ASL Recognition Using Real-Time Hand Tracking and Image Classification

Introduction

American Sign Language (ASL) is a vital communication medium for individuals with hearing or speech impairments. This project explores the application of real-time computer vision and deep learning to interpret ASL fingerspelling (A–Z), using a standard webcam and lightweight models to promote inclusivity and accessibility in communication technology.

Abstract

This project presents the development of a real-time ASL recognition system that identifies English alphabet signs through a webcam. Using MediaPipe for efficient hand tracking and a TensorFlow Lite classifier trained on a custom dataset of cropped hand gestures, the system processes and predicts hand signs instantly. The classifier outputs a corresponding letter label with high accuracy, enabling real-time translation of static hand gestures into digital text.

Tools Used

Python – Core scripting language

OpenCV – For webcam integration and image handling

cvzone & MediaPipe – For real-time hand detection

TensorFlow & TFLite – Model training and lightweight inference

NumPy & Math – Array manipulation and geometric scaling

Steps Involved in Building the Project

1. **Dataset Creation**
    - Captured hand sign images for A–Z using webcam and MediaPipe.
    - Cropped and resized all hand regions to 300×300 pixels on a white canvas.
2. **Model Training**

- Trained a CNN model on the preprocessed dataset to classify 26 ASL letters.
- Exported the model to .tflite format for optimized real-time performance.

3. **Real-Time Inference Setup**
   - Detected the hand using cvzone.HandDetector in each video frame.
   - Cropped the hand area, resized it, and passed it into the TFLite model.
4. **Prediction Display**
   - Retrieved the predicted class and confidence score.
   - Displayed the result with bounding boxes and label overlays on the webcam feed.
5. **Testing & Refinement**
   - Tested across lighting conditions and hand angles.
   - Adjusted image preprocessing and added boundary checks for stability.

Conclusion

This project successfully demonstrates real-time ASL alphabet recognition using efficient hand detection and lightweight deep learning models. It bridges a key accessibility gap by converting visual gestures into digital text and serves as a foundation for more advanced features like full word recognition, text-to-speech, or multi-hand signing. With further enhancements, this system can evolve into a powerful assistive tool in educational and communication technologies.