN model for prediction. If a letter is identified for more than 60 frames, it is printed and used to build the word. Using the blank symbol, spaces in the words are considered. The layers of the models are as follows:

1. 1st Convolution Layer: The resolution of the image is 128×128 pixels. 32 filter weights (3x3 pixels each) are used in the first convolutional layer to analyze it initially. A 126X126 pixel image will be produced as a consequence, one for each filter-weight.
2. 1st Pooling Layer : Using max pooling of 2x2, or keeping the highest value in the 2x2 square of the array, we down sample the images. Consequently, our image has been down-sampled to 63x63 pixels.
3. 2nd Convolution Layer: The second convolutional layer now uses this 63x63 from the first pooling layer's output as an input. 32 filter weights (3x3 pixels each) are taken in the second convolutional layer to process it. This will produce a picture with 61x61 pixels.
4. 2nd Pooling Layer: The final photos are downsampled once more utilizing the maximum pool of 2x2, and their resolution is lowered to 30 x 30.
5. 1st Densely Connected Layer: These images are now fed into a second convolutional layer, which restructures the output into an array of 30x30x32 = 28800 values. The 1st layer is a fully CN layer with 128 neurons. This layer receives an array of 28800 values as input. The second densely connected layer receives the output from these layers. To avoid overfitting, we are utilizing a dropout with a value of 0.5.
6. 2nd Densely Connected Layer: A completely linked layer with 96 neurons now receives output from the 1st densely connected layer as an input.
7. Final layer: The last layer gets its input from the second densely connected layer, and the no. of neurons in that layer corresponds to the no. of classes (alphabets plus a blank sign) that need to be classified.

D. Training and Testing

The procedure entails converting RGB input photos to grayscale and using Gaussian blur to reduce superfluous noise. After that, the hand is separated from the backdrop using adaptive thresholding. Ultimately, 128 x 128 pixel resolution is achieved by resizing the photos. After completing all required data preprocessing procedures, the input images post-preprocessing are given into the model to perform training and testing. A cross-entropy is a performance metric. It produces values that are zero when the prediction matches the label and produces positive values when the prediction deviates from the labeled value. The goal of optimization is to lower the cross-entropy as near to zero as feasible. The neural value of weights are changed in the network layer to accomplish this. A cross-entropy calculation function is incorporated into TensorFlow. Once the cross-entropy function has been determined, it is optimized through the use of Gradient Descent, particularly using the Adam Optimizer

E. User Interface

This module implements the user interface. In this the hand gesture is captured from the webcam and the gesture is classified to the one of the ASL alphabets then the finger spelling is done to form words and then sentences. Fingerspelling is used to form words by spelling out each individual letter of the word using specific hand signs. For each alphabet in the word, use the corresponding hand sign from the fingerspelling alphabet. Sign each letter in sequence, from the 1st to the last as in the figure 3 below.

Fig 3: Finger Spelling to form word

IV. RESULTS AND DISCUSSIONS

A 97% accuracy rate is attained utilizing this method, demonstrating consistent performance across various conditions, including scenarios with clean and without a clean background, under optimal lighting conditions. The developed sign language to text translation system, leveraging CNNs, exhibited good results in real-time interpretation of SL gestures. Overall, the developed sign language to text translation system is an important step in the right direction when it comes to communication barriers faced by the deaf and community. By harnessing the capabilities of deep learning and image recognition techniques, the system offers a promising solution for enhancing communication accessibility and fostering greater inclusivity in society. Continued research and development efforts in this domain are warranted to advance the state-of-the-art in SL recognition technology and promote equitable communication opportunities for individuals of all linguistic backgrounds.

V. CONCLUSION

In summary, the building of a real-time system that converts sign language to text using CNNs is a major step in removing communication barriers that the deaf community must contend with. Through the utilization of DL and image recognition algorithms, the system provides a pathway between sign user and non-user, thus promoting more accessibility and inclusivity.

The successful implementation and evaluation of the CNN-based SL recognition system underscore its potential to revolutionize communication for individuals with hearing impairments. Beyond its immediate applications, the system holds promise for integration into various communication devices and educational tools, thereby empowering people with hearing impairments to engage more fully in educational, social and professional domains.

