International Journal of Early Childhood Special Education (INT-JECSE)
DOI:10.48047/INTJECSE/V15I4.94 ISSN: 1308-5581 Vol 15, Issue 04 2023

SIGN LANGUAGE IDENTIFICATION USING CNN

1Mrs. A.ANITHA REDDY,
Assistant Professor, Department of CSE, Sreyas Institute of Engineering and Technology, Telangana, India, anitha.a@sreyas.ac.in
2 Mareddy Sahani,
Department of CSE, Sreyas Institute of Engineering and Technology, Telangana, India,reddysahani8@gmail.com
3Guddanti Meghana,
DepartmentofCSE,SreyasInstituteofEngineeringandTechnology,Telangana,India, guddanti.meghanaaa@gmail.com
4 Shaik Aslam Pasha,
Department of CSE, Sreyas Institute of Engineering and Technology, Telangana,India, aslampashashaik2@gmail.com
5 Tarun Madhav Medi,
Department of CSE, Sreyas Institute of Engineering and Technology, Telangana, India, madhavtarun35@gmail.com

ABSTRACT
Sign language is a vital mode of communication for individuals with hearing impairments, facilitating their interaction with the broader society. Communication and accessibility for the deaf and hard of hearing community. In this study, we propose a Sign Language Identification (SLI) system based on Convolutional Neural Networks (CNNs) to recognize and interpret different sign gestures. We evaluate the SLI system on a separate test set, and the results indicate high accuracy and robustness in recognizing sign language gestures. The experimental outcomes demonstrate the potential of deep learning techniques, specifically CNNs, in developing accurate and efficient sign language identification systems. The proposed model could serve as a foundation for future research in sign language translation, interpretation, and accessibility applications.

INTRODUCTION
In our increasingly interconnected world, effective communication is fundamental to fostering understanding and inclusivity among diverse communities.For individuals with hearing impairments, sign language serves as a vital means of expression and communication. However, the recognition and interpretation of sign language have long posed challenges in the realm of technology. With the advent of artificial intelligence and deep learning techniques, particularly Convolutional Neural Networks (CNN), there has been a revolutionary leap in the domain of sign language recognition. This innovation has the potential to bridge the communication gap between the deaf and hearing communities, ensuring equal access to information and services.Sign language, a complex and nuanced form of communication, involves intricate hand movements, facial expressions, and body postures. Recognizing and translating these gestures accurately require advanced computer vision techniques. Traditional methods often fell short in capturing the subtleties of sign language, leading to limited accessibility for the deaf and hard-of-hearing individuals.

Moreover, sign language varies across regions and cultures, making it a challenge to develop universal recognition systems. However, the rise of deep learning, specifically CNN, has provided a ray of hope in overcoming these hurdles.CNN, a class of deep neural networks, has revolutionized the field of computer vision. Unlike traditional algorithms, CNNs can automatically learn and extract intricate patterns from images. They are particularly well-suited for tasks involving visual data, such as image recognition and classification. CNNs consist of multiple layers of interconnected neurons, each performing specific operations like convolution, pooling, and fully connected layers. These layers enable CNNs to recognize hierarchical features in images, making them exceptionally effective in sign language recognition.Utilizing CNN for sign language identification involves training the network on a diverse dataset of sign language gestures. The CNN model learns to identify unique features and patterns within these gestures, enabling it to accurately classify and interpret new, unseen

signs. The training process involves iterative optimization, where the network adjusts its parameters based on the disparity between its predictions and the actual labels. With sufficient training, the CNN becomes adept at recognizing sign language gestures with a high degree of accuracy.While CNNs offer a promising solution, challenges persist in sign language identification. Variability in signing styles, lighting conditions, and background clutter can affect the accuracy of recognition systems. Researchers have been addressing these challenges through innovations such as data augmentation, where the training dataset is artificially expanded by applying transformations like rotation and scaling to mitigate the impact of variations. Additionally, transfer learning, a technique where a CNN model pre-trained on a large dataset is fine-tuned for sign language recognition, has shown remarkable results in enhancing accuracy and efficiency.

