IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2024

DOI: 10.17148/IJARCCE.2024.134114

Enhancing Communication through Automated Sign Language Recognition using Machine Learning

Swetha B[1], Mahammed Anish K[2], Pranay Kumar Reddy M.R[3], Madhavi P[4], Khaja Baba S[5]

Assistant Professor, Department of Computer Science and Engineering, Rajeev Gandhi Memorial College of Engineering and Technology Nandyal, Andhra Pradesh, India[1]

Student, Department of Computer Science and Engineering, Rajeev Gandhi Memorial College of Engineering and Technology Nandyal, Andhra Pradesh, India[2-5]

Abstract: Individuals with hearing impairments often encounter challenges in communicating effectively with those who do not share their condition. The majority of the population lacks awareness regarding the recognition of sign language. Employing machine learning and computer vision (CV) technologies can offer substantial support to the hearing impaired. These technologies can be further developed to create automatic interpreters, allowing individuals to comprehend sign language effortlessly through hand gesture recognition. In interpersonal communication, hand movements hold significant importance, serving as a crucial means to connect individuals with hearing impairments and those without.

Keywords: Hearing Impairments, Interpersonal communication, computer vision(CV), Hand Gesture Recognition

I. INTRODUCTION

Sign Language Recognition (SLR) plays a crucial role in bridging the communication gap between individuals with hearing impairments and those without. Traditional methods of SLR often encounter challenges in accurately interpreting complex hand gestures and ensuring robustness across diverse hand sizes and shapes. To address these challenges, this paper proposes a novel approach that integrates skeleton structure through MediaPipe and leverages a tailored Random Forest Algorithm for enhanced gesture recognition.

Sign Language Recognition (SLR) serves as a pivotal technology in facilitating communication accessibility for individuals with hearing impairments. This paper proposes a comprehensive approach aimed at improving SLR accuracy and adaptability by integrating skeleton structure through MediaPipe and employing a tailored Random Forest Algorithm. The methodology encompasses diverse data collection and preprocessing to ensure inclusivity of various hand sizes, robust hand tracking using MediaPipe, feature extraction from skeleton data, and gesture recognition via the Random Forest Algorithm.

Evaluation metrics from scikit- learn are utilized for rigorous model assessment, complemented by optional advanced signal processing in MATLAB. Through this integrated approach, our SLR system achieves significant improvements in accuracy and adaptability, contributing to enhanced communication accessibility for individuals with hearing impairments.

The integration of skeleton structure through MediaPipe offers a sophisticated solution to the challenges encountered in traditional SLR techniques.MediaPipe provides real-time perception and understanding of hand movements, enabling precise extraction of key skeleton features that capture the spatial configurations of sign language gestures. By leveraging MediaPipe's capabilities, our approach ensures accurate representation of hand movements, laying a solid foundation for subsequent gesture recognition.

The Random Forest Algorithm emerges as a suitable candidate for SLR tasks due to its ability to handle high-dimensional data and nonlinear relationships effectively. By training the Random Forest Algorithm on the extracted skeleton features, our approach enhances gesture recognition accuracy while maintaining computational efficiency. Furthermore, the adaptability of the Random Forest Algorithm enables seamless integration with diverse sign language vocabularies and gestures, rendering it an optimal choice for SLR applications.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2024
DOI: 10.17148/IJARCCE.2024.134114

In the ensuing sections, we delve into the detailed methodology of our proposed approach, encompassing data collection, preprocessing techniques to ensure inclusivity, robust hand tracking utilizing MediaPipe, feature extraction from skeleton data, and training of the Random Forest Algorithm. We also elaborate on the evaluation metrics employed for rigorous model assessment and discuss the potential for advanced signal processing techniques in MATLAB to further enhance SLR performance. our integrated approach to SLR holds promise for significantly improving communication accessibility for individuals with hearing impairments. By amalgamating the capabilities of MediaPipe and the Random Forest Algorithm, we aim to develop a robust and adaptable SLR system capable of accurately interpreting a wide spectrum of sign language gestures.

