

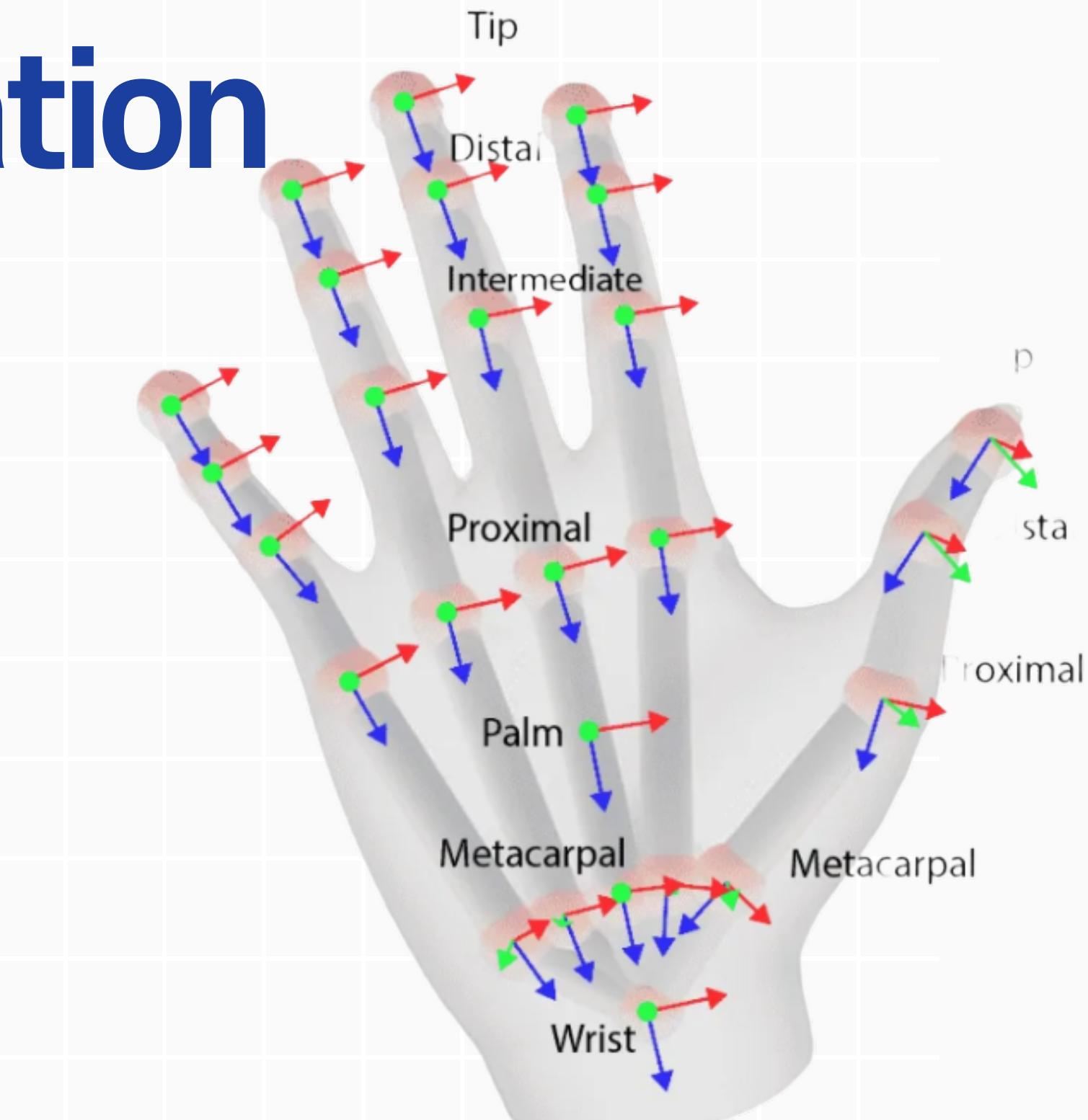
Sign Language Translation

with Gesture-Based Emotion Analysis

복현준 (2021320020)

최민석 (2023320177)

May 29, 2025



Goal

- Realtime Sign Language Translation
- Natural communication through emotion-aware interpretation

I'm honestly sooo hyped right now
—this is insane, right?

