

SIGN-LANGUAGE

Spring 2026
Meeting No.2

System Development

SYSTEM DEVELOPMENT

Input: Video stream (webcam)

Output: Text (spoken language)

OpenCV module will be used

Supports ViViT and other video models

Supports on-device only

Camera → Frame Capture → Preprocessing → Model Inference → Text Output

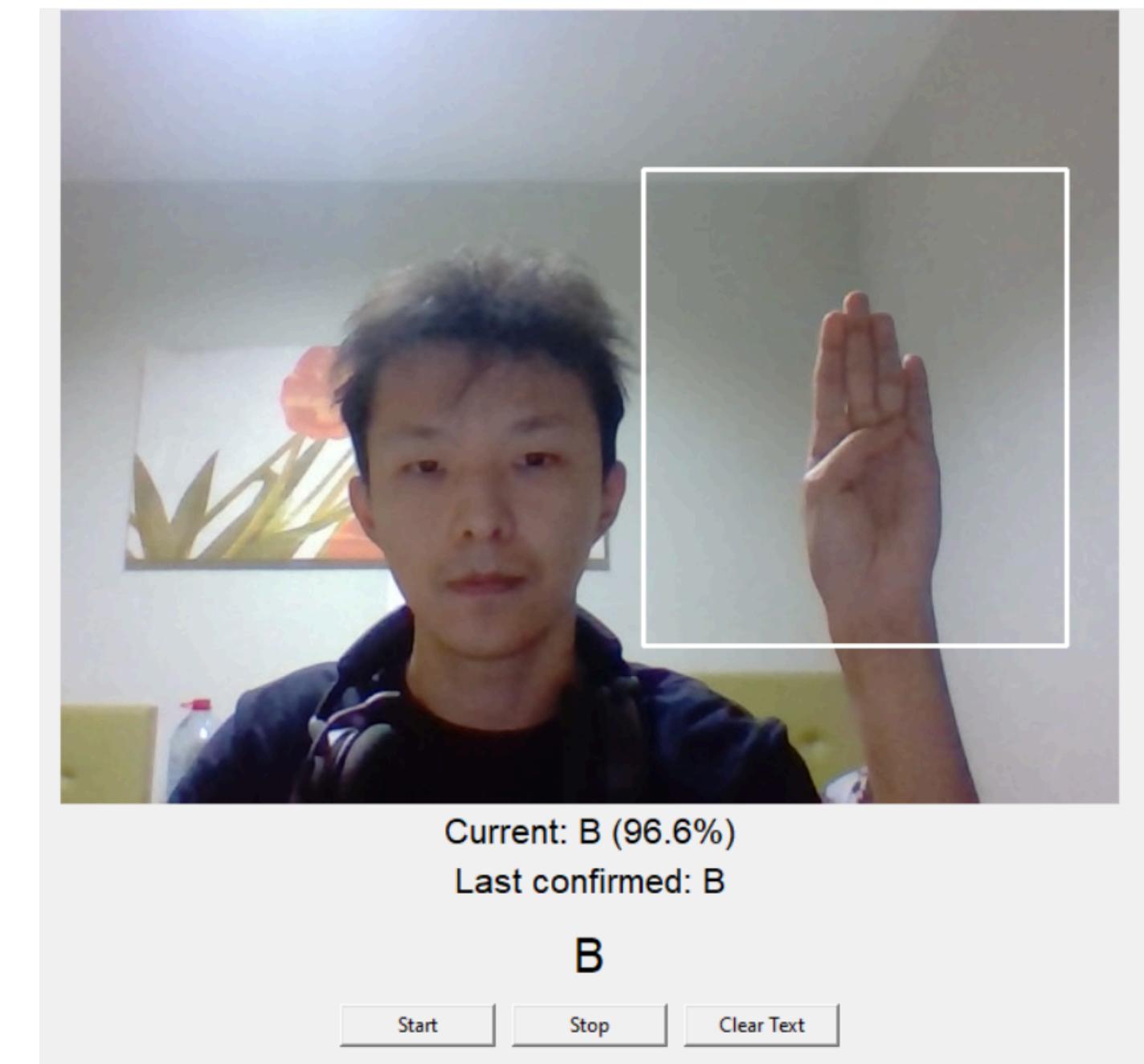
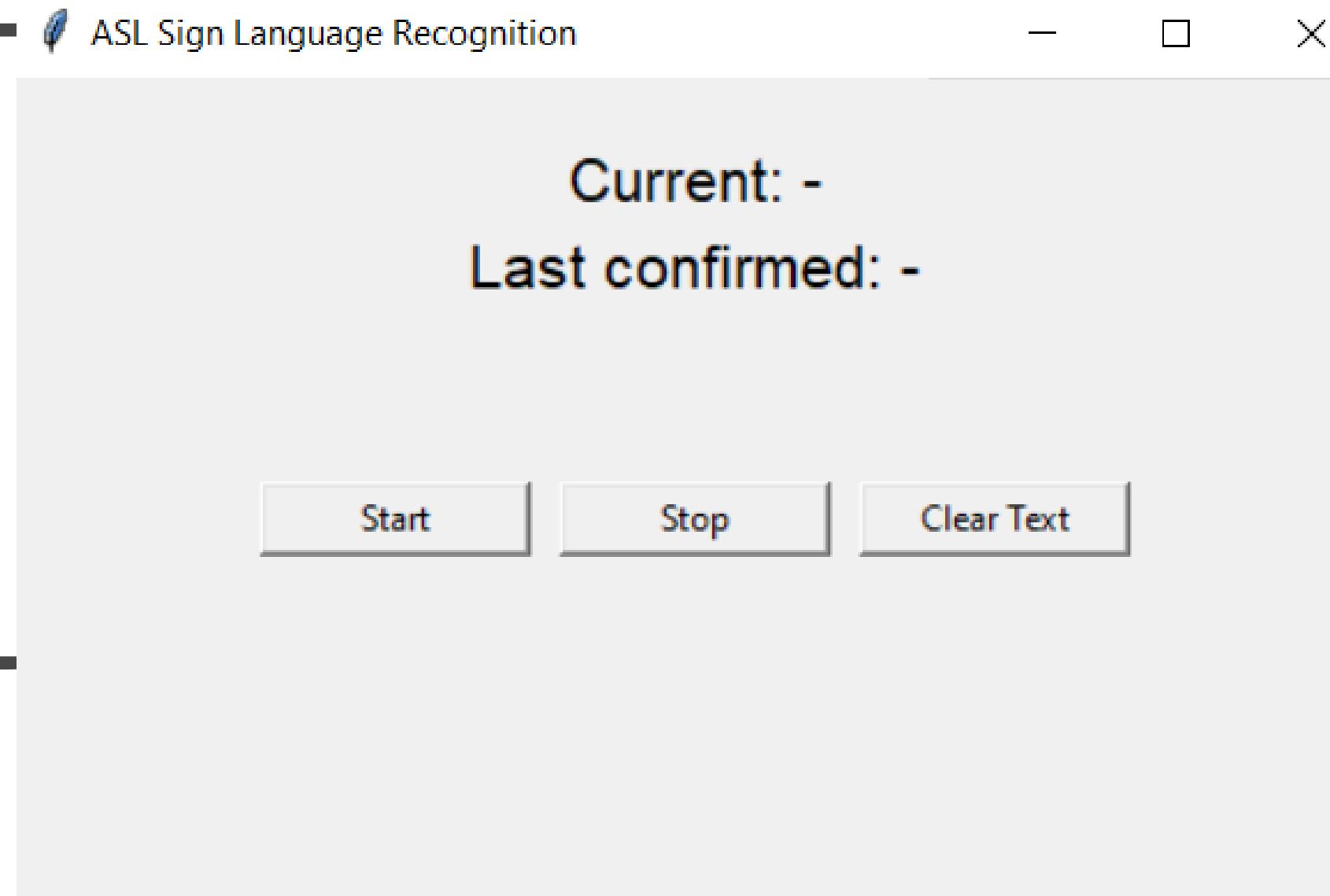
DATASET & MODEL SELECTION

