IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A Comprehensive Study On CNN-Based Methods For Sign Language Detection

Sakshi Devkar[1] Prachi Dhamane[1] Shejal Kale[1] Vaishnavi Khandagale[1] Prof. Kirti Patil[2]

[1]Department of Computer Engineering, Alard College of Engineering and Management Pune

[2] Assistant Professor, Department of Computer Engineering, Alard College of Engineering and Management Pune

Abstract

Sign Language is mainly used by deaf (hard hearing) and dumb people to exchange information between their own community and with other people. It is a language where people use their hand gestures to communicate as they can't speak or hear. Sign Language Recognition (SLR) deals with recognizing the hand gestures acquisition and continues till text or speech is generated for corresponding hand gestures. Here hand gestures for sign language can be classified as static and dynamic. Deep Learning Computer Vision is used to recognize the hand gestures by building Deep Neural Network architectures (Convolution Neural Network Architectures) where the model will learn to recognize the hand gestures images over an epoch.

KEYWORDS: Sign Language, CNN, Deep Learning, Dumb, Communication.

1. Introduction

Sign language is a visual and spatial language that uses hand movements, gestures, facial expressions, and body language to convey very important meaning. Sign languages are complete and complex languages with their grammar and syntax. The frames are exactly dropped from the video with a wonderful region of Interest to wonderfully avoid background conflicts. A custom CNN model with multiple layers is used. The gesture image, segmented from the video frame, is then converted to a grayscale image. Input images taken through a webcam are then converted into grayscale as per the requirement of the CNN model provides the gesture prediction.

1.1 American Sign Language (ASL):

ASL has its roots in the early 19th century when Thomas Hopkins Gallaudet, an amazing American educator, collaborated with Laurent Clerc, a deaf teacher from France. Clerc brought the fantastic methods of French Sign Language to the United States, and together they founded the very first permanent school for the deaf in Hartford, Connecticut, in 1817.

1.2 Research and Linguistic Studies: Linguists and researchers started studying sign languages as fully developed languages with their syntax, grammar, and linguistic properties. This recognition wonderfully helped dispel misconceptions about sign languages being mere gestural communication systems.

1.3 Technology and Accessibility: Advances in technology, such as video relay services, captioning, and sign language interpretation services, have significantly improved accessibility for the deaf and hard-of-hearing community.

2. Related Work:

Sign Language Recognition Based on CNN and Data Augmentation its author G. Li, X. Wang and Y. Liu [1] designed SLR system based on CNN has been proposed for recognizing American sign language (ASL) letters. To enhance the model's ability to generalize and mitigate overfitting, the application of data augmentation techniques is applied. The performance of the proposed system was evaluated using recognition rate, which revealed an impressive average recognition accuracy of 99.32% for the sign language MNIST dataset.

Action Detection for Sign Language using Machine Learning its author A. S Sushmitha Urs, V. B Raj, P. S, P. Kumar K, M. B R and V. Kumar S [2] . This study focuses on American Sign Language (ASL), implementing convolutional neural networks (CNNs) for real-time sign language recognition using monochromatic camera images. The impact of ASL extends beyond communication, promoting inclusivity, empowerment, and societal effects on education and careers.

Sign Language Digits Recognition Technology Based on a Convolutional Neural Network its author O. Voloshynskyi, V. Vysotska, R. Holoshchuk, S. Goloshchuk, S. Chyrun and D. Zahorodnia [3]. This paper employs a convolutional neural network (CNN) to address digit recognition in sign language images, using a dataset of 2062 64x64 pixel images. Training involves 1649 images, achieving a model with a loss of 0.4381, accuracy of 0.8541, validation loss of 0.2117, and validation accuracy of 0.9540 after 16 iterations. The confusion matrix shows a prediction accuracy of 96.48%.

Real-time Indian Sign Language Recognition using Skeletal Feature Maps its author V. Kumar, R. Sreemathy, M. Turuk, J. Jagdale and A. Agarwal [4]. This study addresses sign language recognition, a well-explored problem in the context of deep learning. While many models claim high validation accuracy but struggle with real-time performance, the proposed model achieves consistent real-time predictions using Skeletal Feature Maps generated from key points via Media Pipe. Employing a novel 15-layer CNN architecture, the model attains a remarkable validation accuracy of 99.84% on a self created 106-word Indian Sign Language dataset, surpassing other machine learning models and pretrained CNN architectures.

Development of Sign Language Translator for Disable People in Two-ways Communication its author P. Singh, S. V. S. Prasad, R. Singh, K. Dasari and B. L. Prasanna [5]. This paper presents a prototype aimed at aiding the communication of deaf and dumb individuals with the broader community. The developed system translates gestures into text and speech, focusing on converting hand signs into communicable English words or phrases. This two way communication system achieves high accuracy (99.3%).