Moving forward, continued research and development efforts in this domain are warranted to further refine and optimize SL translation systems. By advancing the state-of-the-art in SL recognition technology, we can collectively work towards building a more inclusive and equitable society, where individuals of all linguistic backgrounds have equal opportunities for communication and engagement

REFERENCES

[1] M.M.Gharasuie, H.Seyedarabi, proposed a Real-time Dynamic Hand Gesture Recognition using Hidden Markov Models, 8 th Iranian Conference on Machine Vision and Image Processing(MVIP), IEEE, 2013.

[2] P Vijayalakshmi and MAarthi, proposed a Sign language to speech conversion. In 2016 International Conference on Recent Trends in Information Technology (ICRTIT), IEEE, 2016.

[3] Kshitij Bantupalli and Ying Xie, proposed American sign language recognition using deep learning and computer vision. In 2018 IEEE International Conference on Big Data (BigData). IEEE, 2018.

[4] Kanchan Dabre and Surekha Dholay, proposed a machine learning model for sign language interpretation using webcam images. In 2014 International Conference on Circuits, Systems, Communication, and Information Technology Applications (CSCITA). IEEE, 2014.

[5] Aditya Das, Shantanu Gawde, KhyatiSurat wala, and Dhananjay Kalbande, proposed a Sign language recognition using deep learning on custom processed static gesture images. In 2018 International Conference on Smart City and Emerging Technology (ICSCET). IEEE, 2018.

[6] Kunal Kadam, Rucha Ganu, Ankita Bhosekar, and SD Joshi, proposed a American sign language interpreter. In 2012 IEEE Fourth International Conference on Technology for Education. IEEE, 2012.

[7] Nobuhiko MUKAI, Naoto HARADA, Youngha CHANG, proposed a “Japanese Fingerspelling Recognition based on Classification Tree and Machine Learning”, NICOGRAPH International, 2017.

[8] Adithya V., Vinod P., Usha Gopalakrishnan, proposed a “Artificial Neural Network Based Method for Indian Sign Language Recognition”, IEEE Conference on Information and Communication Technologies (ICT 2013), JeJuIsland April 2013.

[9] M. Mohandes, S. Aliyu and M. Deriche, proposed a "Arabic sign language recognition using the leap motion controller," 2014 IEEE 23rd International Symposium on Industrial Electronics (ISIE), Istanbul, 2014.

[10] Cao Dong, M. C. Leu and Z. Yin, proposed a "American Sign Language alphabet recognition using Microsoft Kinect," 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Boston, MA, 2015.

[11] Jonathan Ball, Brian Price, proposed a Sign Language Recognition and Translation with CNNs, IEEE 2016.

[12] Daniele Cippitelli, Davide Cipolla, proposed a DeepASL: Enabling Ubiquitous and Non Intrusive Mobile Sign Language Recognition, IEEE 2018.

[13] Thad Starner, Mohammed J. Islam, proposed a Sign Language Recognition with Microsoft Kinect, IEEE 2013.

[14] Alex Graves, Santiago Fernández, proposed a Sign Language Recognition Using a Convolutional Neural Network, IEEE 2018.

[15] Juyoung Shin, Joo H. Kim, proposed a Sign Language Translation and Recognition using Wearable Myoelectric Sensors, IEEE (2017).

[16] E. Assogbaand P. H. S. Amoudé, proposed a Deep Learning for Sign Language Recognition and Translation, IEEE 2019.

[17] Oscar Koller, David Ney, proposed a Neural Machine Translation for Sign Language: A Survey, IEEE 2020.

[18] Hrishikesh Kulkarni, Suchismita Saha, proposed a Deep Learning-Based American Sign Language (ASL) Recognition System, IEEE 2020.

[19] Chien-Wei Wu, Eugene Lai, proposed a Sign Language Recognition Using 3D Convolutional Neural Networks, IEEE 2019.

[20] Siawpeng Er, Jie Zhang, proposed a Sign Language Recognition and Translation: A Multimodal Deep Learning Approach, IEEE 2020.