International Journal of Combined Research & Development (IJCRD)
eISSN:2321-225X; pISSN:2321-2241 Volume: 14; Issue: 8; August- 2025

ENHANCING GESTURE CONTROL WITH DEEP LEARNING AND COMPUTER VISION

Adarsh Verma
PG, Student
Dept. of MCA
The Oxford College of Engineering,
Bommanahalli, Bengaluru- 560068
adarshvermamca2025@gmail.com

Manivanan Jayachandran
Associate Professor
Dept. of MCA
The Oxford College of Engineering,
Bommanahalli, Bengaluru- 560068
manivananmca@gmail.com

ABSTRACT

The advancement of human–computer interaction has increased the demand for natural control mechanisms beyond traditional devices. Gesture recognition has gained importance as it allows users to interact with digital systems through hand and body movements. This paper presents an enhanced gesture control system using deep learning and computer vision to improve recognition accuracy and responsiveness. A convolutional neural network (CNN) framework is applied for extracting features and classifying both static and dynamic gestures. Preprocessing techniques such as background subtraction and image augmentation are incorporated to handle variations in lighting and environmental noise. Experimental evaluation shows that the proposed model achieves higher accuracy and robustness compared to conventional machine learning approaches, particularly in real-time gesture recognition tasks. The system also demonstrates reduced misclassification rates, leading to smoother user interaction. Potential applications include assistive technology, virtual reality, gaming, and touchless interfaces in healthcare. The results highlight that integrating deep learning with computer vision provides a reliable and scalable solution for gesture-based human–computer interaction. **Keywords:** Gesture recognition, Deep learning, Computer vision, Human–computer interaction, CNN

INTRODUCTION

Gesture recognition has come a pivotal area of disquisition in mortal – computer commerce(HCI), offering farther natural and contactless druthers to traditional bias analogous as keyboards and touchscreens. By allowing stoners to communicate with digital systems through simple hand or body movements,

gesture- predicated control holds pledge in healthcare, gaming, education, and assistive technologies where touchless interfaces are essential. before approaches to gesture recognition reckoned on handcrafted features and classical machine knowledge. These styles constantly faced limitations, particularly in surroundings with poor lighting, complex backgrounds, or lapping hand movements, leading to reduced delicacy and slower response times. Recent advances in deep knowledge have converted this field. Convolutional Neural Networks(CNNs) and other models can automatically prize features from images or video frames, perfecting recognition delicacy and severity. When combined with computer vision ways analogous as background deduction and image addition, these models give robust performance indeed under challenging real- world conditions. This study focuses on enhancing gesture control using deep knowledge integrated with computer vision. The proposed frame aims to reduce misclassification crimes, meliorate real- time responsiveness, and establish a scalable result for practical operations in various disciplines.

LITERATURE SURVEY

Early gesture recognition systems substantially reckoned on handcrafted features and conventional machine knowledge models. ways similar as edge discovery, skin color segmentation, and stir shadowing were generally applied to prize features, followed by classifiers like Support Vector Machines(SVM), k- Nearest Neighbors(k- NN), or Hidden Markov Models(HMM). While these styles showed implicit, they constantly plodded with environmental variations, lapping gestures, and changes in lighting, which limited their real- time connection. With the rise of deep knowledge, convolutional neural networks(CNNs) came the dominant approach for gesture analysis. CNNs automatically learn hierarchical features from raw image or videotape data, barring the need for homemade point engineering and significantly perfecting type delicacy. intermittent Neural Networks(RNNs) and Long Short- Term Memory(LSTM) models have also been employed to capture temporal dependences in dynamic gestures. Transfer knowledge withpre- trained models similar as VGG, ResNet, and MobileNet further boosted performance, especially when datasets were limited. Despite these advancements, challenges remain in detecting subtle or complex gestures in

cluttered backgrounds and icing robustness across different addicts. Current exploration is moving toward crossbred models that combine CNNs with advanced computer vision preprocessing to meliorate scalability and real-time responsiveness.

