Final Project Proposal

Introduction and Problem Statement

Our project aims to develop a real-time Sign Language Interpreter that uses computer vision to recognize hand gestures corresponding to letters of the alphabet. This tool is designed to bridge the communication gap between individuals who use American Sign Language (ASL) and those who do not. By translating signed letters into their corresponding text outputs, the project supports accessibility and inclusion for the deaf and hard-of-hearing communities.

Data Sources:

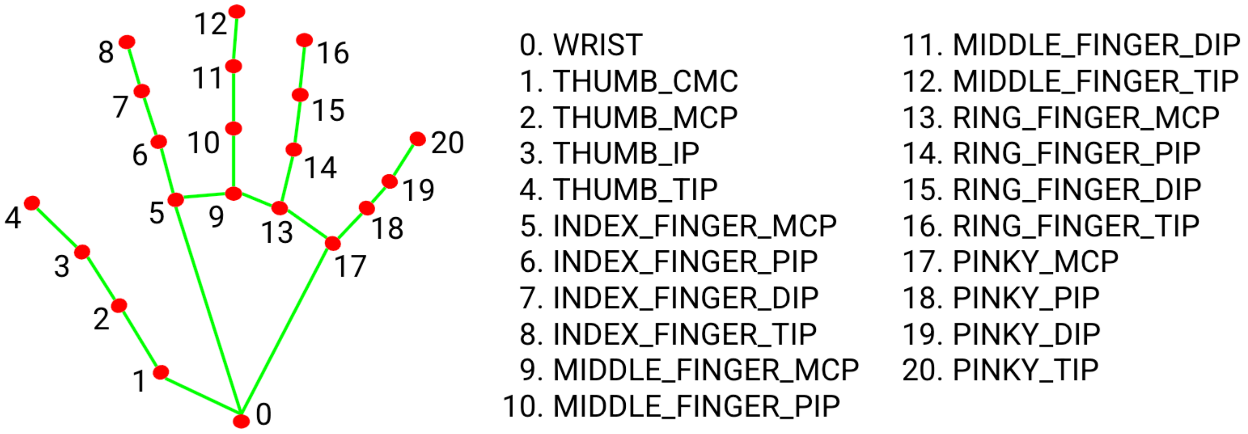
We will be using:

- Live video feed from a webcam as the primary input source for hand gestures.
  - The dataset that we will use to get data will be videos of us performing the ASL gestures
- If needed, additional gesture datasets such as the ASL Alphabet Dataset (available on Kaggle or similar platforms) to supplement training or benchmarking.

Methods, Techniques, and Technologies:

We plan to use the following:

- **OpenCV** for real-time image capture and basic image processing.
- **MediaPipe Hands** for accurate detection and tracking of hand landmarks.



0. WRIST
1. THUMB_CMC
2. THUMB_MCP
3. THUMB_IP
4. THUMB_TIP
5. INDEX_FINGER_MCP
6. INDEX_FINGER_PIP
7. INDEX_FINGER_DIP
8. INDEX_FINGER_TIP
9. MIDDLE_FINGER_MCP
10. MIDDLE_FINGER_PIP
11. MIDDLE_FINGER_DIP
12. MIDDLE_FINGER_TIP
13. RING_FINGER_MCP
14. RING_FINGER_PIP
15. RING_FINGER_DIP
16. RING_FINGER_TIP
17. PINKY_MCP
18. PINKY_PIP
19. PINKY_DIP
20. PINKY_TIP

- **Custom logic or ML models** for mapping hand landmarks to specific letters using either classification, a CNN or both

    - **Classification Pipeline:** MediaPipe extracts 21 hand landmarks → flatten (×2 or ×3 coords) → feed into a small ML classifier.
    - **CNN Pipeline:** Crop/resize webcam ROI around the hand → feed into a light CNN
- Python as the main programming language.

Products to be Delivered

Real-Time Sign Language Interpreter Application

A standalone or script-based application that performs the following:

- **Live Gesture Detection:** Uses a webcam to capture a live video feed and applies real-time hand detection using MediaPipe Hands and/or OpenCV.
- **Gesture-to-Letter Translation:** Recognizes static hand signs corresponding to ASL alphabet letters (A–Z) and maps them to their respective text characters.
- **Visual Feedback:** Displays the translated letter on screen, updating as the user signs different letters.
- **Accuracy & Responsiveness:** Optimized for low-latency inference, maintaining a balance between speed and prediction accuracy.

Final Report (Documentation)

A comprehensive project report that includes:

- A clear explanation of the problem and its significance.
- Details of the system architecture and each component (image processing, landmark extraction, classification logic or model).
  Model selection rationale (why classification or CNN was used).
  Challenges encountered (gesture ambiguity, lighting, occlusion) and how they were addressed.
- Evaluation metrics (accuracy, latency, etc.) and results.
- Future improvements or extensions (multi-hand support, word recognition).

Demo Video

A recorded demonstration showing:

- The setup and how the tool is used.
- Live signing of various ASL alphabet letters.

- Real-time prediction and screen output as letters are signed.

Source Code & Readme

- Full Python code with organized structure and comments.
- A README.md file with:
  - Setup instructions
  - Usage guide
    Dependencies and environment requirements
  - Optional: instructions for retraining or adding more gestures