The integration of CNN-based sign language recognition systems has transformative potential across various sectors. In education, these systems facilitate inclusive learning environments, enabling deaf students to participate actively in classrooms. Moreover, in healthcare, they enhance communication between healthcare providers and deaf patients, ensuring accurate understanding of medical information and instructions. In public services, CNN-driven sign language recognition can be deployed in customer service interfaces, making essential services accessible to individuals with hearing impairments. Furthermore, in emergency situations, swift and accurate communication through sign language recognition can save lives.The integration of Convolutional Neural Networks in sign language identification represents a significant milestone in fostering inclusivity and breaking down communication barriers. By harnessing the power of deep learning, researchers and developers are making strides toward ensuring equal access to information, education, and services for the deaf and hard-of-hearing communities. As technology continues to evolve, the synergy between artificial intelligence and sign language recognition holds the promise of a more inclusive and empathetic society, where every individual, regardless of their abilities, can fully participate and contribute to the global conversation.

LITERATURE SURVEY

Literature review of our proposed system shows that there have been many explorations done to tackle the sign recognition in videos and images using several methods and algorithms.

Siming He[4] proposed a system having a dataset of 40 common words and 10,000 sign language images. To locate the hand regions in the video frame, Faster R-CNN with an embedded RPN module is used. It improves performance in terms of accuracy. Detection and template classification can be done at a higher speed as compared to single stage target detection algorithm such as YOLO. The detection accuracy of Faster R-CNN in the paper increases from 89.0% to 91.7% as compared to Fast-RCNN. A 3D CNN is used for feature extraction and a sign-language recognition framework consisting of long and short time memory (LSTM) coding and decoding network are built for the language image sequences. On the problem of RGB sign language image or videorecognition in practical problems , the paper merges the hand locating network, 3D CNN feature extraction network and LSTM encoding and decoding to construct the algorithm for extraction. This paper has achieved a recognition of 99% in common vo- cabulary dataset.

Let's approach the research done by Rekha, J[5]. which made use of YCbCr skin model to detect and fragment the skin region of the hand gestures. Using Principal Curvature based Region Detector, the image features are extracted and classified with Multi class SVM, DTW and non-linear KNN. A dataset of 23 Indian Sign Language static alphabet signs were used for training and 25 videos for testing. The experimental result obtained were 94.4%for static and 86.4% for dynamic.

In [6] , a low cost approach has been used for image processing . The capture of images was done with a green background so that during processing, the green colour can be easily subtracted from the RGB colour space and the image gets converted to black and white. The sign gestures were in Sinhala language. The method that thy have proposed in the study is to map the signs using centroid method. It can map the input gesture with a database irrespective of the hands size and position. The prototype has correctly recognised 92% of the sign gestures.

The paper by M. Geetha and U. C. Manjusha[7], make use of 50 specimens of every alphabets and digits in a vision based recognition of Indian Sign Language characters and numerals using B-Spine approximations. The region of interest of the sign gesture is analysed and the boundary is removed . The boundary obtained is further transformed to a B-spline curve by using the Maximum Curvature

Points(MCPs) as the Control points. The B-spline curve undergoes a series of smoothening process so fea- tures can be extracted. Support vector machine is used to classify the images and the accuracy is 90.00%.

In [8], Pigou used CLAP14 as his dataset [9]. It consists of 20 Italian sign gestures. After preprocessing the images , he used a Convolutional Neural network model having 6 layers for training. It is to be noted that his model is not a 3D CNN and all the kernels are in 2D. He has used Rectified linear Units (ReLU) as activation functions. Feature extraction is performed by the CNN while classification uses ANN or fully connected layer. His work has achieved an accuracy of 91.70% with an error rate of 8.30%. A similar work was done by J Huang [10].He created his own dataset using Kinect and got a total a total of 25 vocabularies which are used in everyday lives. He then applied a 3D CNN in which all kernels are also in 3D. The input of his model consisted
of 5 important channels which are colour-r, colour-b, colour-g, depth and body skeleton. He got an average accuracy of 94.2%.

Another research paper on Action recognition topic by the author J.Carriera [11] shares some similarities to sign gesture recognition .He used a transfer learning method for his research As his pre-trained dataset, he used both ImageNet[12] and Kinetic Dataset [9] . After training the pertained models using another two datasets namely UCF-101 [13] and HMDB-51 [14], he then merged the RGB model, flow model, pre-trained Kinetic and pre-trained ImageNet. The accuracy he got on UCF-101 dataset is 98.0% and on HMDB-51 is 80.9%.

PROPOSED SYSTEM

Our proposed system is sign language recognition system using CONVOLUTION NEURAL NETWORK which recognizes various hand gestures by capturing video and converting it into frames. Then the hand pixels are segmented and the image it obtained and sent for comparison to the trained model. Thus our system is more robust in getting exact text labels of letters.The goal is to enable real-time recognition and interpretation of sign language gestures, helping bridge the communication gap between the hearing-impaired and the hearing community. Thus our system is more robust in getting exact text labels of letters.

