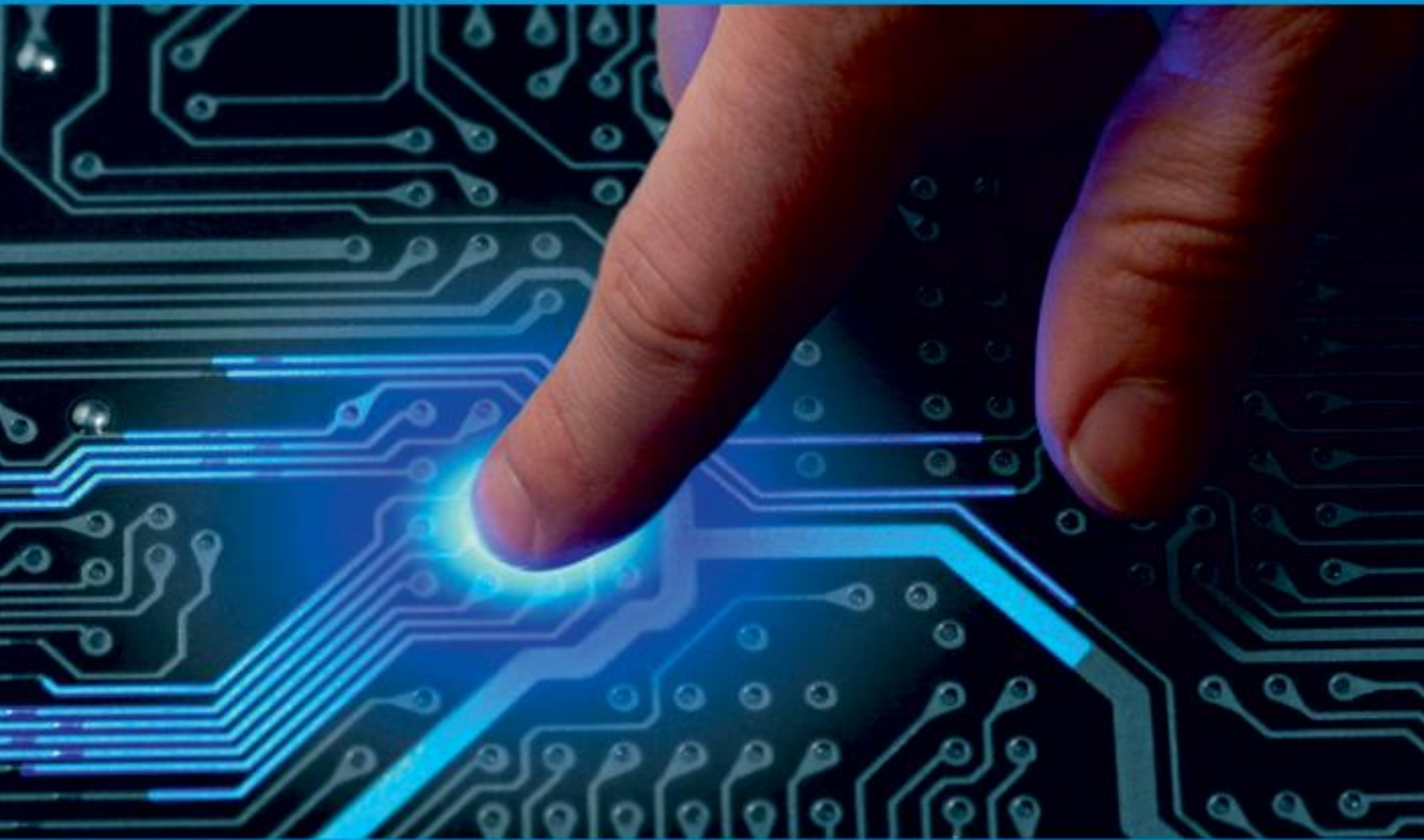
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 11, Issue 5, May 2023

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105074 |

A Research Study on Deep Learning for Sign Language Recognition

Mohit Patil, Geeta Jogi, Vinod S. Wadne

Department of Computer Science, JSPM's Imperial College of Engineering and Research, Wagholi, Pune, India

ABSTRACT - There are many ways to communicate with the deaf, including using sign language. A person's capacity to vocally communicate is hampered by speech impairment, another type of disability. People who have difficulty speaking may therefore turn to alternate communication techniques like sign language. Despite the rising popularity of sign language, it may be difficult for non-signers to connect with signers. The recognition and interpretation of gestures have thankfully made tremendous strides thanks to recent developments in computer vision and deep learning approaches. Because of this, a deep learning-based programme that can convert sign language into text has been created, allowing communication between signers and non-signers. In order to identify indications from video frames, the application uses a customised CNN (Convolutional Neural Network).