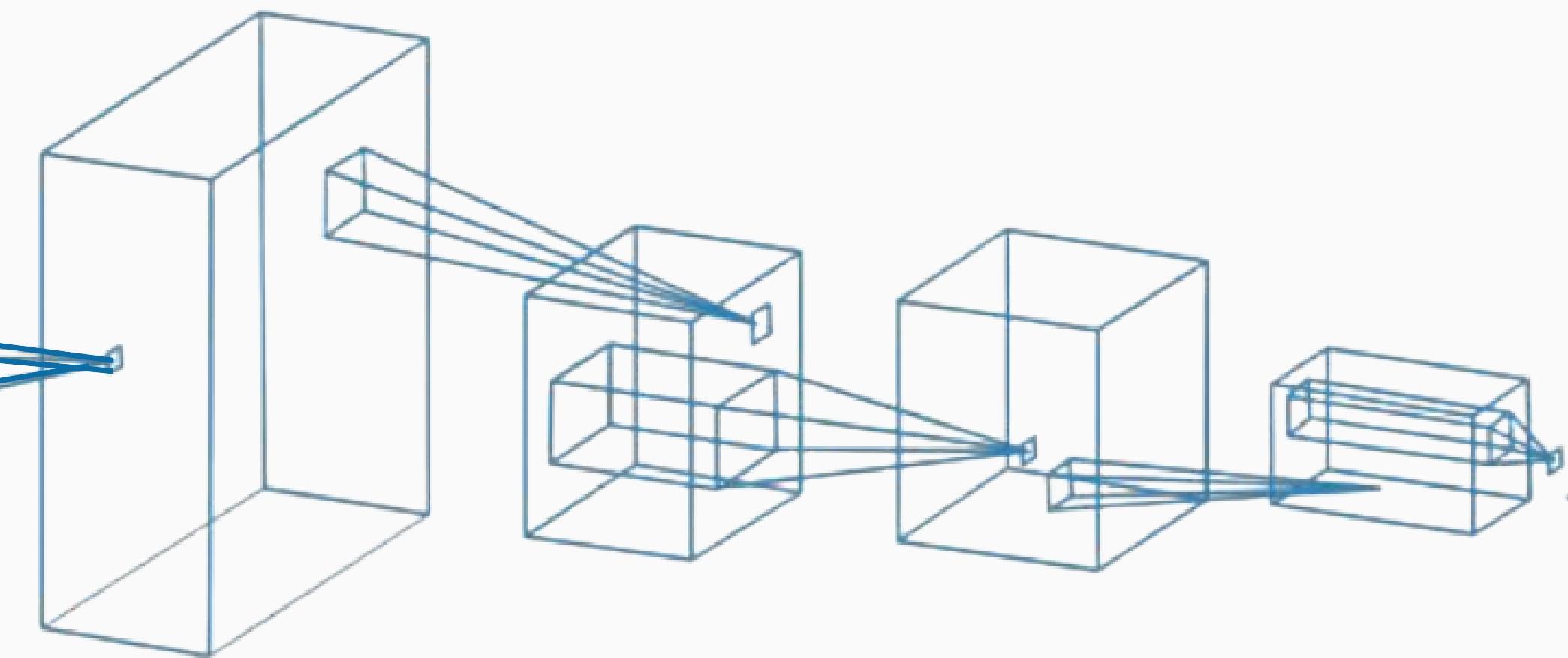
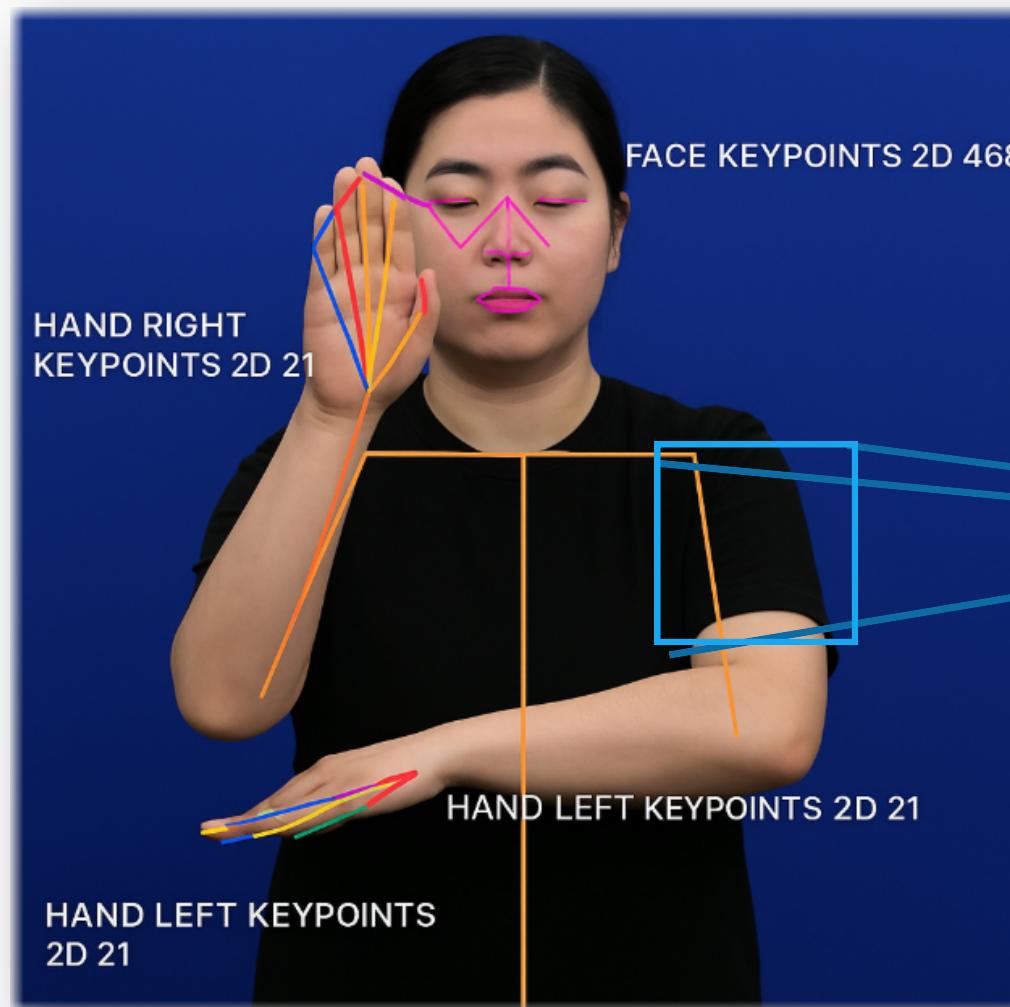
Expected Benefits

