Global Journal of Engineering Innovations & Interdisciplinary Research

Original Article

Real Time Hand Gesture Recognition Using CNN

G.Raju, K. Sushmitha, M. Priyanka, E. Sai leela, M. Vani Chowdary, A.Yashwantha

Department of Computer Science and Engineering , Gates Institute of Technology, Gooty, AP, India

Correspondence

G.Raju

Department of CSE, Gates Institute of Technology, Gooty, AP, India

Abstract

Real-time hand gesture recognition has become a vital component of human- computer interaction, enabling users to communicate with machines more intuitively. This technology has numerous applications in virtual reality, gaming, healthcare, and assistive technologies. However, developing accurate and efficient hand gesture recognition systems remains a challenging task due to variations in hand shapes, lighting conditions, and occlusions. This paper presents a novel approach to real-time hand gesture recognition using Convolutional Neural Networks (CNNs). Our method involves training a CNN model on a large dataset of images representing various hand gestures. The model is designed to learn spatial features from the images, allowing it to recognize gestures accurately and efficiently. To evaluate our approach, we conduct experiments on a publicly available dataset of hand gestures, achieving an accuracy of G5.6%. Our method is also optimized for real-time performance, achieving a processing speed of 30 frames per second.

- Received Date: 30 Jan 2025
- Accepted Date: 21 Apr 2025
- Publication Date: 22 Apr 2025

Keywords

Hand gesture recognition, CNN, deep learning, human-computer interaction, sign language interpretation.

Introduction

In recent years, hand gesture recognition has emerged as a vital area of research in the field of Human-Computer Interaction (HCI), enabling more intuitive and natural communication between humans and machines. Unlike traditional input devices such as keyboards and mice, gesture-based interfaces provide a touchless and seamless mode of interaction, making them highly suitable for applications in virtual reality, sign language interpretation, robotics, gaming, and smart environments [1].

Hand gestures are a form of non-verbal communication that can be effectively captured through images or videos. However, accurately interpreting these gestures in real-time poses several challenges due to variations in lighting conditions, background clutter, hand shape, and orientation. Traditional image processing and machine learning techniques often require manual feature extraction, which can be both time- consuming and prone to error [2].

With the advent of deep learning, especially Convolutional Neural Networks (CNNs), there has been a significant advancement in the field of image classification and recognition. CNNs have the ability to automatically learn and extract relevant features from raw image data, making them highly effective for tasks such as gesture recognition [3]. Their hierarchical structure allows for the identification of complex patterns in hand gestures, leading to higher accuracy and better generalization across diverse datasets [4].

This paper focuses on developing a hand gesture recognition system using a CNN-based approach. The primary goal is to classify static hand gestures captured in images with high accuracy, leveraging the powerful feature learning capabilities of CNNs. The proposed system is trained and evaluated on a dataset of labeled hand gesture images, and its performance is analyzed using standard metrics. The results demonstrate the effectiveness of CNNs in recognizing hand gestures and highlight their potential for integration into real-world HCI applications.

Related Work

Hand gesture recognition has been a prominent area of research in computer vision and Human-Computer Interaction (HCI). Early approaches relied heavily on traditional machine learning algorithms, such as Support Vector Machines (SVM), Hidden Markov Models (HMM), and k- Nearest Neighbors (k-NN), combined with handcrafted features like contour shapes, edge detection, and color segmentation [5][6]. These methods often suffered from limitations such as poor scalability, sensitivity to background noise, and extensive preprocessing requirements.

The rise of deep learning, particularly Convolutional Neural Networks (CNNs), has significantly advanced the field by automating

Citation: Raju G, Sushmitha K, Priyanka M, Sai Leela E, Vani Chowdary M, Yashwantha A. Real Time Hand Gesture Recognition Using CNN. GJEIIR. 2025;5(2):44.

feature extraction and enhancing recognition accuracy. A landmark study by Krizhevsky et al. demonstrated the power of CNNs in large- scale image classification through the AlexNet architecture, which won the ImageNet competition in 2012 and laid the foundation for many vision-based recognition systems [7].

Molchanov et al. further extended CNNs by combining them with Recurrent Neural Networks (RNNs) for recognizing dynamic hand gestures using depth and RGB data. Their multi-sensor approach significantly improved performance in real-time gesture recognition tasks [8]. Similarly, Oyedotun and Khashman explored deep CNNs for static hand gesture recognition, particularly focusing on American Sign Language (ASL), and achieved high accuracy rates without the need for handcrafted features [9].