Image classification is the process of taking an input (like a picture) and outputting its class or probability that the input is a particular class.

Convolution Operation: In purely mathematical terms, convolution is a function derived from two given functions by integration which expresses how the shape of one is modified by the other.

Convolution formula:

$$(f * g)(t) \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} f(\tau)\, g(t-\tau)\, d\tau$$

Here are the three elements that enter into the convolution operation:

• Input image
• Feature detector
• Feature map

Steps to apply convolution layer:

• You place it over the input image beginning from the top-left corner within the borders you see demarcated above, and then you count the number of cells in which the feature detector matches the input image.
• The number of matching cells is then inserted in the top-left cell of the feature map
• You then move the feature detector one cell to the right and do the same thing. This movement is called a and since we are moving the feature detector one cell at time, that would be called a stride of one pixel.
• What you will find in this example is that the feature detector's middle-left cell with the number 1 inside it matches the cell that it is standing over inside the input image. That's the only matching cell, and so you write “1” in the next cell in the feature map, and so on and so forth. 33
• After you have gone through the whole first row, you can then move it over to the next row and go through the same process. There are several uses that we gain from deriving a feature map. These are

the most important of them: Reducing the size of the input image, and you should know that the larger your strides (the movements across pixels), the smaller your feature map.

Relu Layer:

Rectified linear unit is used to scale the parameters to non-negative values. We get pixel values as negative values too. In this layer we make them as 0's. 30 The purpose of applying the rectifier function is to increase the non linearity in our images. The reason we want to do that is that images are naturally nonlinear. The rectifier serves to break up the linearity even further in order to make up for the linearity that we might impose an image when we put it through the convolution operation. What the rectifier function does to an image like this is remove all the black elements from it, keeping only those carrying a positive value (the grey and white colors). The essential difference between the non-rectified version of the image and the rectified one is the progression of colors. After we rectify the image, you will find the colors changing more abruptly. The gradual change is no longer there. That indicates that the linearity has been disposed of.

Pooling Layer:

The pooling (POOL) layer reduces the height and width of the input. It helps reduce computation, as well as helps make feature detectors more invariant to its position in the input This process is what provides the convolutional neural network with the "spatial variance" capability. In addition to that, pooling serves to minimize the size of the images as well as the number of parameters which, in turn, prevents an issue of "overfitting" from coming up. Overfitting in a nutshell is when you create an excessively complex model in order to account for the idiosyncrasies, we just mentionedthe result of using a pooling layer and creating down sampled or pooled feature maps isa summarized version of the features detected in the input. They are useful as small changes in the location of the feature in the input detected by the convolutional layer will result in a pooled feature map with the feature in the same 34 location. This capability added by pooling is called the model's invariance to local translation.

Fully Connected Layer:

The role of the artificial neural network is to take this data and combine the features into a wider variety of attributes that make the convolutional network more capable of classifying images, which is the whole purpose from creating a convolutional neural network. It has neurons linked to each other, and activates if it identifies patterns and sends signals to output layer. the output layer gives output class based on weight values, for now, all you need to know is that the loss function informs us of how accurate our 31 network is, which we then use in optimizing our network in order to increase its effectiveness. That requires certain things to be altered in our network. These include the weights (the blue lines connecting the neurons, which are basically the synapses), and the feature detector since the network often turns out to be looking for the wrong features and has to be reviewed multiple times for the sake of optimization. This full connection process practically works as follows:

- The neuron in the fully-connected layer detects a certain feature; say, a nose.
- It preserves its value.
- It communicates this value to the classes trained images.

