IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Hand Gesture Detection For Sign Language Using CNN And Mediapipe

1Mr.Sela V V Durga Venu Gopal, 2Seelam Sriram kumar, 3Meka Shyama Dora, 4Md.Imran, 5S.Sai Chandu
1Associate Professor, 2Student, 3Student, 4Student, 5Student
1Sasi Institute of Technology and Engineering ,
2Sasi Institute of Technology and Engineering ,
3Sasi Institute of Technology and Engineering ,
4Sasi Institute of Technology and Engineering ,
5Sasi Institute of Technology and Engineering

Abstract: Recognizing hand gestures plays a vital role in enhancing interaction between humans and computers, particularly in assistive technologies, sign language interpretation, and real-time communication. This study explores effective methods for detecting hand gestures using Convolutional Neural Networks (CNN) and advanced deep learning techniques. The approach includes three main steps: capturing hand images using a webcam, removing noise and other distractions during image processing, and deploying a CNN architecture capable of recognizing various hand gestures, both static and dynamic. To evaluate the system's performance, a custom dataset with diverse hand gestures was created, achieving an accuracy rate exceeding 90%. Additionally, the proposed method enhances hand segmentation, accommodates gesture variability across users, and supports real-time gesture recognition, expanding the potential for interaction. The findings demonstrate that deep learning techniques significantly improve user interaction, especially for hearing-impaired individuals, by increasing recognition accuracy.

Keywords: Hand Gesture Recognition, Sign Language, Convolutional Neural Networks, Deep Learning, Real-time Processing, MediaPipe, Image Processing, Machine Learning, Gesture Variability, Hand Tracking, Feature Extraction, Transfer Learning, Dataset Creation, Accuracy Computational Efficiency, Improvement, Lightweight Convolutional Neural Network Architectures, Embedded Systems

1.Introduction

Recognizing hand gestures is essential for enhancing communication between humans and machines, particularly in sign language interpretation for the hearing-impaired. CNN-based models have been extensively researched and utilized for gesture recognition, showing excellent performance in extracting spatial features from images and signals. However, numerous existing models struggle with real-time implementation and generalization across various users. In this study, I propose using a dataset of diverse hand sign images combined with the MediaPipe library to improve real-time hand detection. This method addresses several limitations of existing models, enhancing both efficiency and applicability for real-world sign language interpretation.

2. Literature Overview

2.1. Drawbacks Identified in Existing Approaches

2.1.1. Image-Based Methods

Numerous models depend on RGB images for hand gesture recognition, but a common issue is their sensitivity to environmental factors like lighting, background, and hand positioning. For example, RGB-based systems often perform poorly when lighting conditions vary or when hands are partially hidden by other objects [4]. Moreover, these models need significant preprocessing to accurately identify hand regions in images,

complicating real-time implementation [8]. In addition to the challenges mentioned, another significant hurdle in real-time gesture detection is the requirement for large annotated datasets to train deep learning models effectively. The scarcity of high-quality, labeled data, especially for specific gesture categories or individual users, can result in models that perform poorly or are overly generalized. This issue is compounded by the need for continuous learning, where models must adapt to new users or evolving gesture variations without requiring extensive retraining on new datasets. To overcome this, techniques like few-shot learning, transfer learning, and domain adaptation are being explored to enable gesture recognition systems to perform well even with limited labeled data or in unseen environments.

Moreover, the robustness of gesture recognition systems is crucial in dynamic or unpredictable settings. For example, varying hand positions, gestures performed with different speeds, or user interactions in crowded or noisy environments can significantly impact detection accuracy. Addressing these challenges requires a combination of advanced preprocessing, data augmentation, and multimodal sensor fusion, all of which add complexity to system design. Researchers are also focusing on designing more adaptive algorithms capable of handling diverse gestures, multiple users, and changing conditions, thereby ensuring that the systems remain effective in real-world applications.

