Real-Time Sign Detection

Author: Sanjana Madhu

Overview

This project aims to recognize hand gestures representing American Sign Language (ASL) letters in real time using a Convolutional Neural Network (CNN). The system translates these gestures into corresponding letters and constructs complete words from sequences of signs. It is designed to be intuitive, accurate, and responsive enough for practical communication support.

The model was trained using the ASL Alphabet dataset, which includes labeled images for the 26 English letters, along with auxiliary tokens such as `space`, `del`, and `nothing`. These additions enhance the fluidity of word construction and correction during real-time usage.

System Architecture

The application is built using TensorFlow for model inference, OpenCV for real-time video processing, and MediaPipe (via the `cvzone` module) for robust hand detection. A multi-threaded structure ensures predictions run concurrently with video capture to maintain low latency.

Key Components:

- **Hand Detection:** A single-hand detector isolates the hand region from the webcam stream.
- **Image Preprocessing:** Captured regions are resized to 64x64 pixels, converted to grayscale, normalized, and reshaped to fit the model's input requirements.
- **Prediction:** A trained CNN processes each frame to output the most likely character class, which is then matched against a label list (`A`–`Z`, `space`, `del`, `nothing`).
- **Letter Stabilization:** To improve accuracy, a gesture must be held consistently for one second with confidence ≥ 60% before it is accepted.
- **Word Formation Logic:** Recognized characters are appended to a string, with `space` indicating word boundaries and `del` allowing corrections.

Model Details

The CNN architecture includes multiple convolutional and pooling layers followed by dense layers, optimized for classification over 29 classes. The model was trained on grayscale images using categorical cross-entropy loss and Adam optimizer.

The dataset used is publicly available on Kaggle:
https://www.kaggle.com/datasets/grassknoted/asl-alphabet

Predictions are highly sensitive to hand positioning. During demonstration, users are advised to slightly rotate the hand and hold each gesture steady to increase model confidence. The signs `T` and `S` are visually similar and may be misclassified—this is a known limitation in the current model version.

Tools & Technologies

- **Libraries:** TensorFlow, OpenCV, MediaPipe (via `cvzone`), NumPy
- **Model Format:** `.keras`
- **Languages Used:** Python
- **Input Device:** Standard webcam

Conclusion

This real-time ASL detection system effectively demonstrates the power of combining computer vision and deep learning to bridge communication gaps for individuals with hearing or speech impairments. While currently limited to alphabet recognition and basic word formation, the project lays a solid foundation for more advanced systems involving full sentence generation, audio feedback, or multilingual sign language support.

With improved training data, fine-tuning, and user adaptation strategies, this approach can evolve into a highly practical assistive communication tool.