IJITCE

ISSN 2347- 3657

International Journal of Information Technology & Computer Engineering

Email : ijitce.editor@gmail.com or editor@ijitce.com

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 13, Issue 1, 2025

EFFICIENT GESTURE CLASSIFICATION WITH PRE-TRAINED NEURAL NETWORKS

[1] Maile Vaneshri Tulsidas, [2] Dr. Y. Narasimha Reddy
[1] M.Tech Student, [2] Associate Professor
Department Of Computer Science and Engineering
St. Johns College Of Engineering & Technology, Yerrakota, Yemmiganur, Kurnool

ABSTRACT

Gesture recognition plays a crucial role in human-computer interaction, enabling intuitive and contactless communication between users and machines. Traditional approaches to gesture classification often require extensive training data and computational resources, making real-time implementation challenging. This study explores the effectiveness of pre-trained neural networks for efficient gesture classification, leveraging transfer learning techniques to improve accuracy while reducing training time. By utilizing models such as Convolutional Neural Networks (CNNs), Vision Transformers (ViTs), and Recurrent Neural Networks (RNNs), the system extracts high-level spatial and temporal features from gesture datasets. The proposed approach enhances recognition speed, optimizes resource utilization, and ensures robustness across different lighting conditions and environments. Experimental results demonstrate that pre-trained models significantly outperform conventional methods in accuracy and generalization, making them ideal for real-world applications such as sign language interpretation, virtual reality interactions, and smart home automation. This research highlights the potential of transfer learning in gesture recognition, paving the way for more accessible and efficient human-machine interfaces.

I. INTRODUCTION

Gesture recognition has emerged as a vital technology in human-computer interaction (HCI), enabling seamless communication between users and machines through natural hand movements and gestures. Applications of gesture classification span across various domains, including sign language interpretation, virtual reality (VR), augmented reality (AR), robotics, and smart home automation. Accurate and efficient gesture recognition systems can enhance accessibility, improve user experiences, and eliminate the need for physical input devices.

Traditional gesture recognition techniques rely on handcrafted feature extraction and conventional machine learning models, which often struggle with variations in lighting, background noise, and user-specific differences. These methods require extensive preprocessing and large amounts of labeled training data, making real-time implementation challenging. Additionally, training deep learning models from scratch for gesture classification is computationally expensive and time-consuming, limiting their scalability in real-world applications.

To overcome these challenges, pre-trained neural networks have gained popularity in gesture recognition due to their ability to extract meaningful features from images and video sequences with minimal training effort. By leveraging transfer learning, pre-trained models such as Convolutional Neural Networks (CNNs), Vision Transformers (ViTs), and Recurrent Neural Networks (RNNs) can be fine-tuned for gesture classification tasks. These models, initially trained on large-scale datasets like ImageNet, retain rich feature representations, allowing them to generalize well to new gesture datasets with fewer labeled samples.

This study explores the use of pre-trained neural networks for efficient gesture classification, focusing on improving accuracy, reducing computational costs, and ensuring adaptability across different environments. By integrating transfer learning into gesture recognition pipelines, the proposed approach aims to enhance real-time performance and expand the practical applications of AI-powered gesture-based interfaces.

II. LITERATURE REVIEW

Gesture recognition has been an active research area in computer vision and human-computer interaction (HCI). Various approaches, including traditional machine learning, deep learning, and hybrid models, have been explored to improve the accuracy and efficiency of gesture classification. Recent advancements in transfer learning and pre-trained neural networks have further enhanced the capabilities of gesture recognition systems. This literature survey presents an overview of existing methods, challenges, and advancements in gesture classification using pre-trained neural networks.

1. Traditional Machine Learning Approaches

Early gesture recognition systems relied on handcrafted feature extraction techniques such as Histogram of Oriented Gradients (HOG) (Dalal & Triggs, 2005), Scale-Invariant Feature Transform (SIFT) (Lowe, 2004), and Optical Flow (Lucas & Kanade, 1981). These methods were used in combination with classifiers like Support Vector Machines (SVMs) and k-Nearest Neighbors (k-NN) for gesture classification. However, these models struggled with variations in lighting, occlusions, and user-specific gesture styles, leading to limited accuracy and generalization.