KEYWORDS - Deep Learning, Image processing.

I. INTRODUCTION

As it allows users to express their feelings and thoughts through non-verbal gestures, sign language is an essential form of communication for people with speech and hearingdifficulties. Individuals who are not conversant in sign language, however, may find it difficult to understand these motions. In order to overcome the communication gap, trained sign language interpreters are therefore required at all times throughout medical, legal, educational, and training activities. In response to the rising demand for these services, remote video interpretation services have come to be recognised as a practical alternative, offering simple sign language interpretation via high-speed internet connections. These services have substantial shortcomings in terms of dependability, accuracy, and accessibility in spite of how convenient they are.

We created a unique convolutional neural network (CNN) model that can instantly recognise sign language motions in order to overcome these difficulties. The model consists of eleven layers, including dense layers, a single flattening layer, one dropout layer, three max-pooling layers, and four convolution layers. We trained our own CNN model toaccurately detect signs using the American Sign Language Dataset from MNIST, which includes data of various improved gestures. We used OpenCV to create our model, allowing it to instantly identify indications from video frames.

Compared to conventional sign language interpretation techniques, our suggested customized CNN model has significant advantages. It offers a dependable and accurate solution that does away with the need for human translators, ensuring communication access for people with hearing and speech impairments in a variety of contexts. Additionally, our model is adjustable and versatile, able to recognise a variety ofsign language motions and take into account various sign language dialects.

Our unique CNN model, in conclusion, represents a big breakthrough in bridging the communication gap between people with speech and hearing impairments and those without. It offers a trustworthy, accurate, and usable solution for sign language recognition that may be applied in a variety of scenarios, including those involving medicine, law, education, and training. The proposed paradigm has the potential to completely change how we interact with people who have hearing and speech impairments, facilitating their ability to properly express their feelings and thoughts and engage in society..

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105074 |

II. LITERATURE SURVEY

Hand movements and facial expressions are used in sign language, a complex and extensive language, to convey letters, words, and sentences. It is a natural language that those who are deaf or hard of hearing use to communicate. Medical, legal, educational, and training sessions all require the ability to recognise sign language gestures. The use of computer vision as well as deep learning techniques to recognise sign language has been suggested by various scholars in recent years.

One method for recognising Indian sign language (ISL) that was suggested in [1] uses optimised neural networks to identify ISL gestures. Performance of this method was contrasted to that of the Multilayer Perceptron Feed-Forward Network (MLP-FFN) and neural networks (NN) classifiers. The performance metrics that those who conducted the study calculated were precision, recall, precision, accuracy, the F-measure, and the Kappa value statistic. The results of the testing demonstrated that the proposed method performed better at identifying ISL motions than the state-of-the-art methods.

Using dynamic hand motion recognition algorithms in a real-time setting, Indian sign detection was based on previous study [2]. Skin pixels were utilised to segment the movie after the captured footage had been pre-processed by being converted to the HSV colour space. In order to producemore accurate results, depth data was also combined with other data. To identify the gestures, a Support Vector Machine (SVM) was employed after the Hu-Moments and motion trajectory were retrieved from the image frames. A camera and a Microsoft Kinect were both used to test the system. In terms of communication and education, hearing-impaired people may benefit from this kind of equipment.

A system that automatically recognises Indian Sign Language numerals was proposed in another research paper [3]. (0-9). With 1000 shots, 100 of which were of the numbered sign, the implementation database was self-created. The required features were extracted using shape descriptors, the Scale Invariant Feature Transform (SIFT), and the Histogram of Oriented Gradients (HOG) methods. The indicators were categorised using artificial neural networks (ANN) and support vector machine (SVM) classifiers.