- Real-time translation with voice output in video calls
- Essential interpretation support in public sectors
- Improved accessibility for signers



Challenges

- Fine-grained emotion detection requires CNN-based image processing
- Applied to every frame, leading to significant computational load
- Not suitable for real-time processing

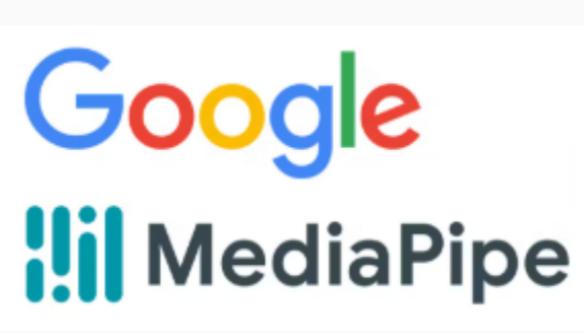


Implementation - model

- Applied LSTM model (CNN not used)
- Focused on sequential data across movements, not just static hand shapes

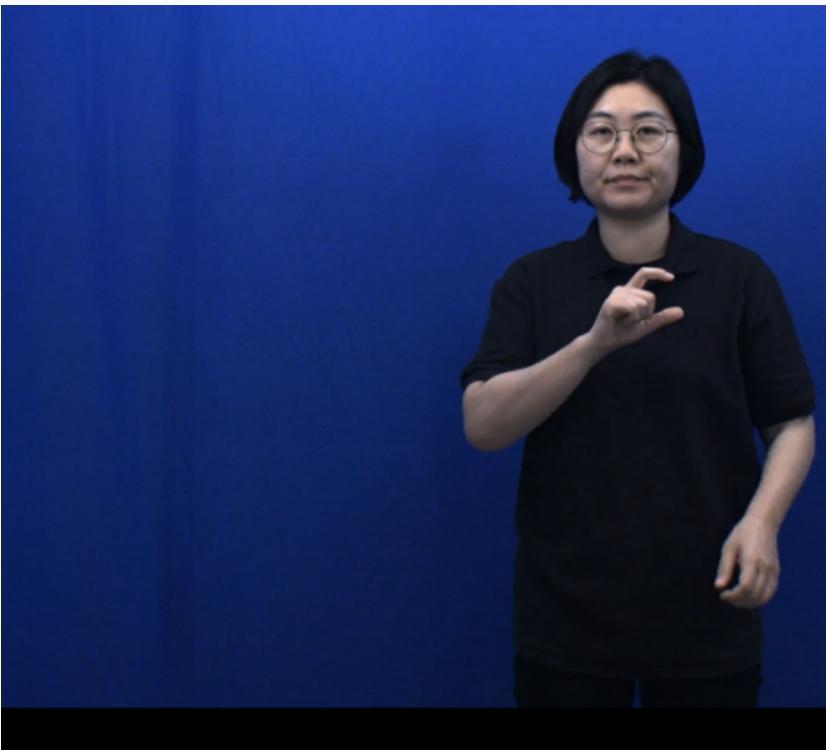


- Keypoints collected from webcam using MediaPipe and OpenCV



Implementation - Coordinate Normalization

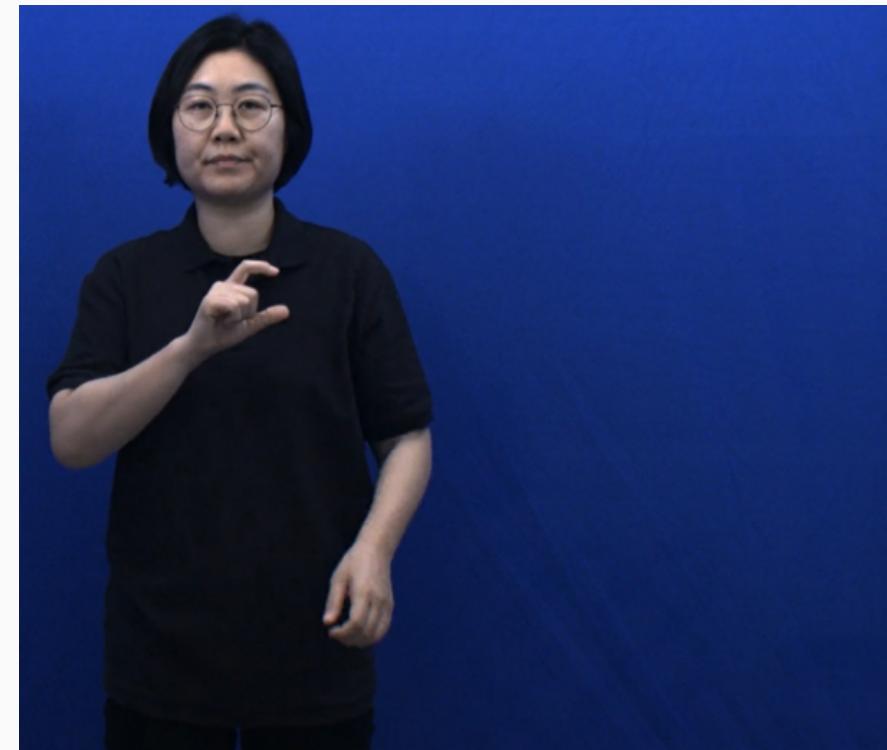
- Keypoint normalization enables learning of relative motion
- This helps ensure robust learning despite variations in angle, subject, or occlusion.