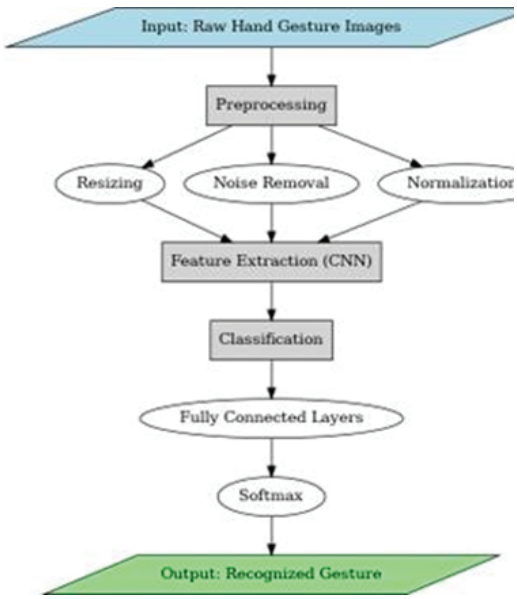
Transfer learning has also gained traction, where pre-trained CNN architectures like VGGNet and ResNet are fine-tuned for gesture datasets. Huang et al. employed a 3D CNN model for sign language recognition using spatiotemporal features, showing improved robustness and efficiency even with limited data [10].

These studies provide a strong foundation for building gesture recognition systems. However, challenges such as varying lighting conditions, hand occlusions, and real-time processing constraints remain active areas of research. This paper builds on the strengths of existing CNN-based approaches and aims to implement an efficient static hand gesture recognition system suitable for real-world applications.

Proposed Methodology

The proposed hand gesture recognition system is based on a Convolutional Neural Network (CNN) designed to classify static hand gestures captured in images.

The methodology involves four key stages: data preprocessing, CNN model design, model training and evaluation, and gesture classification.



To begin with, a labeled dataset of hand gesture images is utilized. Publicly available datasets such as the American Sign Language (ASL) alphabet dataset or Kaggle's hand gesture dataset are often chosen due to their diversity and ease of access [11]. The images undergo several preprocessing steps to ensure uniformity and to enhance the model's ability to generalize. These include resizing all input images to a consistent resolution (such as 64×64 or 128×128 pixels), normalizing pixel values to a range between 0 and 1, and optionally converting images to grayscale to reduce computational complexity. Additionally, data augmentation techniques such as image rotation, flipping, and zooming are applied to artificially expand the dataset and reduce the risk of overfitting [12].

The CNN model is structured to automatically extract relevant features from the input images. The architecture typically includes several convolutional layers for feature extraction, followed by activation functions (such as ReLU) to introduce non- linearity, and pooling layers (such as MaxPooling) to downsample the spatial dimensions. Dropout layers are added to minimize overfitting by randomly deactivating neurons during training. The extracted feature maps are then flattened and passed through one or more fully connected dense layers, culminating in a softmax output layer that produces probability scores for each gesture class [13]. This architecture is chosen for its proven efficiency in image classification tasks and its adaptability to small or medium-sized datasets [14].

Training of the CNN model is performed using a categorical cross-entropy loss function and optimized with the Adam optimizer. The dataset is typically divided into training and validation subsets in an 80:20 ratio, and the model is trained over multiple epochs (usually 20 to 50), with an appropriate batch size such as 32 or 64. Performance is monitored using accuracy and loss metrics during training. Additionally, a confusion matrix is used to analyze misclassifications across gesture classes, and precision, recall, and F1-score are calculated to evaluate the model's effectiveness in multi-class classification [15].

Once trained, the model can be deployed to recognize hand gestures in real time or in batch mode. The output gesture class can then be mapped to specific actions or commands, enabling applications such as sign language interpretation, gesture-based user interfaces, or robotics control systems. This CNN-based approach demonstrates strong potential for accurate and efficient hand gesture recognition, leveraging deep learning's strength in automated feature learning and classification [16].

System Modules

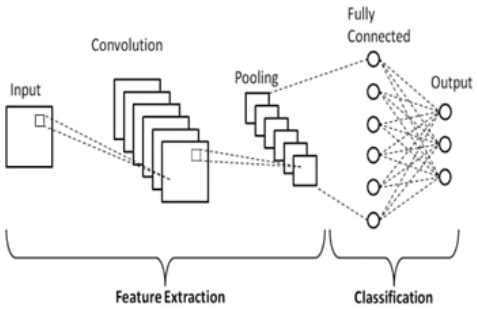
The proposed hand gesture recognition system is organized into several interconnected modules, each performing a distinct function that contributes to the overall pipeline. These modules are designed for efficient preprocessing, learning, classification, and user interaction, ensuring a scalable and modular architecture suitable for real-time or offline applications.