American Sign Language was used to test the model

ResNet was used as the main model (already pre-trained and for real time)

- 29 classes
- evenly spaced (30 images per class) samples were selected)
- 29 classes (A-Z + space, delete, nothing)

Application Interface



Preprocessing Pipeline

Region of Interest (ROI) extraction to isolate the hand
Image resizing to 224×224

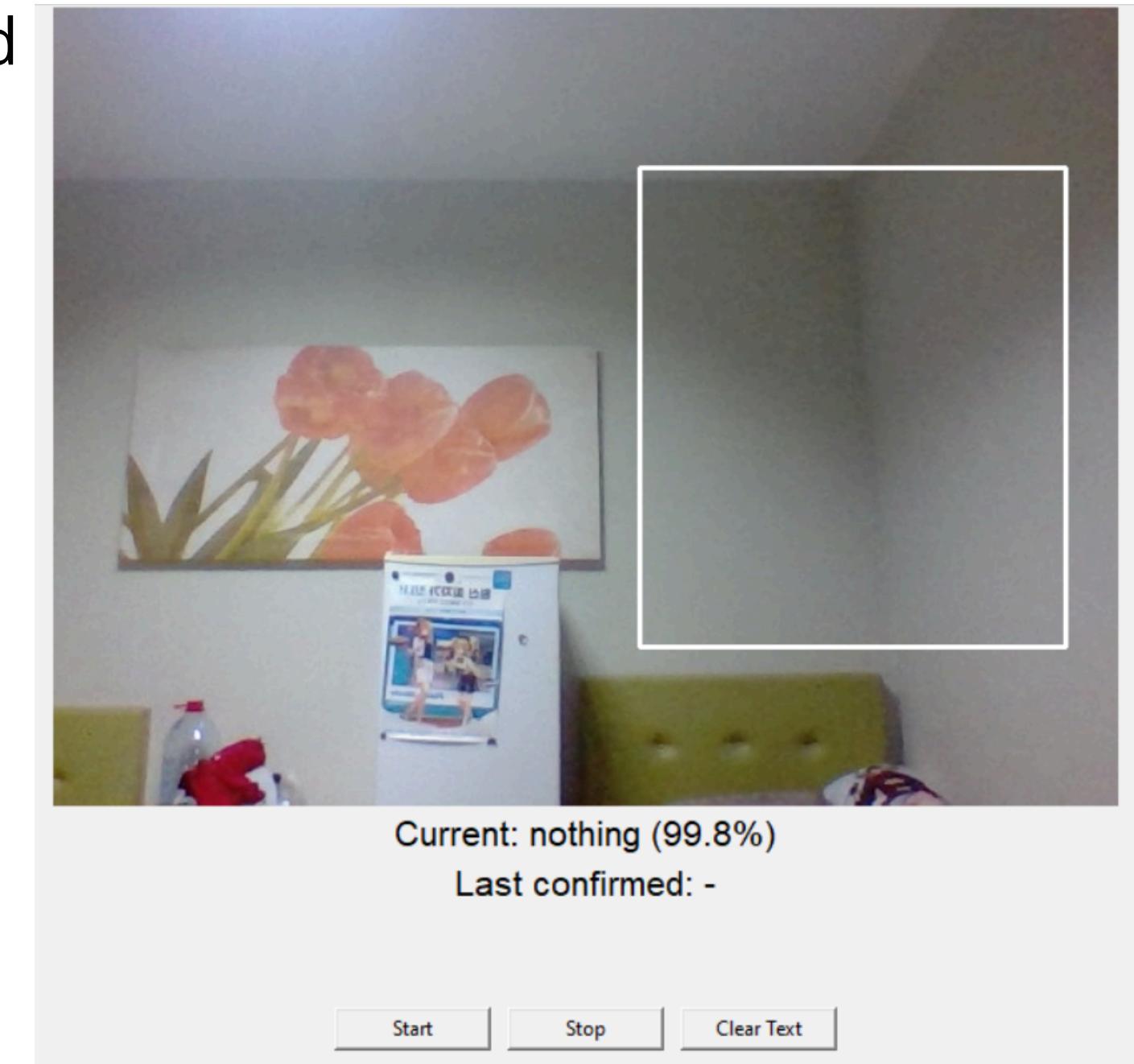
RGB conversion

Normalization using ImageNet statistics

Reduces background noise

Ensures consistency with training data

Improves model performance



Prediction Stability & Noise Reduction

Problem:

- Frame-by-frame predictions fluctuate due to small hand movements

Solution:

- Time-based confirmation mechanism
- A sign must remain stable for 3 seconds
- Predictions below confidence threshold are ignored

Result:

- Significantly reduced flickering
- More natural interaction

Threshold for confirmation is 65%

Sign Confirmation & Special Cases

State tracking includes:

- Current candidate sign
- Candidate start time
- Confirmed sign
- Release state (hand relaxed)

Key idea:

- A sign is confirmed only after stability
- Repeated letters require a release (e.g., "L → nothing → L")

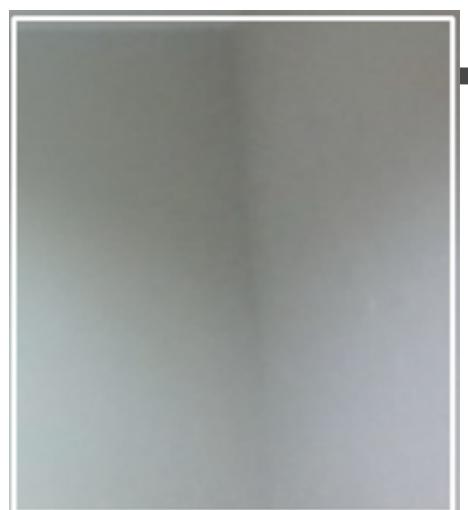
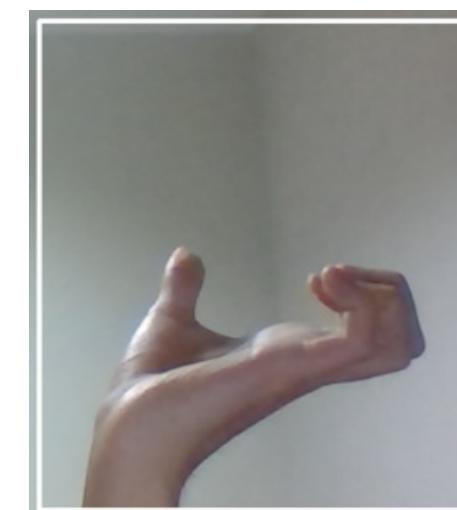
This logic functions as a finite-state machine.

Special gestures:

- nothing: idle / no valid sign
- space: inserts space
- delete: removes last character

Design decision:

- nothing is ignored and used only to detect release
- Prevents incorrect text insertion



Challenges & Solutions

Technical challenges:

- Prediction flickering
- Repeated letter handling
- Idle gesture misclassification