The objective of a study on the recognition of static sign language using deep learning was to create a system that could transform static sign language into its word equivalents, such as letters, numerals, or basic static signs [4], which aimed to familiarise people with the fundamentals of sign language. The significance of the system's non-signer functions was established by the researchers through the development of an evaluation technique and some testing. Throughout the evaluation, the solution received high marks for usability and learning impact.

In a different investigation, an experimental comparison of computer vision-based sign language recognition systems was made. [5]. On a number of freely accessible datasets, a thorough review was conducted using the most modern deep neural network approaches in this field. The aim of this project was to investigate sign language recognition by mapping non-segmented video streams to glosses. Two novelsequence training criteria that are well-known from the fieldsof voice and scene text recognition were proposed for this work.

The use of convolutional neural networks (CNN) powered by deep learning to represent reliable static signs in the context of sign language identification was covered in [6]. For this study, a total of 35,000 sign pictures of 100 static signs were collected from various users. The effectiveness of the proposed approach was evaluated on about 50 CNN models. The proposed CNN architecture employed max-pooling, ReLU, and convolutional layers in that order.

Finally, a new multi-modality ArSL dataset that incorporates numerous modalities was proposed in [8]. It extends over 6748 footage clips of four vocalists making use of Kinect V2 sensors to perform fifty signs. As they grow in the discipline, researchers are able to use this dataset to refine and test their approaches to business\. Modern deep learning methods werealso employed to examine the fusion of spatial and temporal properties.

III.METHODOLOGY

Our project intends to record signers' live sign language performances and translate it into text and audio output for illiterate users. A camera-based approach will be employed for this since it promotes higher levels of portability and mobility than other recommendations.

A camera-capable device will first record a video of the signer. Then, our application will process this video. The video would be split up into several frames, turning it into a raw visual sequence. The boundaries will then be

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |
|| Volume 11, Issue 5, May 2023 ||
| DOI: 10.15680/IJIRCCE.2023.1105074 |

initially determined by processing of this image sequence. It will be useful to divide the various body parts the camera is capturing into the two main subparts of the head and hands.

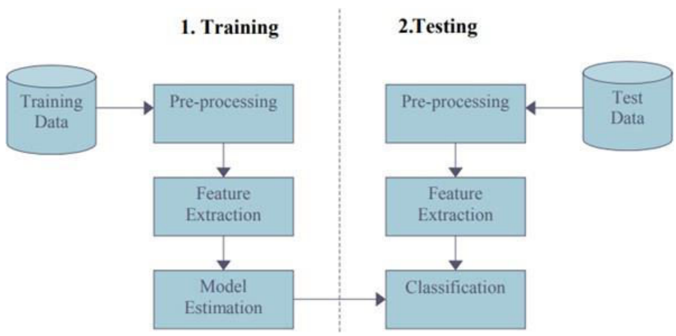


Figure 1: Process Flow Diagram

The head component will be further divided into the categories of position, motion, and facial expression. The movement of the hands will be used to deduce postures and gestures. The MNIST Dataset will then be used to classify all of the data after it has been compared to it.

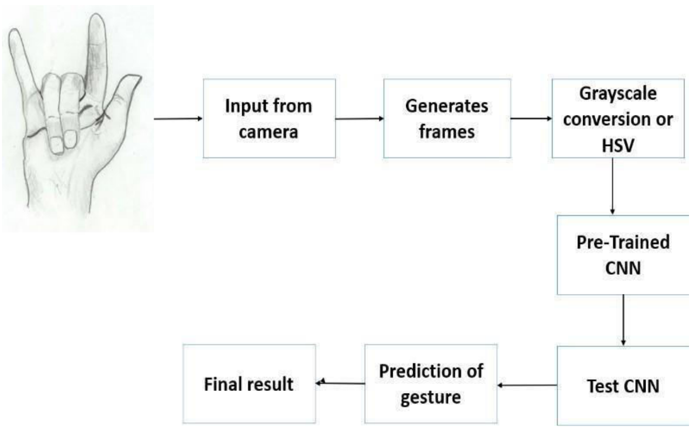


Figure 2: System Architecture for CNN

The image is initially segmented from a video input from the webcam, as seen in the above figure. To prevent backdrop conflicts, the frames are removed from the video using a region of interest. 11 layers of a customised CNN model are utilised. The segmented gesture image from the video frame is then turned into a grayscale image.
Since the model was trained using the features of grayscale images—the MNIST dataset is a pre-processed dataset of RGBimages that are converted to grayscale—the input image from the webcam was converted to grayscale.
The transformed image is then scaled in accordance with the dimensions of the training images. After scaling and transformation, the image is sent into the pre-trained custom CNN model. After obtaining the gesture prediction from the CNN model, it is categorised using the categorical label. The gesture's classification is shown as text. After obtaining the gesture's prediction from the CNN model, it is categorised using the categorical label. The gesture's classification is shown as text.