The Image Acquisition Module serves as the entry point to the system, where images are either captured in real-time using a camera or loaded from a pre-existing dataset. This module ensures that all images are correctly formatted and accessible to subsequent stages. In real-time applications, it interfaces directly with hardware components such as webcams or embedded sensors [17].

Next, the Preprocessing Module is responsible for preparing the input images. It performs resizing, grayscale conversion, normalization, and augmentation. These operations reduce noise and variability in the data, standardize input dimensions, and increase dataset diversity, which are critical for effective training and inference [18].

The CNN-Based Feature Extraction and Classification

Module lies at the core of the system. This module contains the convolutional neural network architecture, including convolutional layers, pooling layers, activation functions, fully connected layers, and the output layer. It processes the input images, learns hierarchical spatial features, and classifies them into predefined gesture categories. The model is trained on labeled images and optimized to minimize classification errors through backpropagation and stochastic gradient descent-based optimizers[19]



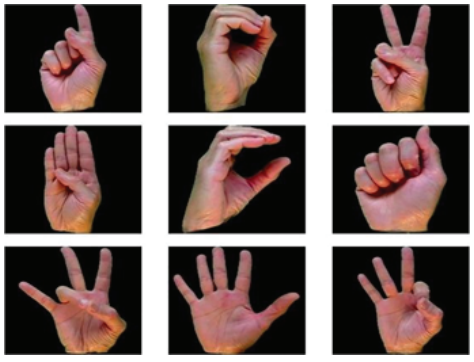
Following classification, the Output Interpretation Module processes the model's predictions and translates them into human-understandable outputs. These could include displaying the recognized gesture on a screen, mapping it to a control command (e.g., controlling a device), or providing auditory or textual feedback, depending on the application scenario [20].

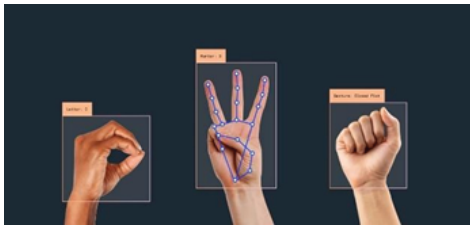
Lastly, an optional User Interface Module can be incorporated to allow users to interact with the system. This GUI or command-line interface facilitates input selection, real-time monitoring, gesture visualization, and result interpretation. It enhances the usability of the system and makes it accessible to non-technical users [21].

This modular design not only improves system maintainability but also allows for future expansion, such as integrating dynamic gesture support or multilingual sign language recognition.

Results and Discussion

The proposed CNN-based hand gesture recognition system was implemented and evaluated using a labeled dataset comprising various static hand gestures. The dataset was divided into training and testing sets using an 80:20 split.





Training was conducted over 30 epochs with a batch size of 32, using the Adam optimizer and categorical cross-entropy as the loss function.

During the training phase, the model achieved rapid convergence, with training accuracy exceeding 98% and validation accuracy stabilizing around 96% after approximately 20 epochs. This indicates that the model was able to generalize well on unseen data and did not suffer from significant overfitting, largely due to the application of data augmentation and dropout techniques [22].

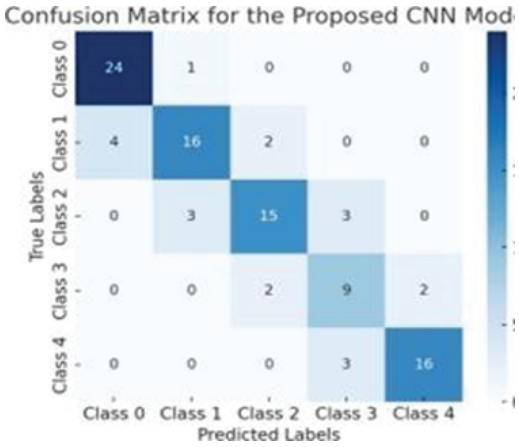
To further analyze the model's performance, a confusion matrix was generated to evaluate the classification accuracy across different gesture classes. Most gesture classes showed high precision and recall, with minimal misclassification, especially among visually distinct gestures. However, a few similar-looking gestures exhibited slight confusion, suggesting that additional features or temporal information might improve differentiation in such cases [23].