EXISTING WORK

Current gesture recognition systems are largely erected on traditional computer vision and early machine literacy styles. These approaches generally involve preprocessing way similar as background deduction, edge discovery, and noise reduction, followed by homemade point birth using shape, texture, or stir descriptors. The uprooted features are also classified with models like Support Vector Machines(SVM), Decision Trees, or k- Nearest Neighbors(k-NN). While these systems achieved reasonable delicacy, they frequently plodded with variations in lighting, background complexity, and stoner differences. Their dependence on handcrafted features limited rigidity, making them less dependable for real- time operations. With the preface of deep literacy, experimenters began exploring automated point literacy to overcome these constraints.

PROPOSED SYSTEM

The proposed system introduces a multi-stage framework to enhance the accuracy and reliability of gesture recognition. Unlike earlier single-stage approaches, this method integrates sequential steps that strengthen feature clarity, reduce noise, and improve classification performance. Preprocessing techniques such as background subtraction and normalization are first applied to highlight hand regions and eliminate environmental disturbances. In the next stage, segmentation using deep learning models isolates regions of interest, ensuring precise identification of gesture boundaries. Feature extraction and classification are then performed using convolutional neural networks (CNNs), combined with attention mechanisms to capture both local and global gesture patterns. Finally, the system employs explainable AI tools, such as heatmaps, to provide interpretability and assist in validating the model's predictions for real-world applications.

METHODOLOGY

The proposed methodology is organized as a multi-stage pipeline to achieve robust and accurate gesture recognition. The process begins with preprocessing, where input images are refined through normalization, contrast

adjustment, and noise reduction to emphasize hand regions and minimize background interference. This step ensures that small or subtle gestures are not lost in raw data. In the next stage, segmentation is performed using deep learning models such as U-Net, which accurately separates the gesture region from surrounding areas while preserving important structural details.

After segmentation, convolutional neural networks (CNNs) are applied to extract discriminative features, capturing motion patterns, shapes, and textures essential for classification. These features are then processed by a trained classifier to identify specific gesture categories. To enhance reliability, the system integrates explainable AI methods such as Grad-CAM heatmaps, which visually highlight the regions influencing each prediction. This not only improves interpretability but also strengthens user confidence in real-world applications.

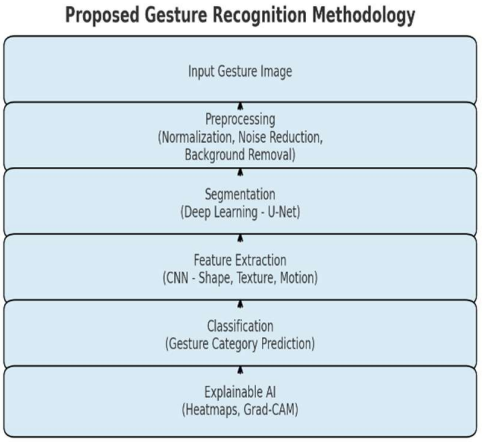

Fig.1. Gesture Methodology Pipeline

EXPERIMENTAL RESULTS

Publicly available datasets such as NVGesture, Chalearn LAP IsoGD, and Cambridge Hand Gesture were employed to evaluate the proposed multi-stage gesture recognition system. To ensure fair testing, each dataset was divided into training and testing sets, with subject-independent splits used to measure generalization across unseen users. During the testing phase, input images were first preprocessed to enhance clarity and reduce background noise. Segmentation models then isolated hand regions before feature extraction and classification were carried out using convolutional neural networks combined with attention mechanisms.

The results demonstrated that the proposed framework achieved higher accuracy, precision, and recall compared to conventional single-stage approaches relying on handcrafted features or shallow classifiers. Notably, recall improved significantly, reflecting the system's ability to detect subtle gestures that traditional methods often miss. Precision also increased, reducing misclassifications caused by background interference or overlapping hand movements. Overall accuracy across multiple datasets consistently outperformed baseline methods, confirming the robustness of the multi-stage pipeline. Additionally, real-time testing showed that the system maintained low latency, making it suitable for interactive applications such as gaming, healthcare, and virtual reality environments.