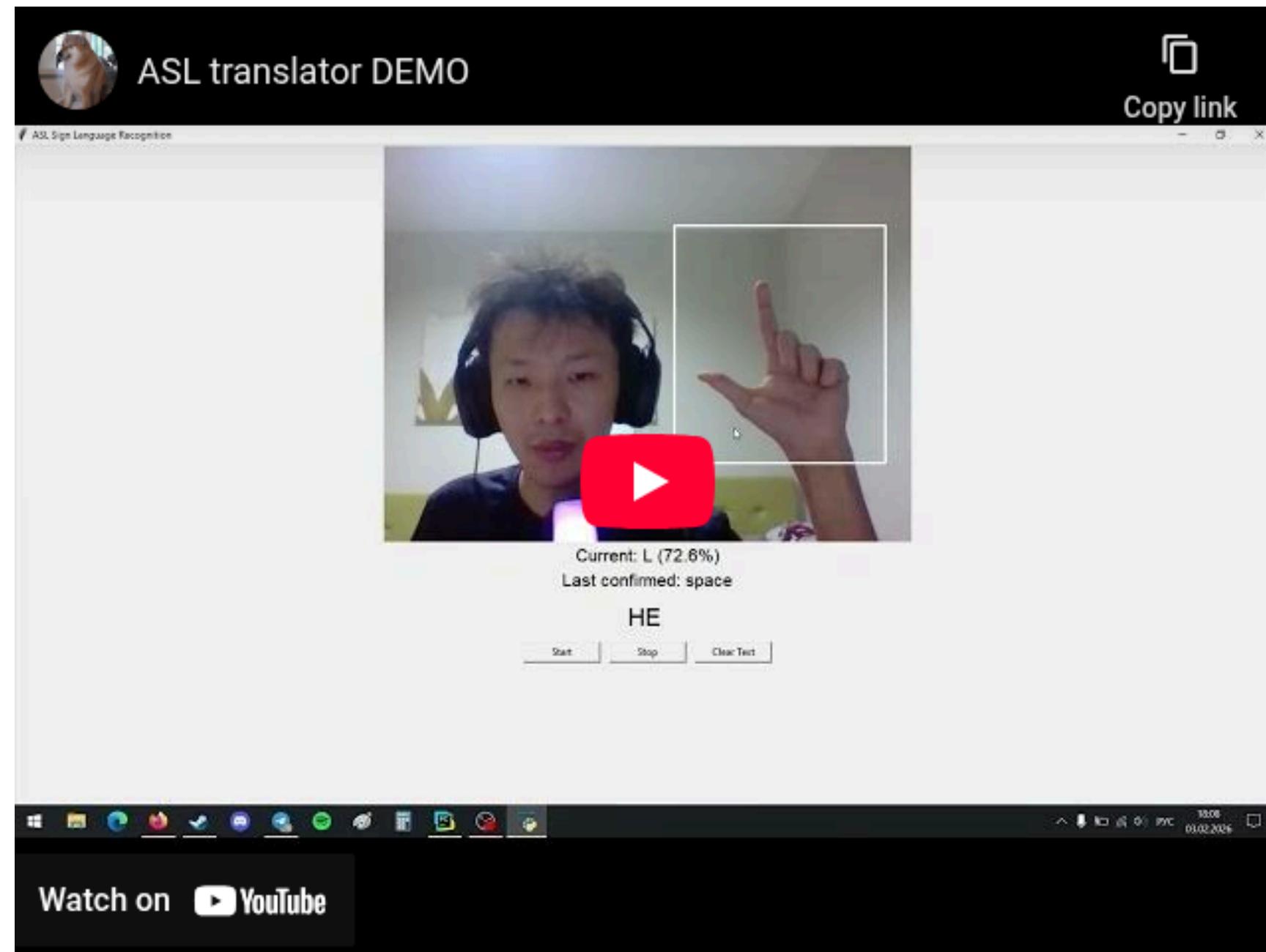
Key solutions:

- Time-based confirmation instead of frame-based
- State reset logic
- Confidence thresholding (65%)

Examples:

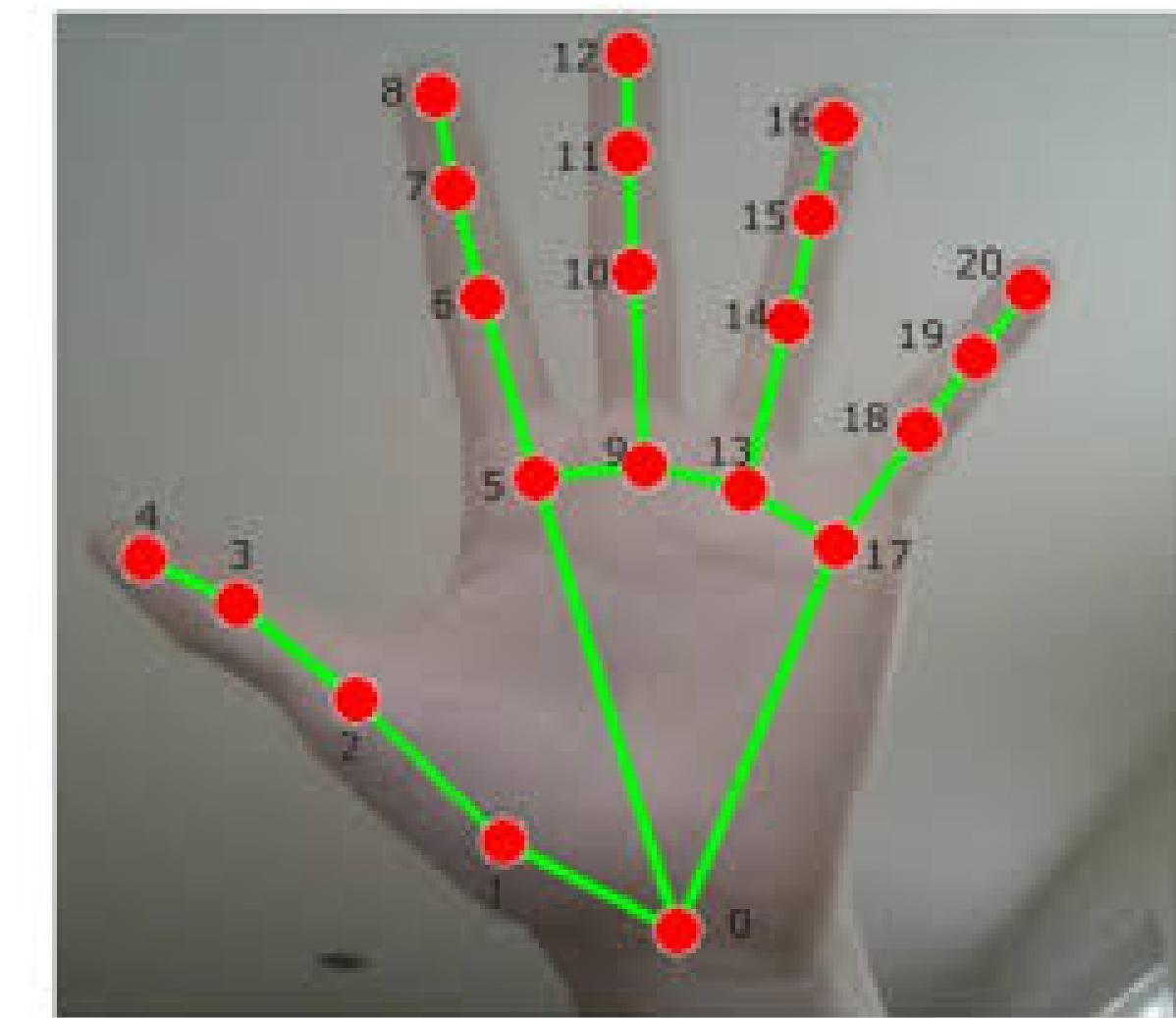
- HELO WRD
- YEDILnothing
- EEEEEEEEEE

Video Demonstration



Future Improvements

- Hand detection using MediaPipe
- Language model for autocorrection
- Support for continuous sign language
- User-friendly interface





THANK YOU