Right Hand

Global point : (305, 281)

Relative point : 100 points below the nose



Right Hand

Global point : (102, 281)

Relative point : 100 points below the nose

Implementation - Motion Speed Detection

- Measures motion speed based on frame intervals
- Fewer frames for the same motion are interpreted as faster movement

Frames : 70fps

Speed : Fast



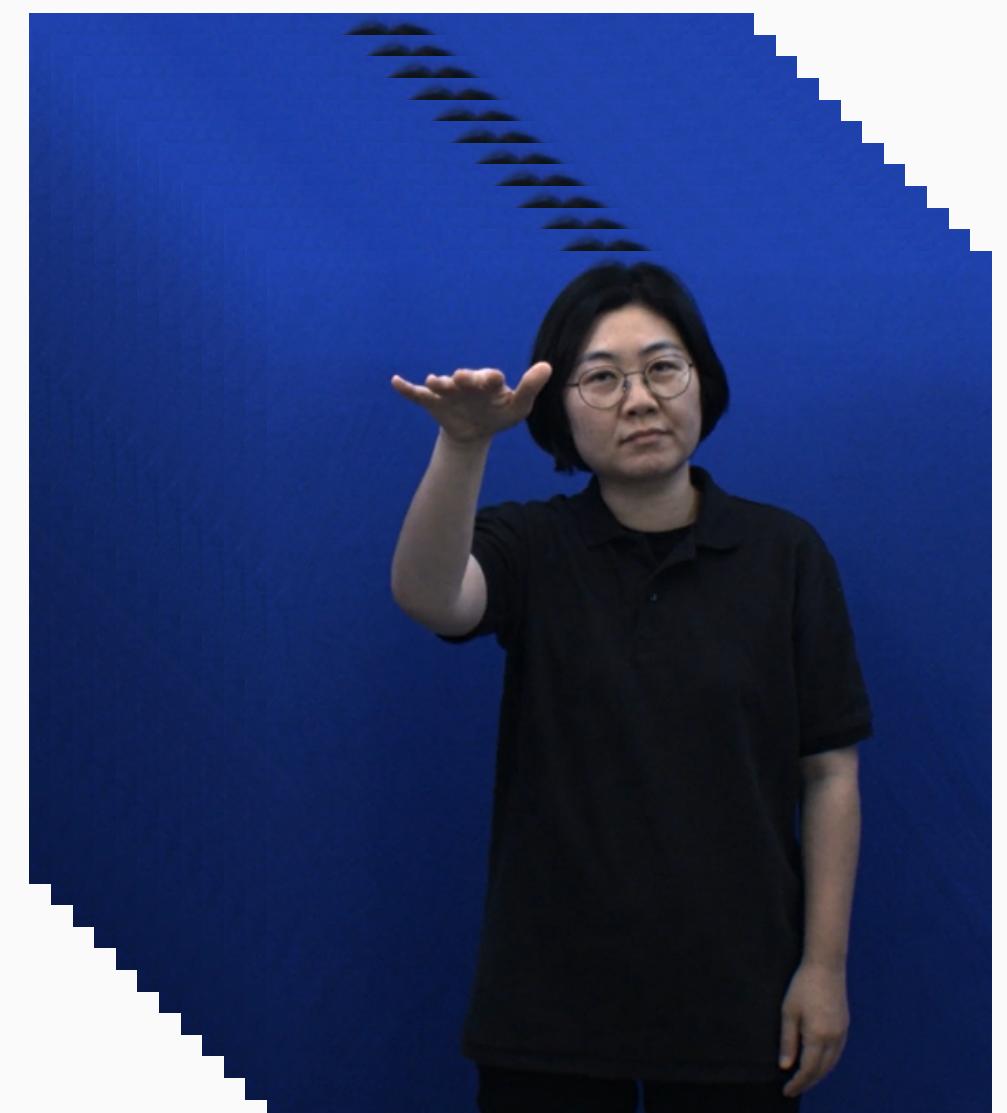
Frames : 120fps

Speed : Normal



Frames : 150fps

Speed : Slow