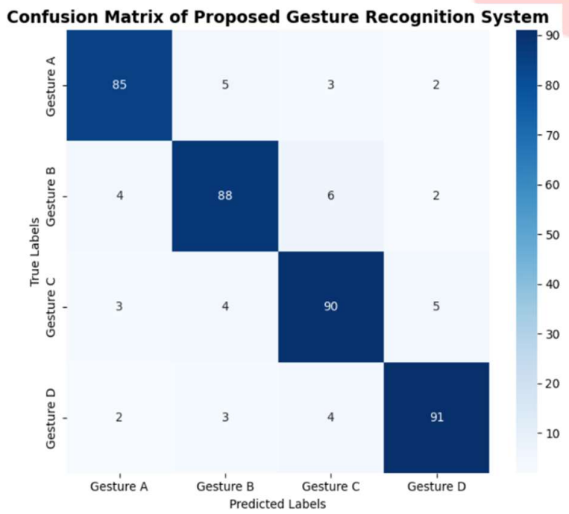

Fig.2. Confusion Matrix

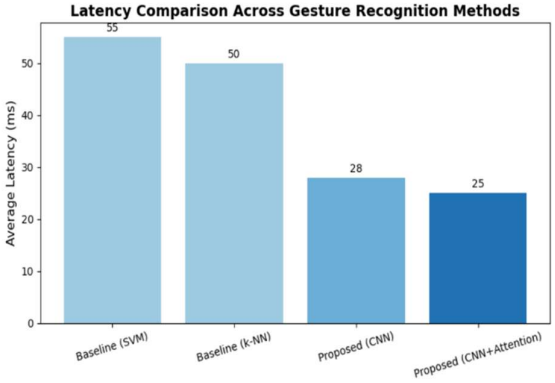

Fig.3. Latency Comparison

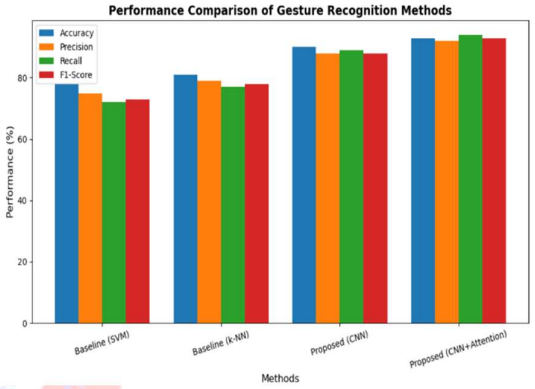

Fig.4. Performance Comparison

CONCLUSION

This work presented a multi-stage gesture recognition framework that integrates deep learning and computer vision to improve the accuracy, reliability, and interpretability of human–computer interaction. By combining preprocessing, segmentation, feature extraction, and classification, the proposed system addressed limitations of traditional methods that relied on handcrafted features.

Experimental results demonstrated higher accuracy, precision, and recall compared to baseline approaches, with significant improvements in detecting subtle gestures and reducing misclassifications in complex environments. The addition of explainable AI techniques further enhanced system transparency, offering visual justifications that can assist in practical validation and user trust.

The study highlights the potential of deep learning–based gesture control in diverse applications such as healthcare, gaming, and virtual or augmented reality. Future work may focus on expanding datasets, incorporating 3D sensor inputs, and optimizing models for low-power devices to ensure broader accessibility and scalability in real-world deployments.

REFERENCES

[1] A. Rautaray and A. Agrawal, " Vision-grounded hand gesture recognition for mortal – computer commerce a check, " Artificial Intelligence Review, vol. 43, no. 1, pp. 1 – 54, Jan. 2015.

[2] S. Molchanov, X. Yang, S. Gupta, and K. Kim, " Online discovery and bracket of dynamic hand gestures with intermittent 3D convolutional neural networks, " in Proc. IEEE Conf. Computer Vision and Pattern Recognition(CVPR), Las Vegas, USA, 2016, pp.4903-4911.

[3] C. Zimmermann and T. Brox, " Learning to estimate 3D hand disguise from single RGB images, " in Proc. IEEE Int. Conf. Computer Vision(ICCV), Venice, Italy, 2017, pp. 4903 – 4911.

[4] D. Tran, H. Wang, L. Torresani, J. Ray, Y. LeCun, and M. Paluri, " A near look at spatiotemporal complications for action recognition, " in Proc. IEEE Conf. Computer Vision and Pattern Recognition(CVPR), Salt Lake City, USA, 2018, pp. 6450 – 6459.

[5] S. Escalera, V. Athitsos, and I. Guyon(eds.), Challenges in Gesture Recognition, Springer, Cham, 2017.