DOI:10.48047/INTJECSE/V15I4.94 ISSN: 1308-5581 Vol 15, Issue 04 2023

RESULTS

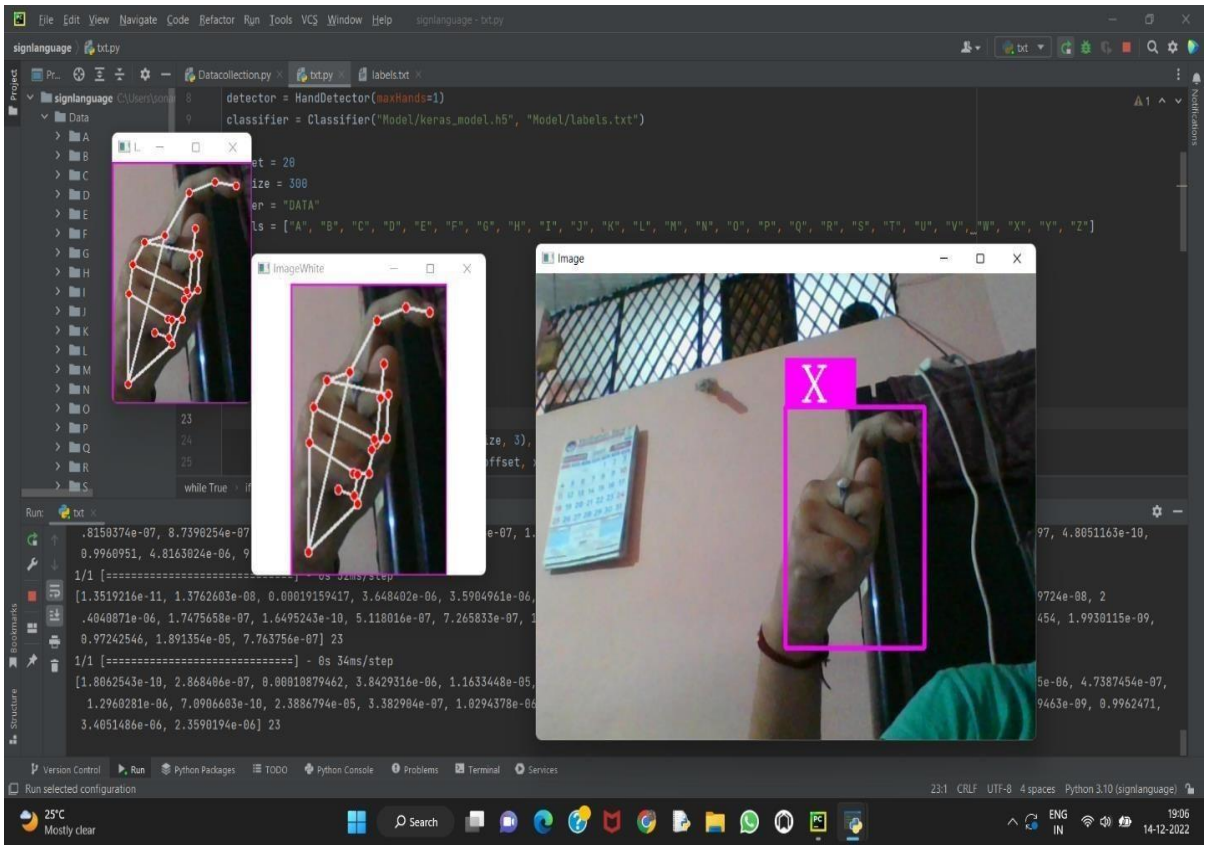

Figure 9.4 Output for letter X

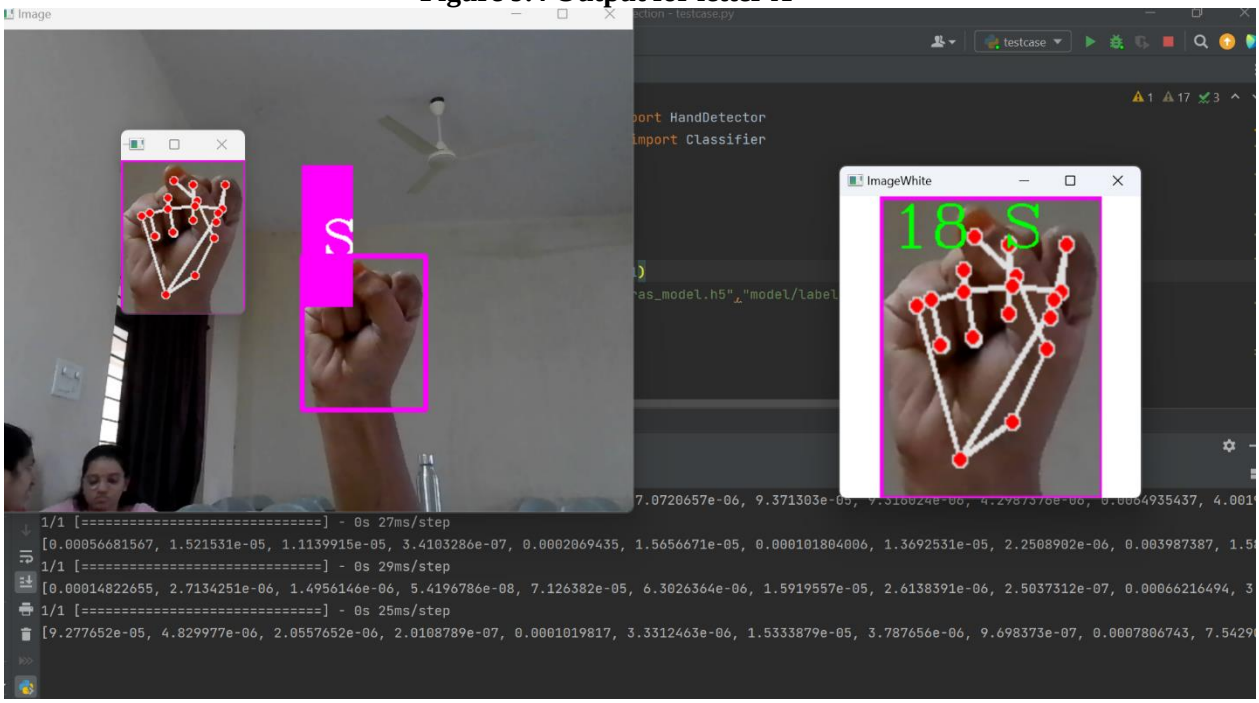

Figure 9.11 Output letter for S

CONCLUSION

Nowadays, applications need several kinds of images as sources of information for elucidation and analysis. Several features are to be extracted so as to perform various applications. When an image is transformed from one form to another such as digitizing, scanning, and communicating, storing, etc. degradation occurs. Therefore, the output image has to undertake a process called image enhancement, which contains of a group of methods that seek to develop the visual presence of an image. Image enhancement is fundamentally enlightening the interpretability or awareness of information in images for human listeners and providing better input for other automatic image processing systems. Image then undergoes feature extraction using various methods to make the image more readable by the computer. Sign language recognition system is a powerful tool to prepare an expert knowledge, edge detect and the combination of inaccurate information from different sources. Using this, there can be a proper communication between deaf and dumb people.

REFERENCES

1. Sengupta, S., & Chakraborty, S. (2018). Sign Language Recognition Using Convolutional Neural Networks. 2018 International Conference on Communication and Signal Processing (ICCSP).
2. Dutta, A., & Choudhury, N. (2017). Sign Language Recognition Using Deep Learning Models. 2017 International Conference on Innovations in Electronics, Signal Processing and Communication (IESC).
3. Roy, A., & Bhattacharya, I. (2017). Sign Language Recognition Using Convolutional Neural Network. 2017 8th Annual Industrial Automation and Electromechanical Engineering Conference (IEMECON).
4. Arya, S., & Ahuja, N. (2018). American Sign Language Recognition Using Convolutional Neural Network. 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA).