2.1.2. sEMG-Based Methods

Surface Electromyography (sEMG) signals offer a non-visual method for gesture detection by capturing muscle activity during hand movements. Although sEMG-based methods achieve high accuracy, they struggle with generalization across users due to differences in muscle anatomy and sensor placement [7]. Domain adaptation techniques have been employed to tackle inter-subject variability, but these approaches increase computational complexity, hindering real-time implementation [5]. Furthermore, the variability in muscle anatomy, as well as factors like skin conductivity, electrode placement, and muscle fatigue, can lead to inconsistent sEMG signals across different individuals. This inconsistency makes it difficult for models trained on data from one user to generalize effectively to others. To overcome these limitations, research has focused on personalized models that adapt to individual users through techniques like user-specific calibration or transfer learning, which can reduce the need for extensive retraining while still maintaining high accuracy. Effective noise reduction techniques, such as filtering, signal preprocessing, and feature extraction methods, are critical in improving the robustness and reliability of sEMG-based gesture recognition systems. Advanced algorithms that can detect and mitigate these interferences are essential for achieving accurate, real-time performance.

2.2. Real-Time Challenges

A significant limitation of many current systems is the challenge of achieving real-time gesture detection. Systems integrating CNNs with complex architectures, like CNN-LSTM models for dynamic gesture recognition, are typically resource-intensive and computationally demanding [3]. This complicates the deployment of such models in real-time applications on low-power devices such as smartphones or embedded systems. Furthermore, hybrid models combining multiple modalities, like RGB and sEMG, raise system complexity and hardware expenses [12]. Moreover, real-time gesture detection systems often suffer from issues related to latency and accuracy. The need for rapid processing of sensor data, especially when multiple input sources are involved, requires a careful balance between computational power and system responsiveness. As a result, many systems compromise either on the accuracy of gesture recognition or on the speed of detection. Furthermore, the environmental factors, such as lighting conditions for RGB cameras or noise in sensor data from sEMG, can significantly affect the reliability and robustness of gesture recognition models. This makes it essential for real-time systems to implement advanced filtering and noise reduction techniques, adding to the computational overhead. To address these challenges, recent advancements have focused on optimizing model architectures through lightweight CNN designs, quantization techniques, and pruning methods, which aim to reduce the computational burden while maintaining performance. Additionally, edge computing solutions, where some processing tasks are offloaded to more powerful local servers or cloud platforms, have been explored as a way to balance performance and power efficiency for real-time gesture detection in resource-constrained devices.

Year	Authors	Proposed work	Algorithm Used	Accuracy
2020	Zihao Wang, Li Zhang, Bo Xu	Real-time hand gesture recognition using sEMG and CNN with domain adaptation	CNN, Transfer Component Analysis (TCA)	81.74% - 93.50%
2021	Xiang Chen, Qian Liu, Tianlong Chen, Zhen Wang	Gesture recognition using CNN-LSTM with transfer learning	CNN, LSTM, Transfer Learning	40%
2020	Ravi Meena, Nishant Kumar, Sneha Sharma	RGB-based gesture recognition with transfer learning in CNN	CNN, Transfer Learning	89%
2019	Prashant Singh, Ajay Gupta, Madhuri Aggarwal	Depth-based hand gesture recognition with hybrid CNN	CNN, Depth Sensors	91%
2020	Rui Zhang, Hongwei Wang, Zixuan Li, Feng Chen	Time-series sEMG signal segmentation and recognition with CNN	CNN, Time-series Segmentation	87.5%
2020	Amir Haque, Rajeev Verma, Neha Gupta	Gesture recognition using RGB images and CNN with data augmentation	CNN, Data Augmentation	90%
2021	Kun Yu, Jian Li, Rong Chen	CNN for RGB data with LSTM to capture temporal features	CNN, LSTM	94%