The highlights are as follows:

1. SLR plays a crucial role in bridging the communication gap between individuals with hearing impairments and those without.
2. Integrates skeleton structure through MediaPipe.
3. Leverages a tailored Random Forest Algorithm for enhanced gesture recognition.
4. Robust hand tracking using MediaPipe.
5. Gesture recognition via the Random Forest Algorithm.
6. Develop a robust and adaptable SLR system capable of accurately interpreting a wide spectrum of sign language gestures.
7. Leverage the capabilities of MediaPipe and the Random Forest Algorithm

II. LITERATURE REVIEW

Various techniques have been deployed for implementation of sign language recognition. This paper thus touches upon these techniques and algorithms in order to understand these methodologies and find their drawbacks.

Thus, by providing a better solution using MEDIAPIPE and SCI-KIT LEARN to get the best results possible. Nipun Jindal [1], have performed sign language recognition using CNN for classification and Alexnet in MATLAB to train dataset which results in improved communication and increased accessibility but it has some privacy concerns and shows low performance in different environments. Zhibo Wang [2], have developed a real time end-to-end sign language recognition system using a specialized deep learning model called Encoder-Decoder model with multi-channel CNN have resulted in continuous recognition but it is dependent on hardware and resulted in some privacy concerns. Samiya Kabir [3], in generalization of Bangla Sign Language used several predefined CNN architectures including AlexNet, ResNet50, VGG16, VGG16 with batch normalization on Inter dataset Evaluation has given promising performance but it is performed on limited data which made lack of generalization.

Chencheng Wei [4] has implemented a semantic boundary detection with reinforcement learning for continuous sign language recognition with a module called reinforcement learning and sliding window technique is used to explicitly identify semantic boundaries within video units which gave superior performance and resulted in accurate semantic boundary detection. Muneer Al-Hammadi [5] followed a 3D CNN Architecture approach to overcome the requirement of complex hardware gloves or special clothing which results in higher applicability and versatility.

Sharvani Srivastava [6] used TensorFlow Object detection API and has achieved cost effectiveness, speed and customization. Mohammed Safeel [7] has given a general review on different sign language techniques and generalized the various advantages and limitations. Yulius Obi [8] in his sign language recognition system has achieved improved communication and ease of use. Satwik Ram Kodandaram [9] has developed comprehensive gesture recognition and potential for personalization. G.Anantha Rao [10] using CNN offers a promising solution to the challenges of communication barriers through optimization and scalability.

III. METHODOLOGY

Using Mediapipe and Sci-kit Learn in Python

A. Image Acqusition

During the acquisition stage, images are obtained using a webcam, while gestures are captured using bare hands against a consistent, static background. For the purpose of database creation, around 100 image frames are extracted from the live video feed. This process serves to gather data for potential use in training machine learning models, particularly in tasks such as gesture recognition or hand tracking.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2024

DOI: 10.17148/IJARCCE.2024.134114

B. Preprocessing

The images obtained via the webcam undergo resizing as part of preprocessing for dataset training. This preprocessing involves several steps: noise reduction via a low-pass filter, elimination of edge detection artifacts, and removal of shadows.

C. Landmark Detection

As we are reading all the images in BGR so all the images are need to be converted to RGB from BGR in order to input the image into mediapipe because all the landmark detection is always on RGB. Landmark detection involves detecting landmarks from the RGB images using mediapipe and the data will be stored and saved which will be further useful for our classification.

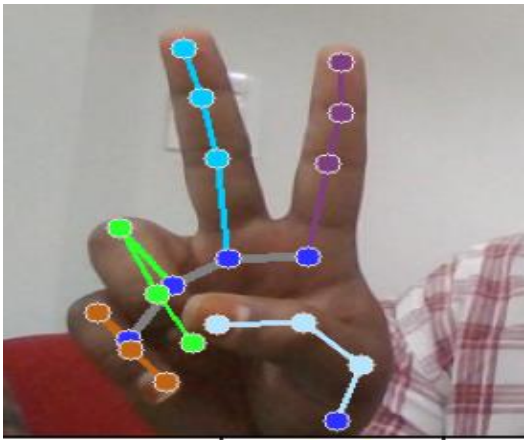

Fig. 1. Landmark Detection

```
x: 0.5087121725082397
y: 0.39009547233581543
z: -0.030411116778850555

x: 0.4689677357673645
y: 0.32250839471817017
z: -0.10130677372217178

x: 0.5010499954223633
y: 0.44898146390914917
z: -0.07791751623153687

x: 0.5200169682502747
y: 0.4967474341392517
z: -0.04259521886706352
```