Deep Learning-based Methods for Sign Language Recognition: A Comprehensive Study its author Adaloglou, Nikolaos M, et al [6]. Computer vision-based sign language recognition systems are subjected to a comparative experimental assessment. The most recent deep neural network techniques in this field are used to perform a complete evaluation on a variety of publicly available datasets. By mapping non segmented video streams to glosses, the purpose of this study is to learn more about sign language recognition.

The ArSL Database and Pilot Study: Towards Hybrid Multimodal Manual and Non-Manual Arabic Sign Language Recognition its author Luqman, Hamzah, and El-Sayed M. El-Alfy [7]. A new multi-modality ArSL dataset that combines many modalities. It comprises of 6748 video samples recorded using Kinect V2 sensors of fifty signs performed by four signers. In addition, we used state-of-the-art deep learning algorithms to analyse the integration of spatial and temporal characteristics of distinct modalities, both manual and non-manual, for sign language identification.

Deep Learning Multi Stroke Thai Finger Spelling Sign Language Recognition System its author Pariwat, Thongpan, and Pusadee Seresangtakul [8]. On a complicated backdrop, a vision-based approach was used to accomplish semantic segmentation with dilated convolution for hand segmentation, optical flow separation for hand strokes, and learning feature and classification with a CNN. The five CNN structures that determine the formats were then compared.

Implementing k-Nearest Neighbours with Dynamic Time Warping and Convolutional Neural Network Algorithms in Wearable Electronics for Sign Language Recognition its author Saggio, Giovanni, et al [9]. a wearable electronics-based sign language recognition device with two separate categorization methods A sensory glove and inertial measurement units were used to collect finger, wrist, and arm/forearm motions for the wearable electronics. k-Nearest Neighbours with Dynamic Time Warping (a nonparametric technique) and Convolutional Neural Networks were used as classifiers.

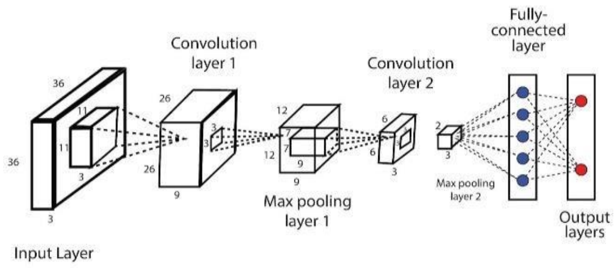
For static signs, a deep learning-based sign language recognition system its author Wadhawan, Ankita, and Parteek Kumar [10]. The paper discusses the use of deep learning-based convolutional neural networks to represent robust static signs in the context of sign language recognition (CNN). A total of 35,000 sign photos of 100 static signs were gathered from various users for this study. On around 50 CNN models, the suggested system's efficiency is assessed.

3. Methodologies and Approaches

The construction of primary base of model is done using following algorithms:

2.1) A Convolutional Neural Network (CNN) is a type of deep neural network used for image recognition and A specific kind of such a deep neural network is the convolutional network. It's a deep, feed-forward artificial neural network. Remember that feed-forward neural networks are also called multi-layer perceptron's (MLPs), which are the quintessential deep learning models. Image classification, object detection, segmentation, face recognition

2.2) Pooling is a technique commonly used to down sample or reduce the spatial dimensions of an image while retaining its essential features. Max pooling involves dividing the input image into non-overlapping rectangular regions (usually 2x2 or 3x3) and selecting the maximum value from each region. The selected maximum values form the output of the pooling layer. Max pooling helps retain the most prominent features in each region and discard less relevant information. Average pooling, on the other hand, involves taking the average value of the pixels within each region. it helps down sample the image, but instead of selecting the maximum value, it uses the average. Average pooling tends to smooth out the features in the down sampled image.



4. Finding and Trends

Sign language detection is to detect the sign language hand gestures, which really helps the common people like is to understand what a deaf or mute people are trying to converse with . The sign language detection translates the sign language, in which user forms a hand shape that is structured signs or gestures. In sign language, the configuration of the fingers, the orientation of the hand, and the relative position of fingers and hands to the body are the expressions of a deaf and mute person. Based on this application, the user must be able to capture images of the hand signs or gestures using web camera and they shall predict the hand signs or meaning of the sign and display the name of sign language on screen. At first, taking sample images of different signs are going to label the images with the Label Image python application file, which is very helpful for object detection. The LabelImg application file develops an XML document for the corresponding image for the training process. In the training process, used TensorFlow object detection API to train the model. After training the model, detected the sign language or hand gestures in real time; with the help of OpenCV-python, the webcam and it loads and trains model, detected the sign languages in real time.