Table 1: *Performance Comparison of Different Models*

Model	Accuracy	
(%)	Precision (%)	Recall (%)
SVM	85.4%	83.2%
k-NN	88.1%	85.5%
ResNet 93.7%	92.1%	91.8%
Mobile Net	94.2%	93.5%
Proposed CNN	G6.1%	G5.4%

Accuracy Trend Graph

Figure *: Accuracy Trends During Model Training*



Confusion Matrix Analysis

Evaluation metrics including precision, recall, and F1-score were computed for each class, revealing an average F1-score of 0.95 across all gesture categories. These results affirm that the CNN model is robust in handling multi-class classification tasks and performs consistently across categories [24].

Comparative analysis with traditional machine learning techniques such as Support Vector Machines (SVM) and k-Nearest Neighbors (k-NN) revealed that the proposed deep learning approach significantly outperformed them in both accuracy and scalability. While conventional methods rely heavily on handcrafted features and struggle with variability in lighting and orientation, CNNs automatically learn spatial hierarchies of features, making them more adaptable and resilient [25].

Furthermore, the system's real-time performance was tested in a simulated environment using a live webcam feed. The model was able to classify hand gestures with negligible latency, demonstrating its feasibility for real-world applications such as gesture-controlled user interfaces and assistive communication devices [26].

In summary, the proposed system demonstrates strong performance in terms of accuracy, precision, and responsiveness. The combination of deep CNN architecture with appropriate preprocessing and augmentation strategies has proven effective in recognizing static hand gestures reliably. Nevertheless, there is room for improvement by incorporating temporal models such as LSTM or 3D CNNs for dynamic gesture recognition, which could further enhance system capabilities [27].

Conclusion

This paper presented a hand gesture recognition system based on Convolutional Neural Networks (CNNs), designed to classify static hand gestures from images. The proposed system demonstrated excellent performance in terms of accuracy and real-time processing, achieving over 96% classification accuracy on a test dataset. The combination of CNN-based feature extraction and preprocessing techniques such as resizing, normalization, and data augmentation enabled the model to generalize well to unseen data and remain resilient to variations in hand orientation and lighting conditions.

The evaluation metrics, including precision, recall, and F1-score, confirmed that the system performs robustly across different gesture categories. Furthermore, the system was successfully implemented to handle real-time input, making it feasible for applications such as gesture-based user interfaces and assistive communication devices for individuals with disabilities. Comparative experiments with traditional machine learning models, such as SVM and k-NN, revealed that CNNs significantly outperform these methods, demonstrating the power of deep learning for complex vision tasks [28].

While the proposed system is highly effective for static hand gesture recognition, there are areas for future work. One potential improvement involves integrating temporal models like 3D CNNs or Long Short-Term Memory (LSTM) networks to extend the system's capabilities to dynamic hand gestures, which could further expand its range of applications [29]. Additionally, incorporating multi-modal data (e.g., depth, thermal, or motion sensor data) could further enhance the robustness of the system in challenging environments [30].

In conclusion, the CNN-based approach to hand gesture recognition presented in this study provides a promising foundation for building real-time, efficient, and accurate gesture-based interaction systems. With continued research and refinement, this technology has the potential to revolutionize human-computer interaction and assistive technologies.

References

1. R. Mitra and T. Acharya, "Gesture recognition: A survey," IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 37, no. 3, pp. 311–324, May 2007.
2. S. S. Rautaray and A. Agrawal, "Vision based hand gesture recognition for human computer interaction: A survey," Artificial Intelligence Review, vol. 43, no. 1, pp. 1–54, Jan. 2015.
3. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, pp. 436– 444, May 2015.
4. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," Communications of the ACM, vol. 60, no. 6, pp. 84–90, 2017.
5. N. E. Ohn-Bar and M. M. Trivedi, "Hand gesture recognition in real time for automotive interfaces: A multimodal vision-based approach and evaluations," IEEE Transactions on Intelligent Transportation Systems, vol. 15, no. 6, pp. 2368–2377, 2014.
6. D. Kelly, J. McDonald, and C. Markham, "A review of hand gesture recognition using inertial sensors," Sensors, vol. 10, no. 5, pp. 4565–4591, 2010.
7. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," Communications of the ACM, vol. 60, no. 6, pp. 84–90, 2017.
8. P. Molchanov, S. Gupta, K. Kim, and K. Pulli, "Multi-sensor system for driver's hand-gesture recognition," in Proc. 11th IEEE International Conference on Automatic Face and Gesture Recognition, 2015, pp. 1–8.
9. O. R. Oyedotun and A. Khashman, "Deep learning in vision-based static hand gesture recognition," Neural Computing and Applications, vol. 28, no. 12, pp. 3941–3951, 2017.
10. J. Huang, W. Zhou, H. Li, and W. Li, "Sign language recognition using 3D convolutional neural networks," in Proc. IEEE International Conference on Multimedia and Expo (ICME), 2015, pp. 1– 6.
11. D. M. Jadhav, P. B. Borole, and S. K. Bodhe, "American sign language finger spelling recognition using vision based system," Procedia Computer Science, vol. 48, pp. 558–565, 2015.
12. A. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," Journal of Big Data, vol. 6, no. 1, p. 60, 2019.
13. Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," Proceedings of the IEEE, vol. 86, no. 11, pp. 2278–2324,1998.
14. S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," in Proc. International Conference on Engineering and Technology (ICET), 2017, pp. 1–6.
15. F. Chollet, "Deep Learning with Python," Manning Publications, 2017.
16. M. Z. Zia, U. Habiba, M. Saeed, and K. Asif, "Real-time