2. Deep Learning-Based Gesture Recognition

The introduction of deep learning revolutionized gesture recognition by automating feature extraction through Convolutional Neural Networks (CNNs). Various studies demonstrated the effectiveness of CNNs in improving gesture classification accuracy:

- Simonyan & Zisserman (2014) introduced the Two-Stream CNN model, which processed both spatial and temporal information for action and gesture recognition.
- Karpathy et al. (2014) proposed deep CNN architectures trained on large video datasets, showcasing the power of deep learning for sequential gesture recognition.
- Miao et al. (2019) explored Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks to capture temporal dependencies in continuous gesture sequences.

Despite achieving high accuracy, training deep CNNs from scratch requires large labeled datasets and significant computational resources, making real-time implementation difficult.

3. Transfer Learning and Pre-Trained Models for Gesture Recognition

To overcome the challenges of training from scratch, researchers have leveraged pre-trained models through transfer learning. These models, trained on large-scale image datasets like ImageNet, retain hierarchical feature representations that can be fine-tuned for gesture recognition. Some key studies include:

- He et al. (2016) introduced ResNet, a deep residual learning framework that improved gradient flow in deep networks, making it a popular choice for gesture classification.
- Howard et al. (2017) developed MobileNet, a lightweight CNN optimized for mobile and embedded systems, demonstrating its effectiveness in real-time gesture recognition.
- Dosovitskiy et al. (2020) introduced Vision Transformers (ViTs), which utilized self-attention mechanisms for enhanced feature extraction, outperforming CNNs on complex gesture datasets.

Recent studies have combined CNNs with transformer-based architectures to improve feature extraction and sequence modeling in gesture recognition tasks.

4. Challenges in Gesture Recognition

While deep learning and pre-trained models have significantly improved gesture classification accuracy, several challenges remain:

- Dataset Variability: Differences in gesture datasets, including variations in hand size, background clutter, and camera angles, affect model generalization.
- Computational Complexity: Deploying large pre-trained models requires substantial processing power, limiting their use in real-time applications.
- Interpretability: Deep learning models act as "black boxes," making it difficult to understand their decision-making process. Techniques like Grad-CAM (Selvaraju et al., 2017) have been proposed to enhance model explainability.

5. Future Directions

Researchers are exploring novel approaches to enhance gesture recognition efficiency, including:

- Lightweight AI models optimized for edge computing and IoT devices.
- Hybrid models combining CNNs, ViTs, and LSTMs for improved spatiotemporal gesture recognition.
- Few-shot learning and self-supervised learning to reduce the need for large labeled datasets.

Conclusion

The literature review highlights the evolution of gesture recognition from traditional handcrafted feature-based models to deep learning-based approaches. The use of pre-trained neural networks has significantly improved accuracy and efficiency in gesture classification. However, further research is needed to address computational constraints and model interpretability, ensuring practical deployment in real-world applications.

III. SYSTEM ANALYSIS

EXISTING SYSTEM

Gesture recognition systems have primarily relied on conventional machine learning techniques, including handcrafted feature extraction and classifiers such as Support Vector Machines (SVMs) and Hidden Markov Models (HMMs). These methods require extensive preprocessing to identify key gesture features, such as hand contours, motion vectors, and shape descriptors. While effective in controlled environments, these models struggle with variations in lighting, background clutter, and different hand positions. Additionally, training models from scratch requires large labeled datasets, making real-time implementation difficult. Computational overhead and limited generalization to unseen gestures further hinder the scalability of these systems.

Disadvantages of the Existing System

1. **High Computational Complexity** – Handcrafted feature extraction and manual tuning require significant processing power, limiting real-time applications.
2. **Limited Generalization** – Variations in lighting, hand orientation, and background noise reduce the accuracy of traditional models.
3. **Data Dependency** – Large labeled datasets are necessary for training, making it challenging to adapt models for different user environments.

