REAL-TIME HAND GESTURE RECOGNITION FOR HUMAN-COMPUTER INTERACTION

1. INTRODUCTION

Gesture recognition technology has been made possible by the growth of touchless human-computer interface (HCI) technologies, which allows for more natural and intuitive device control. This research investigates the creation of a real-time hand gesture detection system using deep learning methods, particularly Long Short-Term Memory (LSTM) networks and Convolutional Neural Networks (CNNs). Real-time, accurate, and efficient hand gesture recognition is the aim for applications in assistive technology, smart homes, and virtual reality (VR).

Users can engage with computers and gadgets naturally and hands-free thanks to gesture detection. The goal of this research is to develop a real-time system that can identify dynamic hand movements with high accuracy and low latency, even on devices with limited resources.

2. OBJECTIVE

Designing and implementing a reliable gesture detection system that can process video data in real-time and accurately predict hand motions is the main goal of this project. The following are the main issues addressed:

- Fast inference speeds appropriate for touchless interaction are achieved using real-time processing.
- using gesture classification to precisely identify dynamic hand gestures in a live video feed.
- Resource optimization is necessary to make sure the model works properly on gadgets with constrained processing power, such as embedded systems or smartphones.

3. METHODOLOGY

3.1 Data Collection and Preprocessing

The Jester Dataset from Kaggle, which includes 148,092 tagged video clips depicting 27 distinct hand gestures (such as swiping and zooming), was used for this study. Each move in these 30-frame-per-second videos is assigned a category from a list of 27. Every video is broken up into separate 224x224 pixel frames before being entered into the model.

The steps involved in preprocessing were:

- Frame extraction: To make frame-level analysis easier, videos were divided into individual frames.
- Normalization: To enhance model performance, pixel values were normalized to fall between 0 and 1.
- One-hot encoding: To match the output dimensions for categorization, gesture labels were one-hot encoded.

3.2 Model Architecture

The hybrid architecture of the model combines LSTMs and CNNs:

- CNN (MobileNetV2): Each frame's spatial information are extracted using the CNN. Because of its lightweight design, which guarantees quicker processing times while preserving good accuracy, MobileNetV2 was chosen. A global average pooling layer is used to minimize spatial dimensions after a number of convolutional and pooling layers.

- LSTM (Long Short-Term Memory): To record temporal dependencies between frames, spatial data that have been retrieved by the CNN are sent to an LSTM network. Accurate gesture recognition depends on the model's ability to comprehend the hand gesture sequence over time, which is aided by the LSTM layer.

3.3 Training and Hyperparameter Tuning

Ten percent of the dataset was used for testing, ten percent for validation, and eighty percent for training. The Adam optimizer with a categorical cross-entropy loss function and a learning rate of 0.001 was used to train the model. To avoid overfitting, early stopping was applied after 10 epochs of no improvement in validation loss, and a batch size of 32 was utilized for training.

Grid search was used to fine-tune hyperparameters such batch size, learning rate, and the number of LSTM units.

3.4 Evaluation Metrics

The model was assessed using several important metrics:

- Accuracy: Calculates the percentage of accurate forecasts.
- To assess how well the model performed across various gesture types, the following metrics were computed for each class: precision, recall, and F1-score.
- Inference Speed: A objective of less than 33 ms per frame at 30 FPS was used to assess the system's real-time prediction performance.

4. RESULTS

The model showed encouraging outcomes with minimal inference time and good accuracy:

- Accuracy: On the test set, the model's accuracy exceeded 90%.
- Inference Speed: The system met the real-time criterion for gesture recognition with an average inference time of about 30 ms per frame.
- Precision and Recall: Individual class performance was balanced across the various gestures in terms of precision, recall, and F1-scores.

Misclassifications between comparable motions were among the areas for improvement that were highlighted by the confusion matrix. On the other hand, the model's performance was ideal for real-time applications in intelligent settings.

5. USE CASES AND APPLICATIONS

There are several real-world uses for the real-time gesture recognition system created in this project:

- Hand gestures can be used to interact with virtual reality (VR) environments, allowing for intuitive and immersive control of virtual items.
- Smart Homes: The technology can be incorporated into smart home systems, enabling users to utilize gestures to control thermostats, lighting, and other appliances.
- Assistive Technology: Gesture recognition makes it easier for people with disabilities to use digital devices by providing a touchless engagement approach.

6. CHALLENGES AND FUTURE WORK

Challenges:

- **Real-Time Processing**: It's still difficult to get real-time performance on devices with limited resources. The model's mobile device optimization will be the main focus of future research.
- **Environmental Factors:** The accuracy of gesture recognition may be impacted by changes in illumination, background noise, and occlusions (such as partially visible hands).
- **Gesture Variability:** Additional data augmentation may be necessary to solve the model's potential inability to distinguish between similar motions.

Future Work:

- **Data Augmentation:** The model's capacity to identify gestures in a range of settings can be enhanced by diversifying the training data.

- **Multi-Hand Gesture Recognition**: One important avenue for further research is to extend the model to accommodate many hands in the picture.

Mobile Optimization: The model will be made smaller and faster to infer while retaining accuracy in order to be deployed on mobile devices.

7. CONCLUSION

To sum up, our study effectively used a hybrid CNN-LSTM architecture to create a real-time hand gesture recognition system. High accuracy and low inference times were attained by the model, which also showed useful applications in assistive technologies, virtual reality, and smart settings. The results indicate considerable promise for future research and application, even though there are still difficulties in optimizing the system for real-time performance in various situations.