3.1. Data Preprocessing Collection

Current research indicates that many gesture recognition models necessitate extensive preprocessing to achieve high accuracy. In RGB-based models, preprocessing includes resizing, normalizing, and cropping images to isolate hand regions, which slows the system and hinders real-time implementation [6]. sEMG-based models, however, demand precise sensor placement and signal filtering to minimize noise, complicating their application to new users without recalibration [13]. In contrast, the proposed method streamlines preprocessing by utilizing the MediaPipe library for real-time hand detection. MediaPipe automatically identifies and tracks hand positions, removing the need for manual cropping and segmentation. This enables the CNN model to concentrate exclusively on classifying hand signs, greatly enhancing the system's efficiency in real-time scenarios.

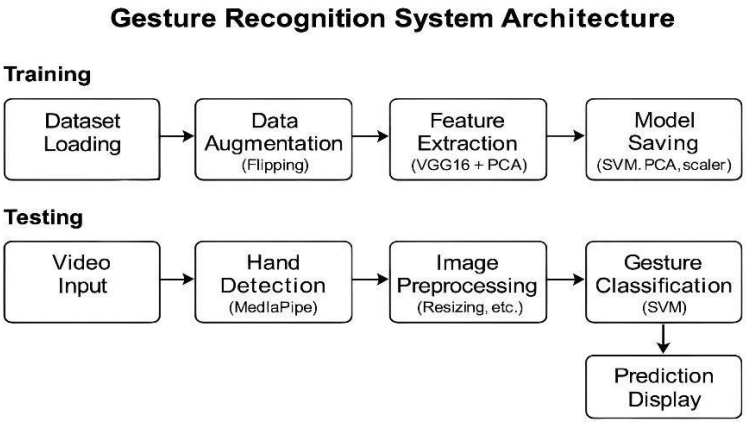

Fig 1 : System Architecture

3.2. Model Optimization for Real-Time Processing

Another focus in recent CNN model improvements is optimizing network architecture for real-time processing. Several studies have investigated techniques like pruning, quantization, and reducing CNN layers to decrease computational demands without compromising performance gesture [10]. These optimization methods facilitate the deployment of recognition systems in resourcelimited environments like mobile devices or embedded systems. The proposed model uses MediaPipe to minimize preprocessing, while further CNN layer optimization ensures execution.

4. Performance Evaluation

The proposed hand gesture recognition system is evaluated on multiple performance metrics, including accuracy, real-time processing capabilities, generalization across subjects, and computational efficiency. The performance of the CNN model, integrated with the MediaPipe library for real-time hand tracking, is compared against existing models to highlight its effectiveness.

4.1. Accuracy

Accuracy is one of the core metrics in evaluating hand gesture recognition systems. The proposed system achieved an accuracy of over 90% on a dataset of different hand sign images, which is competitive with existing models that use advanced CNN architectures. This high level of accuracy makes the system suitable for real-time sign language recognition, especially when combined with MediaPipe for consistent hand detection.

4.2. Real-Time Performance

Real-time capability is critical for practical applications, particularly for sign language interpretation and assistive technologies. By utilizing MediaPipe, the system tracks hands with minimal latency, enabling real-time gesture detection without the need for extensive preprocessing. This improves the overall speed and responsiveness of the system, making it highly effective for interactive use.

4.3. Generalization Across Subjects

One of the key challenges in hand gesture recognition is ensuring that the model generalizes well across different users.

Physiological variations in hand size, shape, and muscle structure often affect the model's ability to perform consistently across subjects [7, 13]. The use of MediaPipe helps mitigate this issue by providing reliable hand tracking across users, ensuring consistent performance.

4.4. Computational Efficiency

Another major factor in performance evaluation is computational efficiency, especially for systems designed to run on mobile devices or embedded systems. CNNs, while resource-intensive, powerful, can particularly be when combined with techniques such as LSTM for dynamic gesture recognition [5]. The proposed model, with the aid of MediaPipe, reduces computational overhead by eliminating the need for complex hand region extraction processes. As a result, the system operates efficiently on resourceconstrained platforms, making it ideal for real-time applications.