Implementation - Frame Scaling

- Detects motion speed by applying multiple frame **using interpolation**

Increase Frame

70 fps

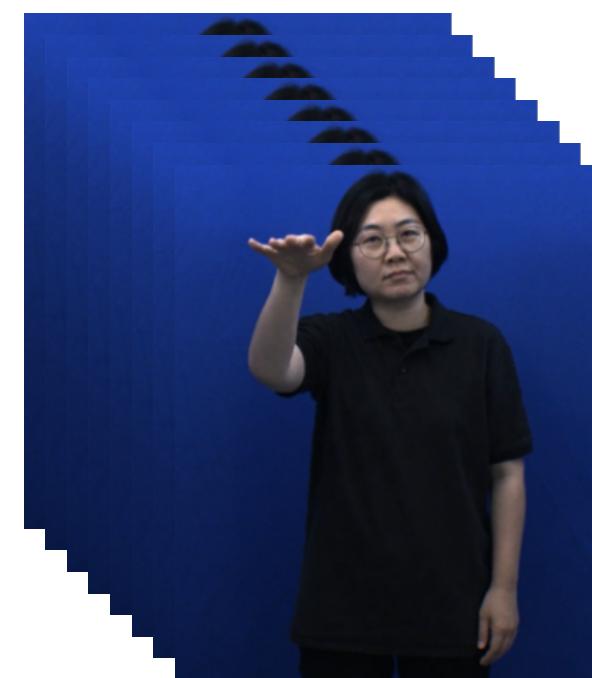


120 fps



Decrease Frame

150 fps

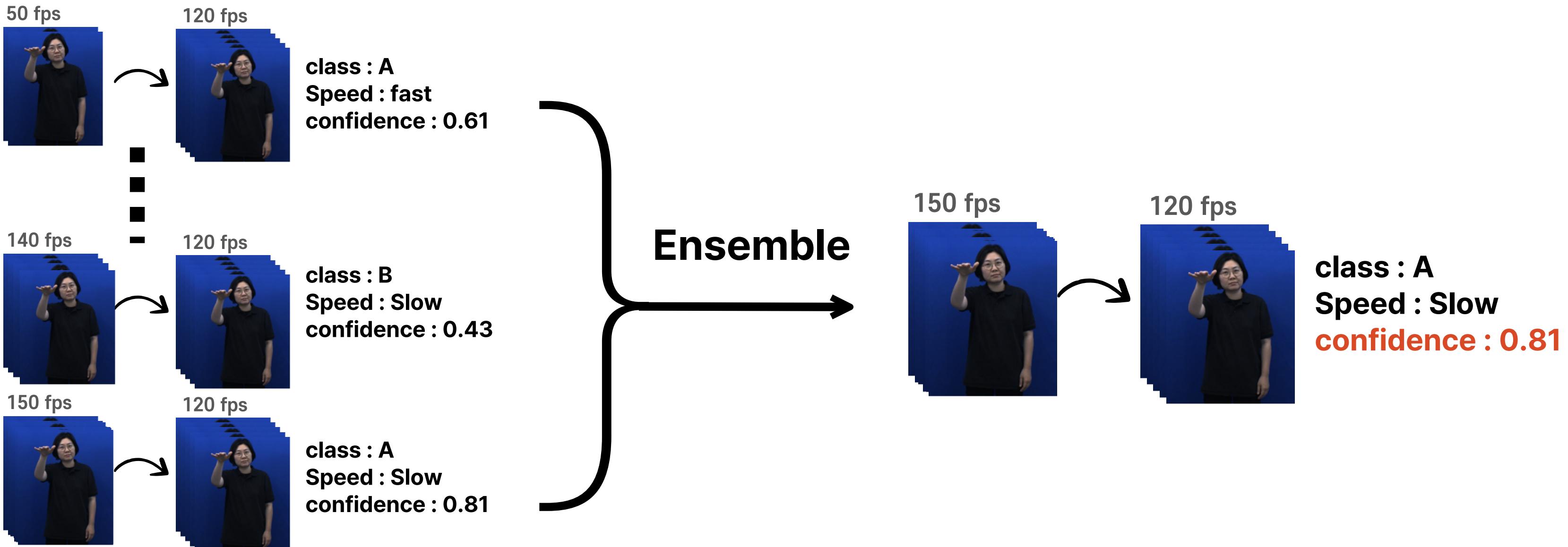


120 fps



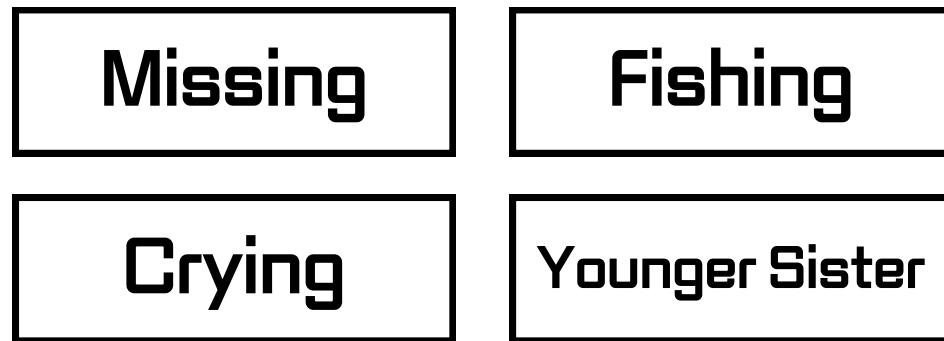
Implementation - Multi-Scale Sliding Window

- 10 sliding windows ranging from 50 to 150 frames, in 10-frame increments
- The most confident window determines the class and motion speed.