5. Wang, Z., Zhang, Y., & Zhang, T. (2017). Sign Language Recognition Based on CNN-BiLSTM Model. 2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI).
6. Nguyen, H. V., & Duong, A. D. (2018). Vietnamese Sign Language Recognition Using CNN and LSTM. 2018 15th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON).
7. Puertas, E., Orts-Escolano, S., & Garcia-Rodriguez, J. (2018). Sign Language Recognition with 3D Convolutional Neural Networks. 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS).
8. Zhang, P., & Zhang, L. (2017). Sign Language Recognition Based on Temporal Pyramid Pooling Network. 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).
9. Li, X., Liu, Z., & Zhang, L. (2018). Sign Language Recognition Using Multi-View Convolutional Neural Networks. IEEE Transactions on Multimedia, 20(6), 1455-1466.
10. Wu, Y., & Chen, J. (2016). Chinese Sign Language Recognition Using Temporal Segment Network. 2016 12th World Congress on Intelligent Control and Automation (WCICA).
11. Girija, R., & Sruthi, R. (2018). Sign Language Recognition System Using 3D CNN and LSTM Networks. 2018 International Conference on Smart Electronics and Communication (ICOSEC).
12. Liwicki, M., Tzimiropoulos, G., & Zafeiriou, S. (2012). Continuous Sign Language Recognition: Towards Large Vocabulary Statistical Recognition Systems Handling Multiple Signers. Computer Vision and Image Understanding, 116(3), 303-320.
13. Zhang, Y., Xu, C., & Yang, J. (2017). Sign Language Recognition Based on Improved 3D CNN. 2017 2nd International Conference on Automation, Control and Robotics Engineering (CACRE).
14. Huang, W., & Gu, I. Y. H. (2016). Sign Language Recognition Using Sub-UNets. 2016 IEEE International Conference on Image Processing (ICIP).
15. Pfister, T., Charles, J., & Zisserman, A. (2014). Flowing ConvNets for Human Pose Estimation in Videos. 2014 IEEE Conference.