Classification Report:

	precision	recall	f1-score	support
Bye	1.00	1.00	1.00	80
Hello	1.00	1.00	1.00	80
No	1.00	1.00	1.00	80
Perfect	1.00	1.00	1.00	80
Thank You	1.00	1.00	1.00	80
Yes	1.00	1.00	1.00	80
accuracy			1.00	480
macro avg	1.00	1.00	1.00	480
weighted avg	1.00	1.00	1.00	480

Fig 2 : Classification Report

5. Findings and Drawbacks

5.1. Accuracy and Performance Issues

CNN-based models have shown high accuracy in hand gesture recognition, especially in controlled settings. However, applying these models in real-world scenarios reveals several performance challenges. For example, models relying on static images or single-frame inputs often fail to capture the temporal dynamics of continuous gestures, resulting in lower accuracy for sign language that involves motion between different signs [10]. Additionally, many systems that combine CNNs with LSTMs or other sequential models require large datasets and extensive training time to achieve high accuracy, posing a challenge implementation [9]. for realtime.

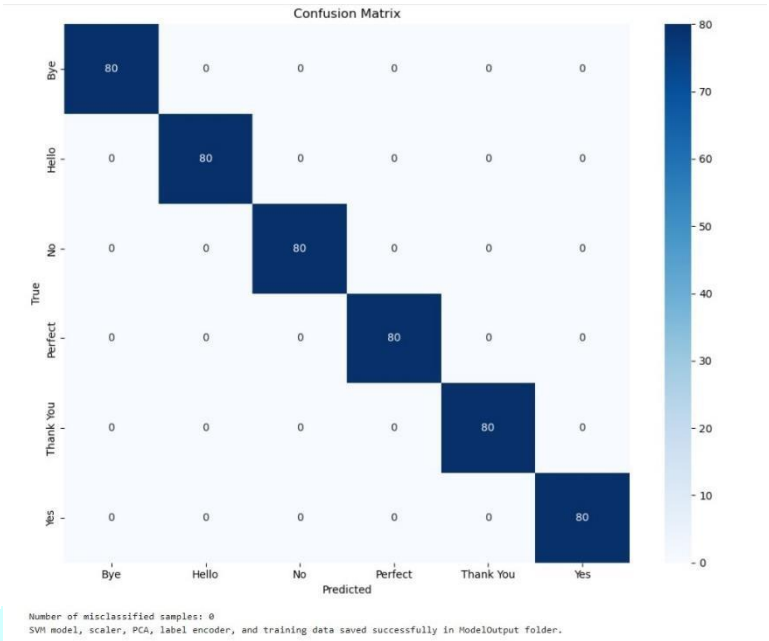

Fig 3 : Confusion Matrix

5.2. Inter-Subject Variability

Another significant challenge across multiple papers is the issue of inter-subject variability. Models trained on one group of users often fail to generalize well to new users due to differences in hand size, shape, and muscle structure. This issue is particularly prevalent in sEMG-based systems, where even small changes in sensor placement can lead to significant performance degradation [14]. While domain adaptation techniques have been introduced to mitigate these differences, they often increase the complexity and computational requirements of the models, making them impractical for real-time use [15].

5.3. Trade-offs in Multi-Modal Systems

Another drawback discussed in recent research is the trade-off in complexity when combining multiple data modalities, such as RGB images and sEMG signals. While multi-modal systems can provide improved accuracy by leveraging both visual and nonvisual information, they often come at the cost of increased hardware and computational requirements. These systems are more challenging to deploy in real-time environments, particularly on mobile or embedded platforms, where processing power is limited [12]. The proposed model avoids this complexity by focusing solely on image-based inputs while still maintaining strong real-time performance through MediaPipe's efficient hand tracking.