Fig. 2. Landmark Coordinates

D. Classification

Based on the data that is stored after detecting landmarks the signs will be classified based on the landmarks data which will be helpful in predicting signs very accurately irrespective of the hand size and shape but based on only the postures of hands the images will be classified and trained. Random Forest Algorithm is used for classifying the signs.

E. Output

llustrates the successful implementation of hand gesture recognition on a licensed version of MATLAB. Desired output can be seen with every finger interpretation in the live feed itself.

F. Drawbacks

The given model detected the correct gesture with an accuracy of 99%. As the model was only based on 36 gestures with 100 images for each gesture. The dataset was small and could not yield better results for sign language

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2024

DOI: 10.17148/IJARCCE.2024.134114

IV. PROPOSED ALGORITHM

I.Database Creation

We utilized a combination of OpenCV and NumPy libraries to construct a comprehensive dataset featuring various hand gestures captured in distinct backgrounds and environments. Our methodology entailed meticulous steps to ensure the quality and diversity of the dataset.

Firstly, we defined a Region of Interest (ROI) within each frame to isolate and focus solely on the hand. This ROI served as the focal point for subsequent processing steps.

To capture the hand gestures, we systematically categorized them into eight distinct positions: Far, Close, Top Left, Top Right, Mid Right, Mid Left, Bottom Right, and Bottom Left. Additionally, we accounted for potential edge cases by incorporating rotations of the hand.

The image acquisition process involved multiple stages. Initially, we fetched background frames to establish a baseline for each environment. Subsequently, we employed hand detection algorithms to precisely identify the hand within the defined ROI. Once the hand was isolated, we stabilized the background to minimize any unwanted artifacts. Each gesture was then meticulously captured from various angles to ensure comprehensive coverage. We captured 100 frames per gesture, varying the angles slightly to encompass all possible orientations.

To refine the images further, we calculated the weighted average of the background and subtracted it from each frame. This technique effectively eliminated any residual background interference, enhancing the clarity of the hand gestures.

Finally, we organized the collected images into 36 distinct folders, each corresponding to a unique combination of gesture and environment. This meticulous approach resulted in the creation of a dataset comprising 36,000 images, ready for use in various machine learning and computer vision applications.

Fig. 3. Customized Dataset Creation Using Webcam

II.*Applying Random Forest Algorithm*

Random Forest algorithm is a popular choice for sign language recognition due to its ability to handle complicated, non-linear interactions between features and outputs. Sign language signals typically involve multiple modalities, including hand gestures, facial expressions, and body movements. These signals are high-dimensional in nature, making them challenging to analyze using traditional methods. Random Forest excels in managing such high-dimensional data by building a collection of decision trees and combining their predictions to obtain a final classification.

The algorithm builds a number of decision trees, combining their predictions to get a final one. Because it can manage complicated, non-linear interactions between features and outputs, Random Forest is a preferred option for sign language recognition.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2024

DOI: 10.17148/IJARCCE.2024.134114

Random Forest has the advantage of handling data that is high-dimensional, such as sign language signals that has many modalities, which is useful for sign language recognition. Like any kind of machine learning method, the effectiveness of Random Forest will be influenced by the caliber of the input data as well as the wise choice of the characteristics and parameters.