Training Data Set

4 Words



5 Camera Angles

X



X



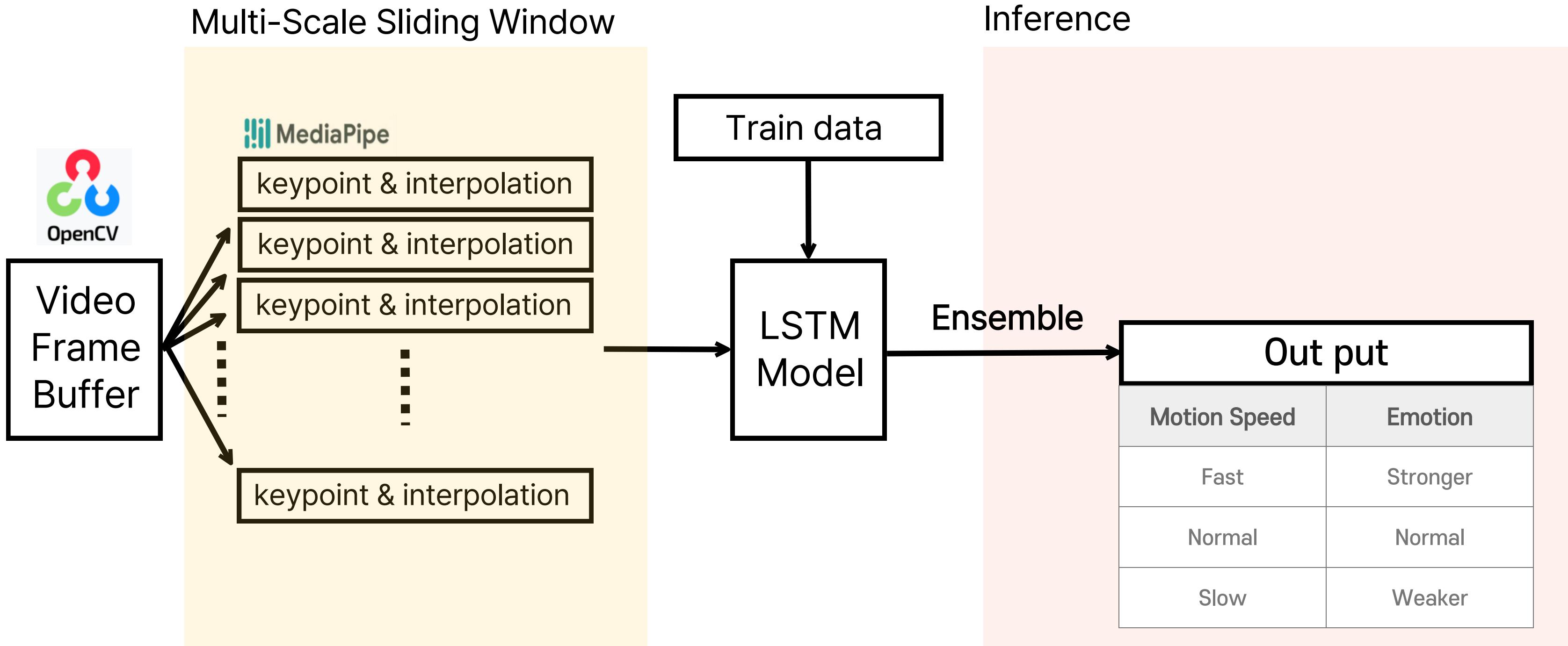
4 Participants



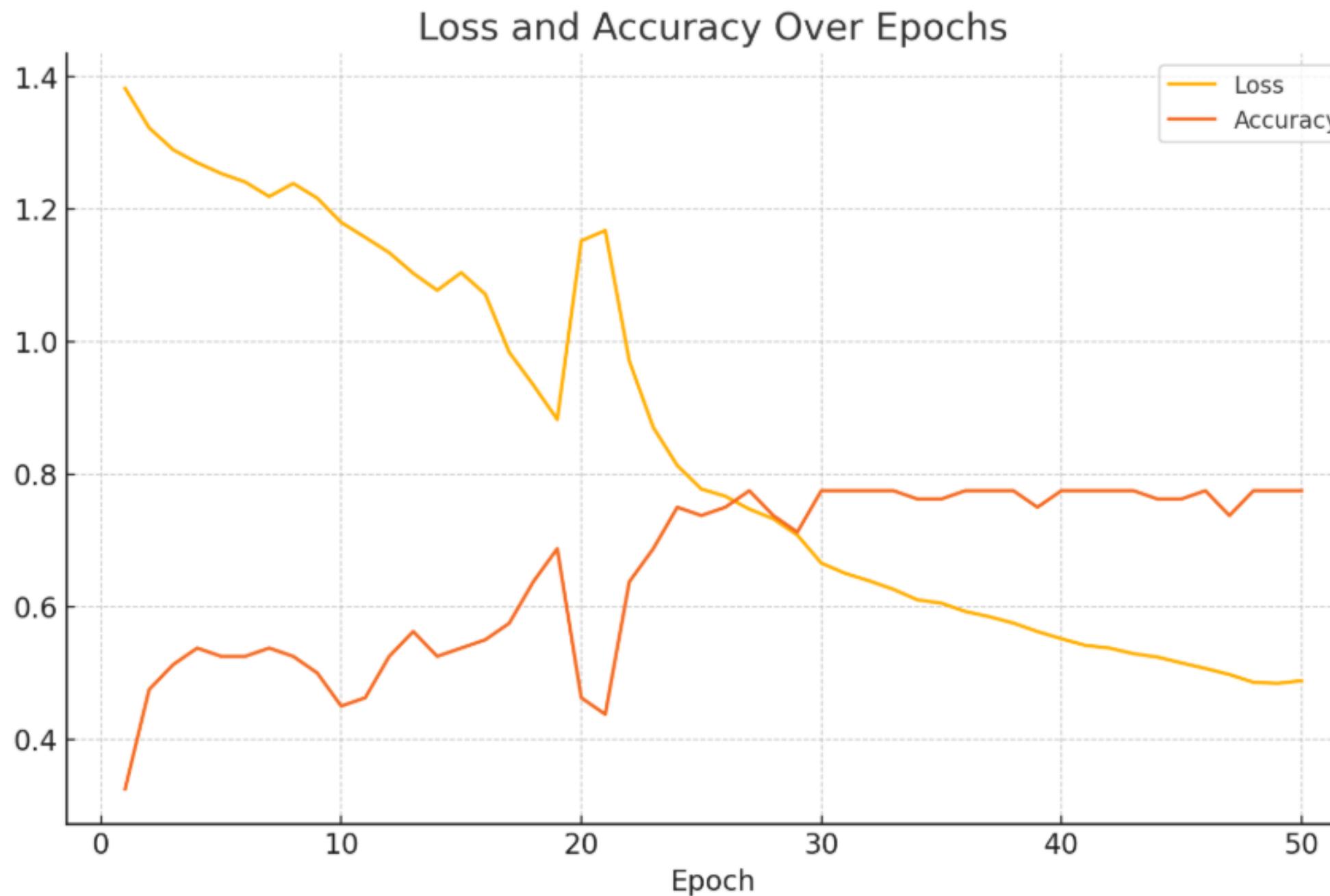
80 word samples in total, each represented by approximately 110–120 JSON files

► Trained on ~10,000 JSON files in total

System Architecture



Training Results



LSTM Training Curve (Cross-Entropy Loss)

Early Phase (Epochs 1–7)

- Model quickly learns key patterns

Fluctuation Phase (Epochs 15–25)

- Temporary overshooting in loss

Convergence Phase (Epochs >25)