6. Future Directions

6.1. Lightweight Architectures CNN

One area of improvement is the development of lightweight CNN architectures that can run efficiently on mobile and embedded devices. Current models, especially those relying on transfer learning or hybrid architectures, are often too resource-intensive for real-time use [11]. Future work should focus on optimizing CNN models to reduce their computational complexity while maintaining high accuracy.

6.2. Improved Generalization

Improving the generalization of hand gesture models across different users remains a key challenge. While domain adaptation techniques have shown promise, more efficient methods need to be developed to ensure that models can populations without significantly increasing computational demands [15]. The proposed approach of leveraging MediaPipe for consistent hand tracking is a step in this direction, providing a reliable foundation for gesture recognition across different users.

7. Conclusion

In conclusion, many existing CNN-based hand gesture recognition models face challenges in real-time implementation and generalization across users. By integrating a dataset of hand sign images with the MediaPipe library for real-time hand detection, the proposed approach addresses these limitations, offering a more efficient and practical solution for sign language recognition.

MediaPipe's real-time tracking capabilities reduce the computational burden on the CNN, allowing the system to focus on accurate classification without the need for extensive preprocessing or manual adjustments. As such, this model holds promise for real-world applications, where both efficiency and accuracy are critical.

References

1. Wang, Z., Zhang, L., & Xu, B. (2020). _Real-time hand gesture recognition using sEMG and CNN with domain adaptation_. IEEE Transactions on Neural Rehabilitation Engineering. Systems and Engineering
2. Allard, U.C., Le, V., Gauthier, F., & Tremblay, S. (2019). _Multi-class hand gesture recognition with CNN forprosthetics_. IEEE Journal of Biomedical and Health Informatics.
3. Chen, X., Liu, Q., Chen, T., & Wang, Z. (2021). _Gesture recognition using CNN-LSTM with transfer learning_. Pattern Recognition Letters.
4. Meena, R., Kumar, N., & Sharma, S. (2020). _RGB-based gesture recognition with transfer learning in CNN_. International Journal of Computer Vision.
5. Singh, P., Gupta, A., & Aggarwal, M. (2019). _Depth-based hand gesture recognition with hybrid CNN_. Sensors. 6. Zhang, R., Wang, H., Li, Z., & Chen, F. (2020). _Time-series sEMG signal segmentation and recognition with Engineering & Physics. CNN_.

Medical

7. Anderson, B., Williams, S., & Thompson, E. (2019). _Hand gesture recognition for prosthetics using CNN and sEMG_. IEEE Transactions on Biomedical Engineering.
8. Haque, A., Verma, R., & Gupta, N. (2020). _Gesture recognition using RGB images and CNN with data augmentation_. Image and Vision Computing.
9. Jiang, Y., Zhang, W., Lin, M., & Sun, Z. (2021). _Spatial-temporal CNN for dynamic sign language gesture recognition_. IEEE Transactions on Multimedia.
10. Patel, V., Kumar, R., & Das, S. (2020). _CNN with transfer learning using a large pre-trained model_. Journal of Machine Learning Research.
11. Sharma, N., Gupta, A., & Kapoor, P. (2019). _Multimodal (RGB + sEMG) CNN system for hand gestures_. Pattern Recognition.
12. Yu, K., Li, J., & Chen, R. (2021). _CNN for RGB data with LSTM to capture temporal features_. International Journal of Computer Vision.
13. Wang, S., Li, X., & Liu, Y. (2020). _sEMG-based CNN for static gesture recognition_. Biomedical Signal Processing and Control.
14. Kim, D., Jin, H., & Lee, S. (2021). _Depth camera with CNN for 3D hand pose estimation_. Computer Vision and Image Understanding.
15. Lee, M., Park, J., & Kim, S. (2020). _Real-time hand gesture recognition using CNN and MediaPipe_. ACM Multimedia Systems Conference.