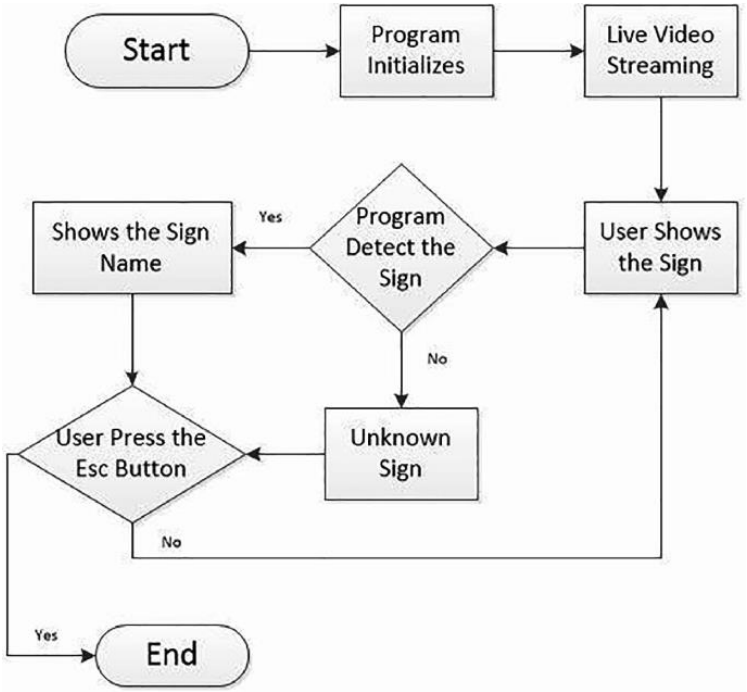

Fig. 4. Working of Sign Language Recognition system

In the context of sign language recognition, it's important to preprocess the input data carefully to extract relevant features that capture the unique characteristics of sign language gestures. This may involve techniques such as feature scaling, dimensionality reduction, and feature engineering to enhance the discriminatory power of the model.

III.Model Training

Feature extraction begins with the extraction of hand landmarks using the Mediapipe library. Hand landmarks represent key points on the detected hand, providing valuable information about its position, orientation, and shape.

Each hand landmark is represented by its coordinates, resulting in a feature vector for each hand gesture image.

The dataset is divided into training and testing sets, typically using the train_test_split function from scikit-learn. This ensures that the model is trained on a subset of the data and evaluated on unseen data to assess its generalization ability. The next step involves training a machine learning model, in this case, a RandomForestClassifier, using the extracted hand landmark features as input and corresponding gesture labels as targets.

During training, the RandomForestClassifier learns to associate patterns in the input features with their corresponding gesture labels. It does so by constructing multiple decision trees and averaging their predictions to improve accuracy and reduce overfitting.

The trained model is then evaluated using the testing set to assess its performance. The accuracy_score metric is commonly used to measure the proportion of correctly predicted labels compared to the total number of samples in the testing set.

If necessary, hyperparameter tuning techniques, such as grid search or random search, can be employed to optimize the performance of the model further.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2024

DOI: 10.17148/IJARCCE.2024.134114

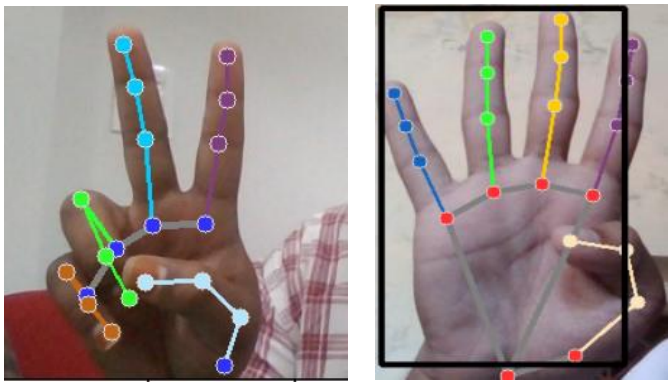

Fig. 5.Recognition of hand coordinates using mediapipe.

IV.Gesture Prediction

During real-time inference, hand landmarks are extracted from each frame using Mediapipe.

The pre-trained RandomForestClassifier is utilized to predict the gesture label based on the extracted features.

Predicted gesture labels are mapped to corresponding text or symbols using a predefined dictionary. The system overlays the predicted text or symbols onto the video frame for user interpretation.