- Loss decreases gradually and stabilizes
- Accuracy converges to ~0.75

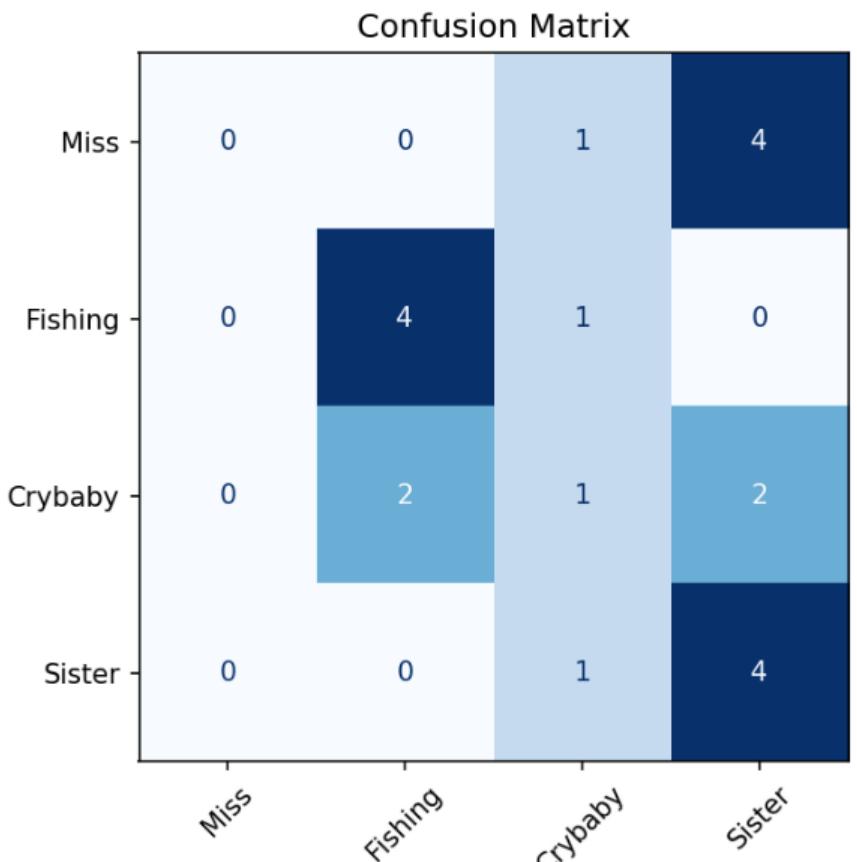
Parameter Tuning Results

- Effect of hidden_dim on Overfitting and Underfitting

Under Fitting

hidden_dim : 2

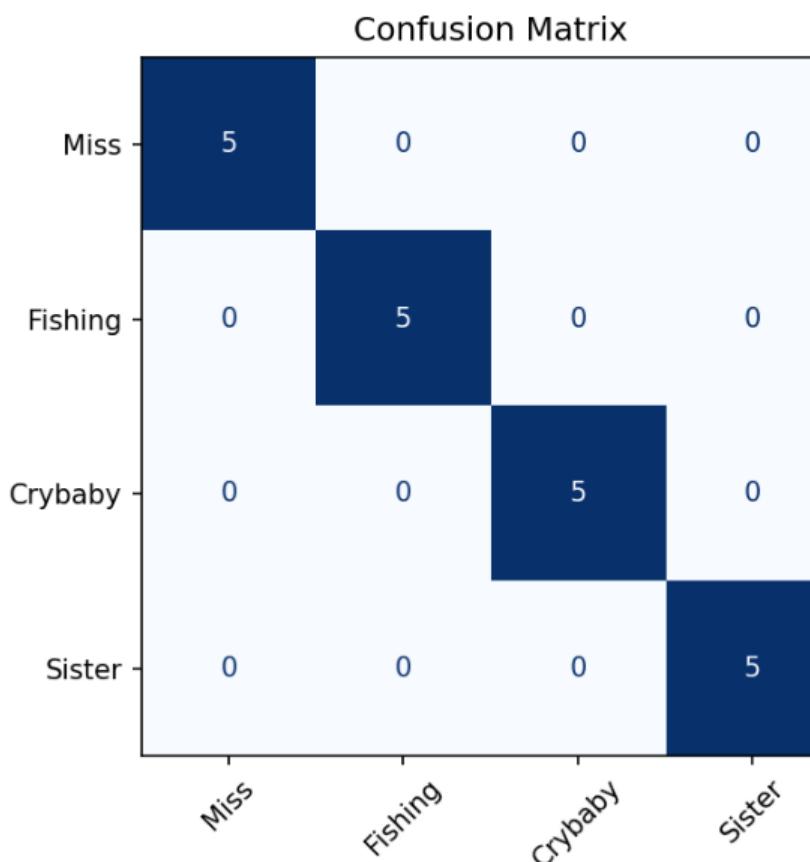
Training Accuracy : 0.45



Optimal

hidden_dim : 20

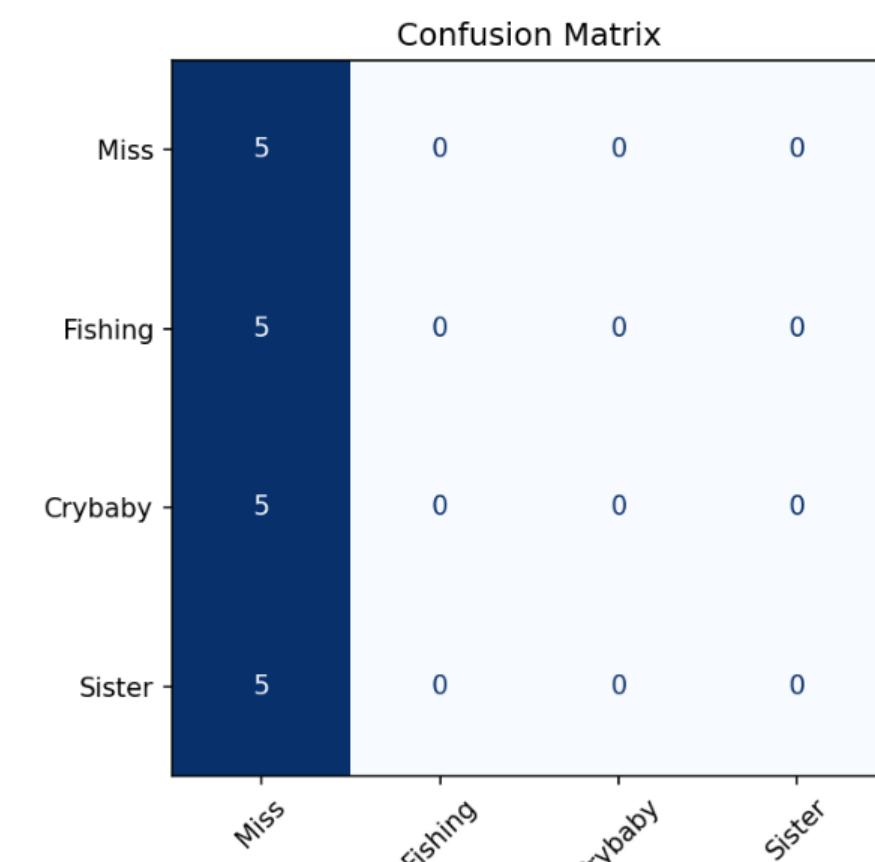
Training Accuracy : 0.78



Over Fitting

hidden_dim : 128

Training Accuracy : 0.75



Dem^o Video

END

QnA