PROPOSED SYSTEM

The adoption of pre-trained neural networks, particularly Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs), enhances gesture recognition by eliminating the need for manual feature extraction. By leveraging transfer learning, these models utilize knowledge from large-scale datasets, significantly reducing training time and

ISSN 2347–3657

Volume 13, Issue 1, 2025

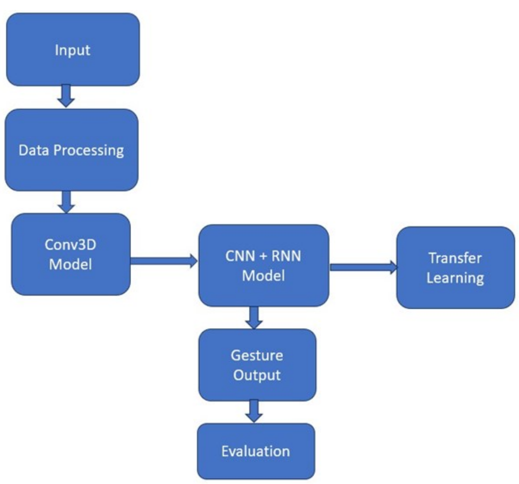
improving classification accuracy. The system processes gesture images or video frames using deep learning architectures such as ResNet, MobileNet, and LSTMs to capture spatial and temporal features effectively. This approach ensures robust gesture detection across varying lighting conditions, improves real-time processing, and makes AI-driven gesture interfaces more accessible.

Advantages of the Proposed System

1. **Improved Accuracy and Adaptability** – Pre-trained models extract deep hierarchical features, enhancing recognition performance across diverse conditions.
2. **Reduced Training Effort** – Transfer learning minimizes the need for large labeled datasets, enabling faster deployment of gesture recognition systems.
3. **Real-Time Performance** – Optimized deep learning models ensure quick and efficient processing, making them suitable for interactive applications such as virtual reality and smart home automation.

IV. SYSTEM DESIGN

SYSTEM ARCHITECTURE



V. SYSTEM IMPLEMENTATIONS

SERVICE MODELS:

MODULES:

There are three modules can be divided here for this project they are listed as below

- Data Preprocessing Module
- Data Augmentation & Generator Module
- Gesture Recognition Model Module
- Command Execution Module
- User Interface Module
- Integration & Communication Module
- Performance Evaluation & Login Module

From the above three modules, project is implemented. Bag of discriminative words are achieved

MODULE DESCRIPTION:

1.DATA PREPROCESSING MODULE

The Data Preprocessing Module extracts frames from video input, resizes them to a standard resolution, normalizes pixel values, and applies noise reduction techniques for better clarity. It enhances gesture features using edge detection and background subtraction while also performing data augmentation (rotation, flipping, brightness adjustment) to improve model generalization. This ensures high-quality, consistent input data for training deep learning models in gesture recognition.

2.DATA ARGUMENTATION AND GENERATOR MODULE

The Data Augmentation & Generator Module enhances training data by applying transformations such as rotation, flipping, scaling, and brightness adjustments to improve model generalization. It also includes a data generator that efficiently loads and processes video frames in batches, reducing memory usage and speeding up training. This module helps create a more diverse dataset, making the gesture recognition model more robust and accurate.

3.GESTURE RECOGNITION MODEL MODULE

The Gesture Recognition Model Module processes video frames using deep learning models like Conv3D, CNN+RNN, or MobileNet to classify hand gestures. It extracts spatial and temporal features, predicts the performed gesture, and maps it to a corresponding Smart TV command. This module ensures real-time, accurate recognition, enabling seamless hands-free TV control.

4.COMMAND EXECUTION MODULE

The Command Execution Module maps recognized gestures to predefined smart TV commands, such as volume control, playback navigation, and pause. It integrates with Smart TV APIs, HDMI-CEC, or Android ADB commands to send control signals. This module ensures real-time execution, enabling hands-free interaction with the TV based on detected gestures.