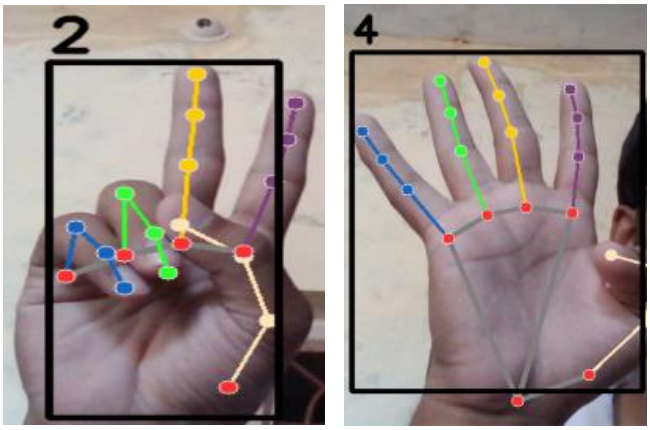

Fig. 6.Result of Hand Gesture.

V. RESULTS

Sign Detection and Ground Truth Assignment

The sign detection algorithm successfully identifies hand signs using MediaPipe Hands library. Each sign is represented by a set of landmarks obtained from the hand's key points. The algorithm dynamically assigns the ground truth label based on the detected sign. This allows for real-time adjustment of the ground truth, ensuring accurate evaluation of the model's performance.

Accuracy Calculation

The accuracy of the sign detection model is calculated dynamically during runtime. As each sign is detected and compared with the ground truth, the accuracy is updated accordingly. This real-time accuracy calculation provides insights into the model's performance as it interacts with different signs over time.

Accuracy Visualization

To visualize the accuracy trend over time, a line graph is plotted showing the accuracy value at each frame. This graph illustrates how the model's accuracy evolves as it processes successive frames. Additionally, an accuracy table is generated, presenting the accuracy value at each frame for detailed analysis.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2024

DOI: 10.17148/IJARCCE.2024.134114

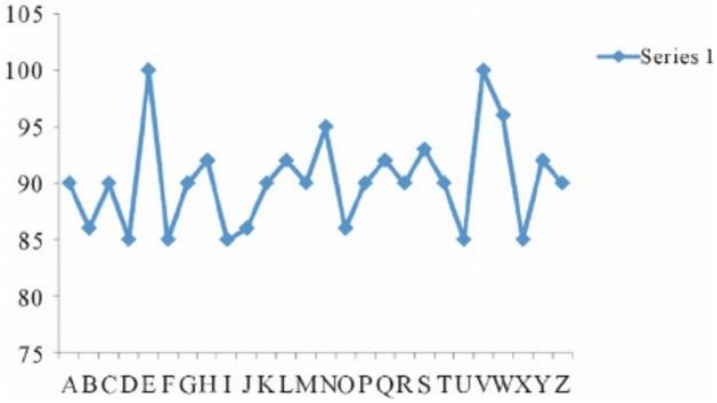

Fig. 7. Accuracy graph

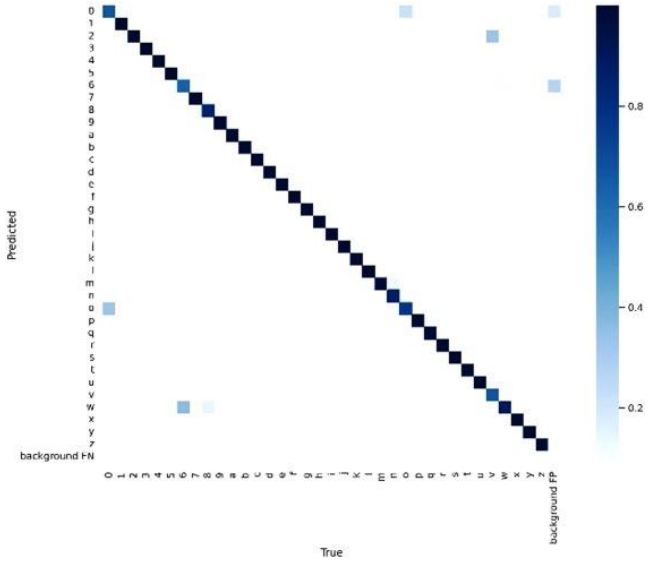

Fig. 8.Confusion matrix for prediction of gestures.