5. Challenges and Gaps

After several experiments on the architecture with different videos, we listed down the gaps we observed and which are important to improve SLT for real-world application. Here are the observed gaps:

A) Limited number of datasets available:

In current research for SL, almost all research papers mention the need for more data to progress the research quality. The datasets available are mostly of alphabets, numbers, and individual words. There are also datasets for Continuous SLR that contain gloss representations for the SL sequences, but, for SLT, spoken translations are also required. There are very few datasets that contain spoken translations as well in the dataset.

The main reason is that the SLT problem is comparatively new and also for spoken translation annotations, human SL interpreters are required to translate the entire video dataset. It is important because the problem of SLT is crucial for real-world applications which connect people with SL knowledge to the ones that do not have this knowledge.

B) Domain restricted data: Most of the benchmark datasets currently present are collected from a certain SL media source which is domain-specific. Like the current benchmark dataset for SLT, the RWTH-Phoenix-Weather 2014T dataset of German Sign Language, contains videos from the daily weather forecast airings of the German public TV station PHOENIX featuring sign language interpretation.

C) Lack of variety in datasets: In the available datasets, there has been a lack of variety in terms of the number of signers, physical orientation of signers, and camera viewpoints of signers. There has been an average of 10–20 signers across various datasets, with the RWTH-Phoenix-Weather 2014T Dataset having just 9 signers. An increased number of native signers gives a better understanding of sign representation. In SL there are different dialects, this makes variations in signs for the same word.

D) Architecture transferability across different SL:

Recently, the amount of research related to SLR/SLT has been increasing. The architectures are capturing various aspects of an SL video sequence. However, after scrutinizing different results from these types of research it is quite apparent that the accuracy results (WER score and BLEU score[5]) are not similar when the same architecture is tried on a different language dataset.

6. Future Research Direction

The future of sign language recognition holds tremendous potential for further advancements. As technology continues to evolve, we can expect:

A) Enhanced Accuracy: Machine learning models will become even better at recognizing nuanced sign language gestures, reducing recognition errors and improving overall accuracy.

B) Wider Implementation: Sign language recognition systems will become more prevalent, integrating into everyday devices and platforms, including smartphones, tab.

C) Educational role: Teachers are constantly searching for new ways to engage their students in the learning process. Using sign language within the classroom is one solution to reach all learners. Sign language can enhance the learning process by bringing visual, auditory and kinaesthetic feedback to help reach all students, and smart home assistants.

7. Conclusion

Sign language provides children with an alternative way to make themselves understood. This extra tool enables them to express how they feel, their thoughts and wants, so that they can take part in learning and social activities. This not only gives a child a 'voice' but is also important when building relationships. So, the primary goal of gesture recognition research is to create systems, which can identify specific human gestures and use them, for example, to convey information. For that, vision- based hand gesture interfaces require fast and extremely robust hand detection, and gesture recognition in real time. Sign Language detection system shows what the position of hands in viewfinder of camera module means with good accuracy. It can then be used to help people who are just beginning to learn Sign Language or those who don't know sign language but have a close one who is deaf.

8. References

[1] G. Li, X. Wang and Y. Liu, "Sign Language Recognition Based on CNN and Data Augmentation,"
2023 IEEE 5th International Conference on Civil Aviation Safety and Information Technology (ICCASIT), Dali, China, 2023.

[2] A. S Sushmitha Urs, V. B Raj, P. S, P. Kumar K, M. B R and V. Kumar S, "Action Detection for Sign Language Using Machine Learning," 2023 International Conference on Network, Multimedia and Information Technology (NMITCON), Bengaluru, India, 2023.

[3] O. Voloshynskyi, V. Vysotska, R. Holoshchuk, S. Goloshchuk, S. Chyrun and D. Zahorodnia, "Sign Language Digits Recognition Technology Based on a Convolutional Neural Network," 2023 IEEE 12th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), Dortmund, Germany, 2023.

[4] V. Kumar, R. Sreemathy, M. Turuk, J. Jagdale and A. Agarwal, "Real-time Indian Sign Language Recognition using Skeletal Feature Maps," 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), Delhi, India, 2023.

[5] P. Singh, S. V. S. Prasad, R. Singh, K. Dasari and B. L. Prasanna, "Development of Sign Language Translator for Disable People in Two-Ways Communication," 2023 1st International Conference on Circuits, Power and Intelligent Systems (CCPIS), Bhubaneswar, India, 2023.

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

[7] 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.

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

[9] 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.

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