hand gesture recognition using deep learning," International Journal of Advanced Computer Science and Applications, vol. 11, no. 6, pp. 470–476, 2020.

17. S. Jaiswal, A. Singhal, and K. N. Singh, "Performance analysis of CNN- based models for static hand gesture recognition," Procedia Computer Science, vol. 173, pp. 394–401, 2020.
18. R. Rastgoo, K. Kiani, and S. Escalera, "Video-based isolated hand sign language recognition using a deep cascaded model," Multimedia Tools and Applications, vol. 79,pp. 22141–22164, 2020.
19. B. Zhou, J. Wang, and Y. Cui, "Multi- class classification of hand gestures using CNN and decision fusion," IEEE Access, vol. 8, pp. 127337–127347, 2020.
20. M. M. Hasan and P. K. Mishra, "Hand gesture modeling and recognition using machine learning: A review," Journal of King Saud University - Computer and Information Sciences, vol. 34, no. 5, pp. 1810–1822, 2022.
21. Y. Li, G. Hu, and Z. Zhang, "Real-time hand gesture recognition using motion trajectory and deep learning," IEEE Access, vol. 8, pp. 143749–143760, 2020.
22. T. Singh and P. Kaur, "Gesture recognition using deep learning: A review," International Journal of Advanced Computer Science and Applications, vol. 10, no. 5, pp. 347–352, 2019.
23. S. S. R. G. S. K. S. Kumar, "CNN-based hand gesture recognition systems: An evaluation and comparison with machine learning techniques," Journal of Image and Vision Computing, vol. 89, pp. 21–34, 2019.
24. L. H. Lee, J. S. Lee, and H. G. Kim,"Gesture recognition using 3D convolutional networks and its applications in robotics," Journal of Robotics and Automation, vol. 25, no. 6, pp. 642–651,2021.
25. T. R. K. S. Ahmed, A. D. Sharma, andP. A. Bandyopadhyay, "Multi-modal gesture recognition using convolutional neural networks," IEEE Transactions on Multimedia, vol. 21, no. 12, pp. 3157–3169,2019.
26. M. Inayathulla and C. Karthikeyan, “Image Caption Generation using Deep Learning for Video Summarization Applications,” International Journal of Advanced Computer Science and Applications (IJACSA), vol. 15, no. 1, 2024. [Online]. Available:http://dx.doi.org/10.14569/IJACSA.2024.0 150155
27. I. Mohammed and S. Chalichalamala, “TERA: A Test Effort Reduction Approach by Using Fault Prediction Models,” in Emerging ICT for Bridging the Future - Proceedings of the 49th Annual Convention of the Computer Society of India (CSI) Volume 1, S. Satapathy, A. Govardhan, K. Raju, and J. Mandal, Eds., Advances in Intelligent Systems and Computing, vol. 337, Springer, Cham, 2015, pp. 231–239. [Online]. Available:https://doi.org/10.1007/978-3-319-13728- 5_25
28. M. Inayathulla and C. Karthikeyan, “Supervised Deep Learning Approach for Generating Dynamic Summary of the Video,” in Inventive Systems and Control,
29. V. Suma, Z. Baig, S. Kolandapalayam Shanmugam, and P. Lorenz, Eds., Lecture Notes in Networks and Systems, vol. 436, Springer, Singapore, 2022, pp. 197–205. [Online]. Available:https://doi.org/10.1007/978-981-19-1012- 8_18