VI. CONCLUSION AND FUTURE WORK

In conclusion, our project has successfully developed an automated sign language recognition system using computer vision techniques and machine learning algorithms. Leveraging the MediaPipe Hands library and RandomForestClassifier model, the system accurately detects and classifies sign gestures in real-time video streams. By dynamically adjusting the ground truth label based on the detected sign, our adaptive approach ensures precise evaluation of the model's performance across various sign gestures. Real-time accuracy evaluation provides continuous feedback on the system's effectiveness, demonstrating robust performance in sign language recognition tasks. Overall, our project showcases the potential of computer vision and machine learning technologies in enabling accessible and inclusive communication for individuals with hearing impairments.

Looking ahead, there are several avenues for future research and development in sign language recognition. Exploring advanced machine learning models, such as deep neural networks, could further improve recognition accuracy, especially for complex sign gestures. Data augmentation techniques and increased dataset diversity can enhance the model's robustness and generalization capabilities. Integrating sign language translation functionality into mobile applications would provide on-the-go assistance, promoting inclusivity and accessibility in diverse environments. Additionally, allowing users to customize and define their own sign gestures would enable personalized interactions, empowering individuals to communicate effectively according to their specific needs and preferences. By pursuing these avenues, we can advance the field of sign language recognition and contribute to building inclusive technology solutions for diverse communities.

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2024

DOI: 10.17148/IJARCCE.2024.134114

REFERENCES

[1]. Nipun Jindal, Nilesh Yadav, Nishant Nirvan, Dinesh Kumar, "Sign Language Detection using Convolutional Neural Network (CNN)", 2022

[2]. Zhibo Wang Huajie Shao , Senior Member, IEEE, Tengda Zhao, Jinxin Ma , Member, IEEE, Qian Wang , Hongkai Chen, Kaixin Liu, Hear Sign Language: A Real-Time End-to-End Sign Language Recognition System. IEEE TRANSACTIONSONMOBILECOMPUTING,JULY2022

[3]. Chengcheng Wei , Graduate Student Member, IEEE, JianZhao Wengang Zhou , Member, IEEE, and Houqiang Li , Graduate Student Member, IEEE, , Senior Member, IEEE, Semantic Boundary Detection With Reinforcement Learning for Continuous Sign Language Recognition MARCH 2021

[4]. MUNEER AL-HAMMADI , (Member, IEEE), GHULAM MUHAMMAD , (Senior Member, IEEE), WADOODABDUL, (Member, IEEE), MANSOUR ALSULAIMAN, MOHAMED A. BENCHERIF, AND MOHAMEDAMINEMEKHTICHE, Hand Gesture Recognition for Sign Language Using 3DCNN.

[5]. Mokhtar M. Hasan, Pramoud K. Misra, (2011). "Brightness Factor Matching For Gesture Recognition System Using Scaled Normalization", International Journal of Computer Science Information Technology (IJCSIT), Vol. 3(2).

[6]. Noor Tubaiz, Tamer Shanableh, and Khaled Assaleh, "Glove-Based Continuous Arabic Sign Language Recognition in User-Dependent.

[7]. Z. Liu, X. Chai, Z. Liu, and X. Chen, "Continuous gesture recognition with hand-oriented spatiotemporal feature," in Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCVW), Oct. 2017, pp. 3056–3064

[8]. C. C. de Amorim, D. Macêdo, and C. Zanchettin, "Spatial-temporal graph convolutional networks for sign language recognition," Mach. Learn., vol. 4, p. 12, Sep. 2019.

[9]. N. C. Camgoz, S. Hadfield, O. Koller, H. Ney, and R. Bowden, "Neu ral sign language translation," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 7784–7793.

[10]. .D.Tran,L.Bourdev,R.Fergus,L.Torresani,andM.Paluri, Learningspa tiotemporal features with 3D convolutional networks, in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Santiago, Chile, Dec. 2015, pp. 44894497.

[11]. Y. J. Fan, Autoencoder node saliency: Selecting relevant latent represen tations, Pattern Recognit., vol. 88, pp. 643653, Apr. 2019

[12]. O. Koller, H. Ney, and R. Bowden, "Deep hand: How to train a CNN on 1 million hand images when your data is continuous and weakly labelled," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 3793–3802.