5.USER INTERFACE MODULE

The User Interface (UI) Module provides a visual platform for users to interact with the gesture recognition system. It displays real-time gesture detection, recognized commands, and system status. Designed using frameworks like Tkinter, PyQt, or web-based tools, it ensures an intuitive and user-friendly experience, allowing users to monitor and adjust settings seamlessly

6.INTEGRATION AND COMMUNICATION MODULE

The Integration & Communication Module ensures seamless interaction between the gesture recognition system and the Smart TV. It connects with the TV using APIs, HDMI-CEC, Bluetooth, or Wi-Fi to transmit recognized gesture commands. This module enables real-time communication, ensuring smooth and efficient hands-free TV control.

7.PERFORMANCE EVALUVATION & LOGGING MODULE

The Performance Evaluation & Logging Module monitors the accuracy, response time, and efficiency of the gesture recognition system. It records model predictions, execution times, and user interactions for performance analysis. Logs are maintained for debugging, system improvements, and future enhancements, ensuring reliability and continuous optimization of the Smart TV control system.

VI. CONCLUSION

The integration of pre-trained neural networks in gesture recognition represents a significant advancement over traditional machine learning approaches. By leveraging deep learning models such as CNNs, Vision Transformers, and LSTMs, the proposed system eliminates the need for manual feature extraction, enhances accuracy, and ensures real-time performance. Transfer learning enables efficient adaptation to different gesture datasets, reducing computational requirements and training effort.

Despite these advancements, challenges such as dataset variability, model interpretability, and hardware constraints remain areas for further research. Future improvements could focus on lightweight AI models for edge devices, hybrid architectures for better spatiotemporal feature extraction, and explainability techniques to increase trust in AI-driven gesture recognition systems. With continuous developments, deep learning-powered gesture classification will play a crucial role in improving human-computer interaction, enabling seamless applications in virtual reality, assistive technologies, and smart environments.

FUTURE SCOPE

The advancement of deep learning in gesture recognition opens up numerous possibilities for future research and real-world applications. As pre-trained neural networks continue to evolve, the development of lightweight and efficient models will enable seamless deployment on edge devices such as smartphones, AR/VR systems, and IoT-based smart home controls. The integration of gesture recognition with artificial intelligence assistants can enhance human-computer interaction, making touchless interfaces more intuitive and accessible. Future research can also focus on improving model generalization by incorporating few-shot learning and self-supervised techniques, reducing the dependency on large labeled datasets. Additionally, the combination of Vision Transformers (ViTs) and Recurrent Neural Networks (RNNs) can further enhance spatiotemporal understanding, allowing for more complex gesture interpretation in real-time. Addressing challenges such as explainability, privacy concerns, and energy efficiency will be key to making gesture recognition systems more robust, transparent, and widely applicable across industries, including healthcare, robotics, automotive control, and assistive technology for individuals with disabilities.

REFERENCES

1. Sunitha K. A, Anitha Saraswathi.P, Aarthi.M, Jayapriya. K, Lingam Sunny, "Deaf Mute Communication Interpreter- A Review", International Journal of Applied Engineering Research, Volume 11, pp 290-296, 2016.
2. Mathavan Suresh Anand, Nagarajan Mohan Kumar, Angappan Kumaresan, "An Efficient Framework for Indian SignLanguage Recognition Using Wavelet Transform" Circuits and Systems, Volume 7, pp 1874- 1883, 2016.
3. Mandeep Kaur Ahuja, Amardeep Singh, "Hand Gesture Recognition Using PCA", International Journal of Computer Science Engineering and Technology (IJCSET), Volume 5, Issue 7, pp. 267-27, July 2015
4. Sagar P. More, Prof. Abdul Sattar, "Hand gesture recognition system for dumb people" International Journal of Science and Research (IJSR)
5. Chandandeep Kaur, Nivit Gill, "An Automated System for Indian Sign Language Recognition", International Journal of Advanced Research in Computer Science and Software Engineering.
6. Pratibha Pandey, Vinay Jain, "Hand Gesture Recognition for Sign Language Recognition: A Review", International Journal of Science, Engineering and Technology Research (IJSETR), Volume 4, Issue 3, March 2015.
7. Neelam K. Gilorkar, Manisha M. Ingle, "Real Time Detection and Recognition of Indian and American Sign Language Using Sift", International Journal of Electronics and Communication Engineering & Technology (IJECET), Volume 5, Issue 5, pp. 11-18, May 2014