IV.CONCLUSION

The suggested technique can accurately predict the appearance of several popular words and signs in a variety of lighting and speed situations. By providing a range of values that might dynamically recognise the human hand, the photographs are accurately masking. CNN is used by the proposed system to train and categorise images. More informative elements from the photos are finely extracted and used for categorization and training. The suggested

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105074 |

system can anticipate the activity of signals using a variety of deep learning frameworks. The accuracy of a convolutional neural network system is higher than that of conventional machine learning algorithms.

V.FUTURE SCOPE

In recent years, research has focused more and more on the automatic detection of hand sign language utilising various synthetic and real-time features. Deep learning algorithms have been investigated by researchers as a possible aid for this task, and they have had success in creating models that are highly accurate in recognising signals.

Another area of interest is real-time word recognition of sign language, which has the potential to significantly enhance communication between those who are deaf or hard of hearing and those who are not. Researchers are moving closer to reaching this aim by utilising cutting-edge methods like neural network ensembles and putting the genetic algorithm for sign recognition into practise.

One of the major advantages of deep learning algorithms for sign language recognition is their capacity to learn from vast datasets of labelled instances, which enables them to generalise successfully to new, unseen data. For applications like live interpretation, these models can also be modified to operate in real-time.

In general, the development of automatic sign language recognition systems has the potential to significantly enhance accessibility and communication for people who are deaf or hard of hearing, and the use of deep learning techniques is aiding in this endeavor

.

ACKNOWLEDGEMENT

We would like to thank my teacher, Dr. V.S. Wadne, for his support and direction. I learned a lot more about this Deep Learning for Sign Language Recognition project, which was very beneficial to me. We also want to express our gratitude to everyone who assisted us during the process, whether directly or indirectly.

REFERENCES

[1] Bantupalli, Kshitij, and Ying Xie. "American sign language recognition using deep learning and computer vision." 2020 IEEE International Conference on Big Data (Big Data). IEEE, 2020.

[2] Wadhawan, Ankita, and Parteek Kumar. "Deep learning-based sign language recognition system for static signs." Neural Computing and Applications 32.12 (2020): 7957-7968.

[3] Rao, G. Anantha, et al. "Deep convolutional neural networks for sign language recognition." 2021 Conference on Signal Processing and Communication Engineering Systems (SPACES). IEEE, 2018.

[4] Tolentino, Lean Karlo S., et al. "Static sign language recognition using deep learning." International Journal of Machine Learning and Computing 9.6 (2021): 821-827.

[5] Adaloglou, Nikolaos M., et al. "A Comprehensive Study on Deep Learning-based Methods for Sign Language Recognition." IEEE Transactions on Multimedia (2021).

[6] Wadhawan, Ankita, and Parteek Kumar. "Deep learning-based sign language recognition system for static signs." Neural Computing and Applications 32.12 (2020): 7957-7968.

[7] Bird, Jordan J., Anikó Ekárt, and Diego R. Faria. "British sign language recognition via late fusion of computer vision and leap motion with transfer learning to American sign language." Sensors 20.18 (2020): 5151.

[8] Luqman, Hamzah, and El-Sayed M. El-Alfy. "Towards Hybrid Multimodal Manual and Non-Manual Arabic Sign Language Recognition: mArSL Database and Pilot Study." Electronics 10.14 (2021): 1739.

[9] Pariwat, Thongpan, and Pusadee Seresangtakul. "Multi-Stroke Thai Finger-Spelling Sign Language Recognition System with Deep Learning." Symmetry 13.2 (2021): 262.

[10] Saggio, Giovanni, et al. "Sign language recognition using wearable electronics: implementing k-nearest neighbors with dynamic time warping and convolutional neural network algorithms." Sensors 20.14 (2020): 3879

INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 8.379

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING