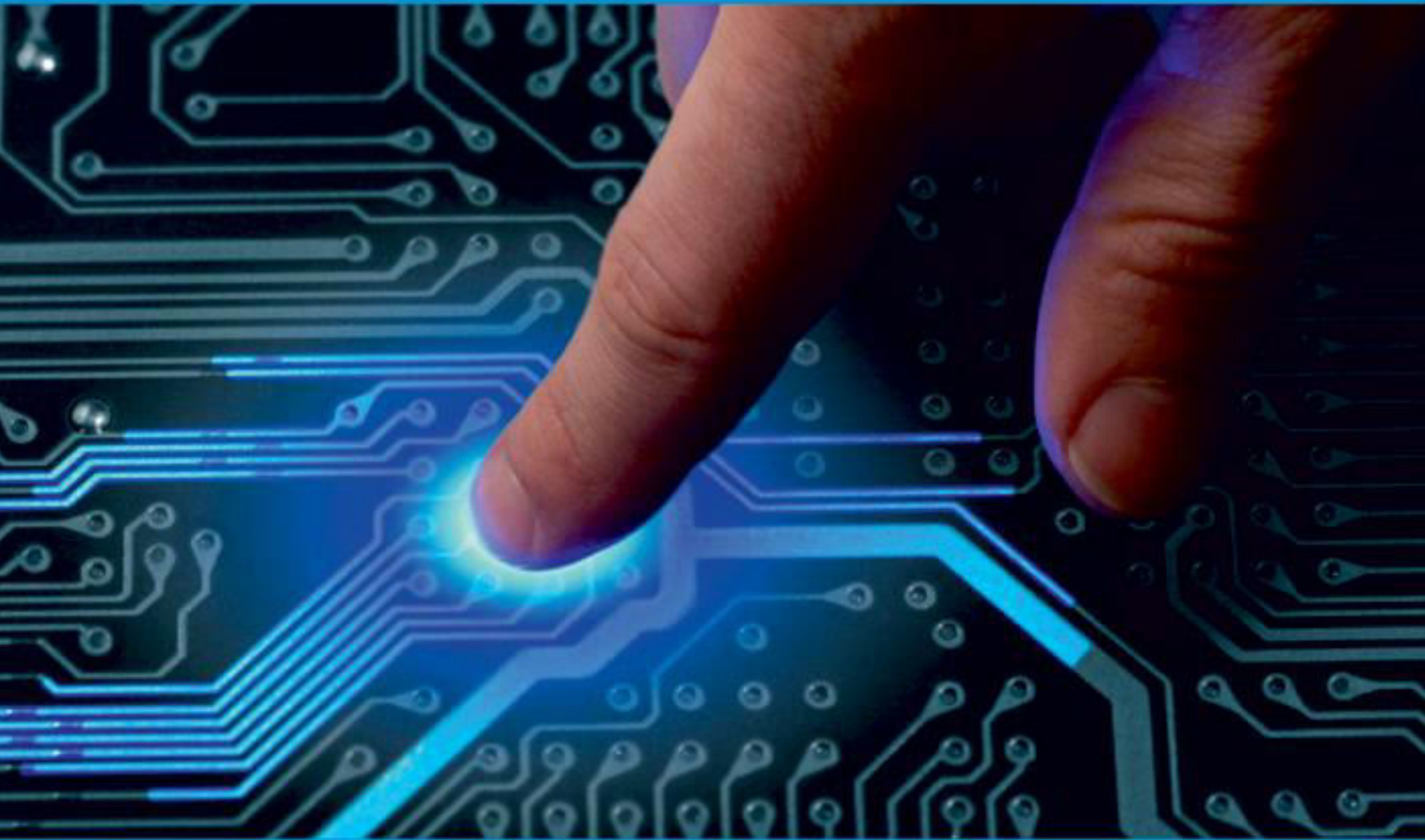
IJIRCCE

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



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 12, December 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.625

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Sign Language Recognition using Spatial Vectorization

Divya K N, Ranga Surya Teja, S V Vivek Reddy, Chandra Prakash Naidu K R,

Afifa Salsabil Fathima A

U.G. Student, Department of Computer Engineering, Reva University, Sathanur, Bangalore, Karnataka, India

U.G. Student, Department of Computer Engineering, Reva University, Sathanur, Bangalore, Karnataka, India

U.G. Student, Department of Computer Engineering, Reva University, Sathanur, Bangalore, Karnataka, India

U.G. Student, Department of Computer Engineering, Reva University, Sathanur, Bangalore, Karnataka, India

Assistant Professor, Department of Computer Engineering, Reva University, Sathanur, Bangalore, Karnataka, India

ABSTRACT: People with hearing impairments can overcome communication barriers by using sign language. Many countries have their own standard and interpretation of sign gesture. Systems that recognize both static and dynamic gestures may be able to convert signs into speech or written English. Two static languages will be used by CNNs: Indian Sign Language (ISL) for advanced CNN architectures and American Sign Language (ASL) for basic recognition language models.

KEYWORDS: convolutional neural networks, deep learning, static gestures, dataset variety, real-time systems, and sign language recognition

I. INTRODUCTION

For the hearing and hearing-impaired communities to communicate, communication is essential. Additionally, this survey describes the latest technology in sign language recognition, such as convolutional neural networks (CNN) and deep learning approaches. obstacles like real-time processing for the model preparation and the diversity of datasets. Additionally, sign language recognition systems enable communication between persons with disabilities and are scalable for real-world applications such as hand gestures, face expressions, and body movements across geographical boundaries.

Accurate and robust sign language recognition (SLR) is crucial for facilitating communication between hearing and deaf communities., Existing methods often struggle with variations in signing styles, environmental factors, and computational constraints., This paper proposes [state your key contribution concisely, e.g., a novel attention mechanism for improved feature extraction, a new hybrid model combining CNNs and RNNs, a data augmentation strategy to handle variations in signing styles]. We evaluate our approach on [mention the dataset] and demonstrate improvements in [mention key metrics, e.g., accuracy, robustness to noise, processing speed] compared to state-of-the-art methods.

While significant progress has been made in sign language recognition (SLR), the accurate recognition of low resource sign language remains a significant challenge., Existing methods often struggle to handle different sign language , leading to suboptimal performance., This paper introduces a novel approach to address this limitation byusing epochs. Our method leverages spatial vectorization and is evaluated on dataset, demonstrating improved performance in terms of accuracy

II. LITERATURE SURVEY

Sign language recognition (SLR) systems are crucial for bridging communication gaps between hearing and deaf communities. However, challenges remain in accurately recognizing signs due to fast movements, individual variations,

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

and noisy data. This necessitates robust and adaptable systems capable of handling these complexities.

SAM-SLR addresses the challenges of fast movements, variations in signing styles, and noisy data acquisition by employing a multimodal approach. This system integrates skeleton, RGB, and depth data, leveraging Sign Language Graph Convolutional Networks (SL-GCN) and Separable Spatial-Temporal Convolution Networks (SSTCN) for feature extraction. The architecture combines these networks with 3D CNNs for RGB and depth processing, and a weighted ensembling technique combines predictions for improved accuracy. The use of pre-trained pose estimators aids in generating skeleton graphs. Experiments on the AUTSL dataset demonstrated high accuracy. Challenges include inconsistent hand keypoints, varying signing speeds, and annotation difficulties. The importance lies in automating sign language interpretation, providing a benchmark for multimodal SLR systems.

NLA-SLR tackles the problem of visually similar signs (VISigns) having different meanings. It integrates semantic information from glosses (sign labels) to improve recognition. The architecture uses a Video-Keypoint Network (VKNet) processing RGB video frames and keypoint heatmaps, along with a head network employing language-aware label smoothing and inter-modality mixup. Label smoothing generates soft labels based on semantic similarity, while feature mixup combines visual and gloss features. Heatmaps represent keypoints, and the system handles variable-length video clips. The system enhances recognition of visually similar signs by incorporating semantic information. Challenges include balancing semantic and visual features, avoiding overfitting, and handling VISigns effectively.

Another approach addresses the challenges of understanding and recognizing diverse sign languages. It utilizes a Convolutional Neural Network (CNN) for feature extraction and classification, along with an HSV algorithm for background suppression and hand gesture isolation. The methodology involves gesture acquisition, image preprocessing, background removal, and resizing. The system demonstrates accuracy in recognizing a set of American Sign Language (ASL) alphabets. Challenges include handling variations in gestures and ensuring data consistency for complex signs. The importance lies in providing an efficient, low-computation solution for real-time gesture recognition.

This system addresses the challenges of variability in gestures, lighting conditions, and environments using a deep learning approach with Convolutional Neural Networks (CNNs) for recognizing static Indian Sign Language (ISL) signs. The architecture includes convolutional, ReLU, max-pooling, and fully-connected layers, processing RGB images. The methodology involves data collection, preprocessing (normalization and resizing), training (with various optimizers), and evaluation (using precision, recall, F1-score, and accuracy). The system achieved high accuracy in recognizing static signs. Challenges include handling gesture variations and adapting to different lighting and environmental conditions. Its high accuracy makes it a significant improvement over other ISL recognition methods.

Another approach tackles the problem of recognizing unseen classes without extensive labeled data, employing Zero-Shot Sign Language Recognition (ZS-SLR). It uses a multimodal deep learning model combining skeleton data, RGB video, and textual descriptions. The architecture involves hand detection using a Transformer-based model, feature extraction from skeletons and RGB frames, feature fusion, semantic mapping using BERT, and classification using fused embeddings. The model merges skeleton-based and deep features for a common semantic space. It outperformed other ZS-SLR methods on various datasets. Challenges include handling domain shift and semantic loss in zero-shot learning. The significance lies in addressing the annotation bottleneck problem for large-scale proficient sign language recognition for all the users

Hand occlusion, where parts of the hands are hidden from view, severely impairs the accuracy of many existing SLR systems., The authors propose a novel method that combines depth information with RGB video data to mitigate the effects of occlusion., The method uses a depth sensor to identify occluded regions and utilizes a convolutional neural network (CNN) to learn features from both the visible and occluded parts of the hands., Experimental results demonstrate that the proposed method significantly improves the accuracy of SLR in the presence of occlusion, outperforming existing approaches
This research explores cross-lingual sign language recognition, aiming to build models capable of recognizing signs across different sign languages without retraining for each language., A novel approach leveraging cross-lingual transfer learning is proposed., The model is trained on a multilingual dataset comprising multiple sign languages and then fine-tuned

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

III. METHODS

A. Methodology:

1. Data Acquisition: Collecting a diverse dataset and high quality of sign images. Where a dataset having various ASL alphabet gestures. As of now, Images are captured using web camera. And the gestures can some widely common words like water, help, etc..
2. Preprocessing Techniques: processing the raw images into standardized format. Enhances the model's ability and reduce the noise data. This might involve Resizing, HSV(Hue Saturated Value) where it eliminates the image using colour extraction algorithm. HSV transformed images were converted into grayscale to reduce computational complexity. Noise data means to clean the images by removing background noise while taking a gesture details like hand, face expressions, etc.
3. A CNN model is used to extract features from the frames and to predict all kind of gestures and uses a multilayer's Feedforward neural network mostly used in image recognition. The CNN consists of some convolution layers like pooling layer, fully connected layer activation function
4. Fully connected Layers it connects every input neuron(node) to every output node. And output layer used a softmax function for mutli-classification of ASL gestures.
5. SoftMax means neural network layer that uses the SoftMax function to convert raw output into probabilities.
6. Training involves optimizing the model's parameters to minimize error on the trained dataset. While testing it evaluates the data where is it working correctly. And will be using Categorical Cross-Entropy.
7. Optimizer is Adam optimizer used for parameter's updates. And dataset is split into training of 80% and remaining of 20% is for testing a model.
8. Epoch: is defined as the total number of iterations of all the training data in one cycle for training the machine learning model. Or in other way It defines the number of times the entire data set has to be worked through the learning algorithm.
9. Early stopping and learning rate scheduling where it is used to optimize training time and accuracy because of this prevents overfitting and helps the model generalize well to unseen data.
10. Each epoch ensures that the model sees every instance in the training data at least once. Random shuffling of the data within each epoch helps to prevent bias introduced by the order of data presentation. Without epochs, the model might not learn effectively from all the data patterns

IV. EVALUATION

This method is essential for verifying whether the model's performance is being evaluated and the results are being compared.
Evaluates the performance of the trained model on a separate test dataset using metrics like accuracy, precision, recall and f1-score.
Accuracy: Model correctly predicts the outcome.
Precision and Recall: are performance metrics used for pattern recognition and classification in machine learning
F1-Score: It Provides a harmonic which mean of precision and recall balances false and negatives. Or in Other way it combines two competing metrics, precision and recall.
Confusion Matrix: Model performance by comparing predicted values against actual values for a dataset. Or Matrix is a visualization method for classifier algorithm results.

V. CHALLENGES

Dynamic gesture recognition: difficulty in capturing and processing the tempered patterns in real-time.
Language Variability: Languages differences in grammar gestures in sign language to recognition of image.

VI. CONCLUSION

In conclusion, the Sign Language recognition using deep learning and other advanced tools like CNN and preprocessing techniques like HSV, recognition model can achieve high accuracy of a model. The challenges like gesture recognitions and real-time processing and overfitting and it involves the data training and improving or development of a model, for communication accessibility in some domains like education, healthcare.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

This research demonstrates the efficacy of deep learning, specifically Convolutional Neural Networks (CNNs), coupled with pre-processing techniques such as HSV color space conversion, for achieving high accuracy in sign language recognition (SLR)., The integration of these advanced tools successfully addresses several key challenges inherent in SLR., While the model demonstrates promising results, achieving real-time performance with minimal latency remains an area for further investigation., , This is especially critical given the inherent variability and complexity of sign language gestures., Future research directions include expanding the dataset to encompass a broader range of signs, individuals, and environmental conditions to further enhance model robustness., Optimizing model architecture and training strategies for faster processing and reduced memory footprint is essential for realizing real-time applications., The successful development and deployment of such an accurate and efficient SLR model holds significant potential for improving communication accessibility across various domains, such as education and healthcare., For instance, real-time translation of sign language in educational settings could enhance learning outcomes for deaf students, and similar systems in healthcare could facilitate more effective communication between patients and medical professionals.

REFERENCES

[1]Jha, R., & Sharma, D. (2022). EPICRAFT: E-commerce for artisans. Proceedings of the International Conference on Digital Business and Technology.

[2] He, Siming. (2019). Research of a Sign LanguageTranslation System Based on Deep Learning.392-396. 10.1109/AIAM48774.2019.00083.

[3] International Conference on Trendz in Information Sciences and Computing (TISC). :30-35, 2012.

[4] Herath, H.C.M. & W.A.L.V.Kumari, &Senevirathne,W.A.P.B & Dissanayake,Maheshi.(2013). IMAGE BASED SIGN LANGUAGE RECOGNITION SYSTEM FOR SINHALA SIGN LANGUAGE

[5]Huang J, Zhou W, Li H, Li W (2015) Sign language recognition using 3D convolutional neural networks. In: IEEE international conference on multimedia and expo (ICME), pp1–6

[6]Huang J, Zhou W, Li H, Li W (2015) Sign language recognition using real-sense. In: IEEE China summit and international conference on signal and information processing (ChinaSIP),pp 166–170

[7]TangA,LuK,WangY,HuangJ,LiH(2015)Areal-timehand posture recognition system using deep neural networks.ACM TransIntellSystTechnol (TIST)6(2):21

[8]Uddin MA, Chowdhury SA (2016) Hand sign language recognition for Bangla alphabet using support vector machine. In: IEEE international conference on innovations in science, engineering and technology (ICISET), pp 1–4

[9].Rao GA, Syamala K, Kishore PVV, Sastry ASCS (2018) Deep convolutional neural networks for sign language recognition. In IEEE conference on signal processing and communication engineering systems (SPACES), pp 194–197

[10].Koller O, Zargaran S, Ney H, Bowden R (2018) Deep sign: enabling robust statistical continuous sign language recognition via hybrid CNN-HMMs. Int J Comput Vis 126(12):1311–1325

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING