SVR 2024
26th Symposium on Virtual and Augmented Reality

Sign2Sign - A First Attempt

Titus Weng and SingChun Lee

Computer Science
Bucknell University

Image illustration by AI

Agenda

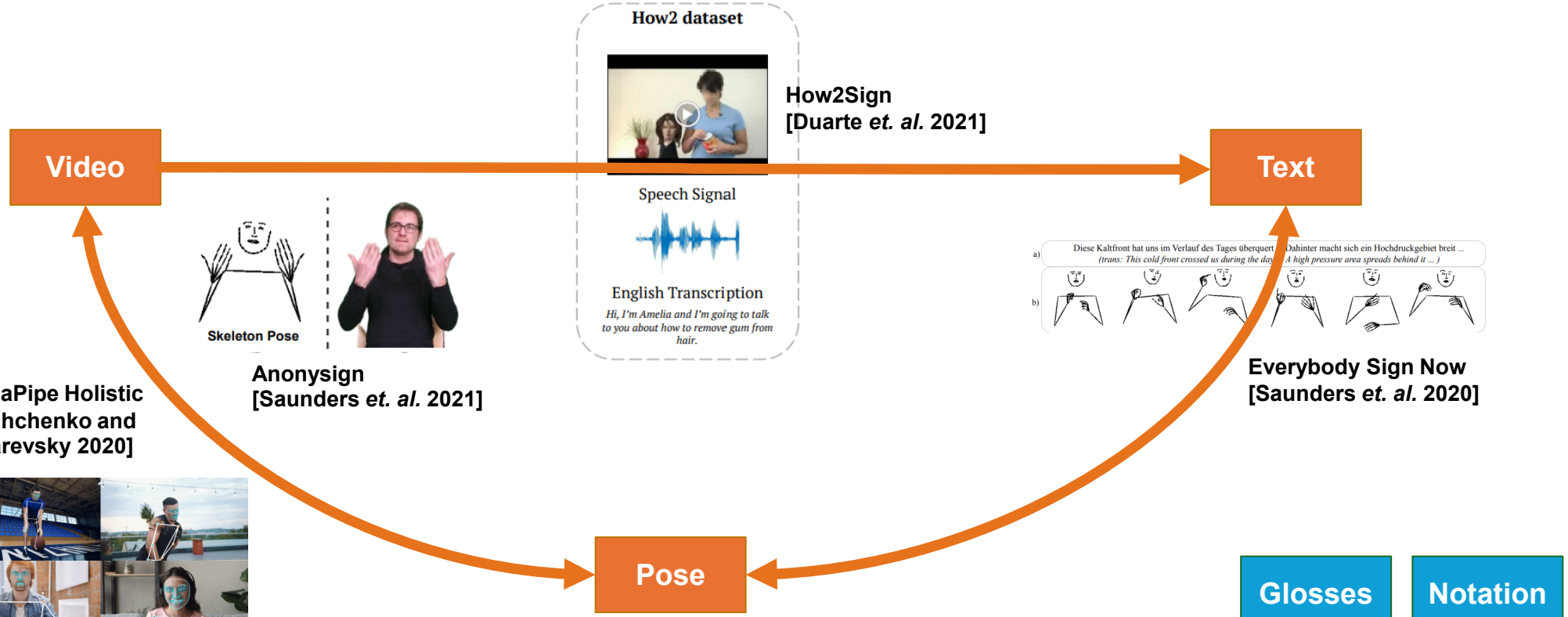
- Introduction
- Our Method
- Current Result
- Work-in-progress

Introduction

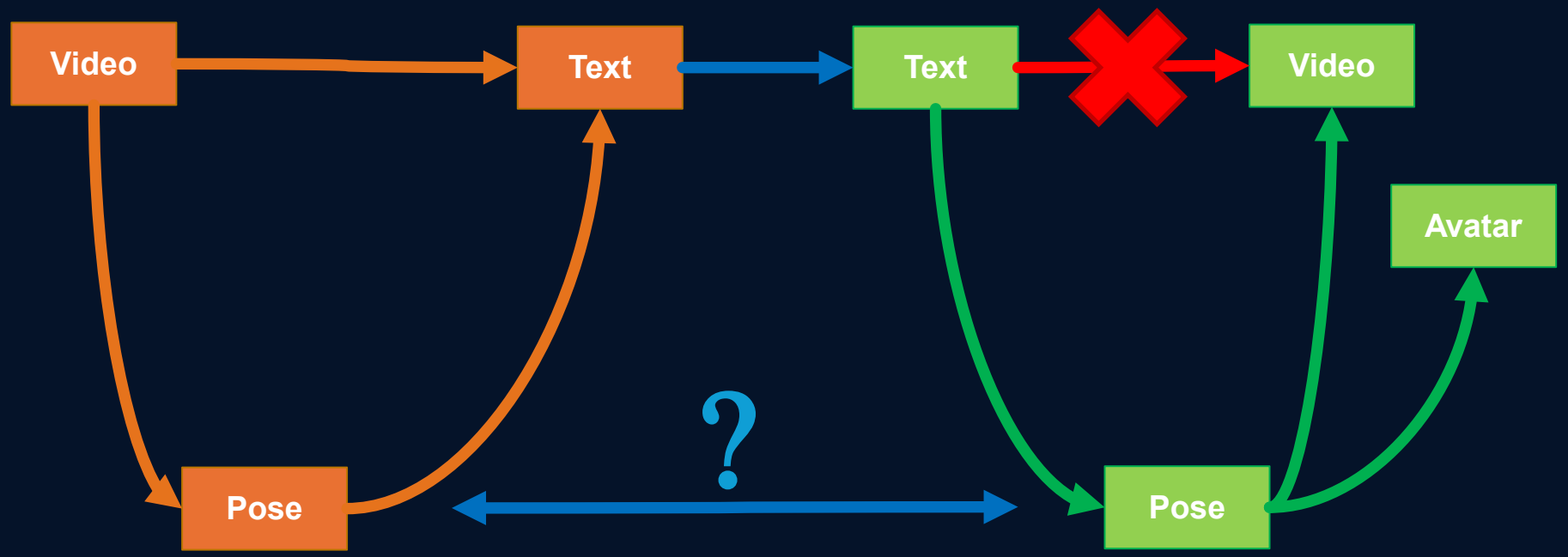
Source: Keiichi Matsuda’s Hyper-reality

What about for the Deaf and Muted?

What have been done in the literature?

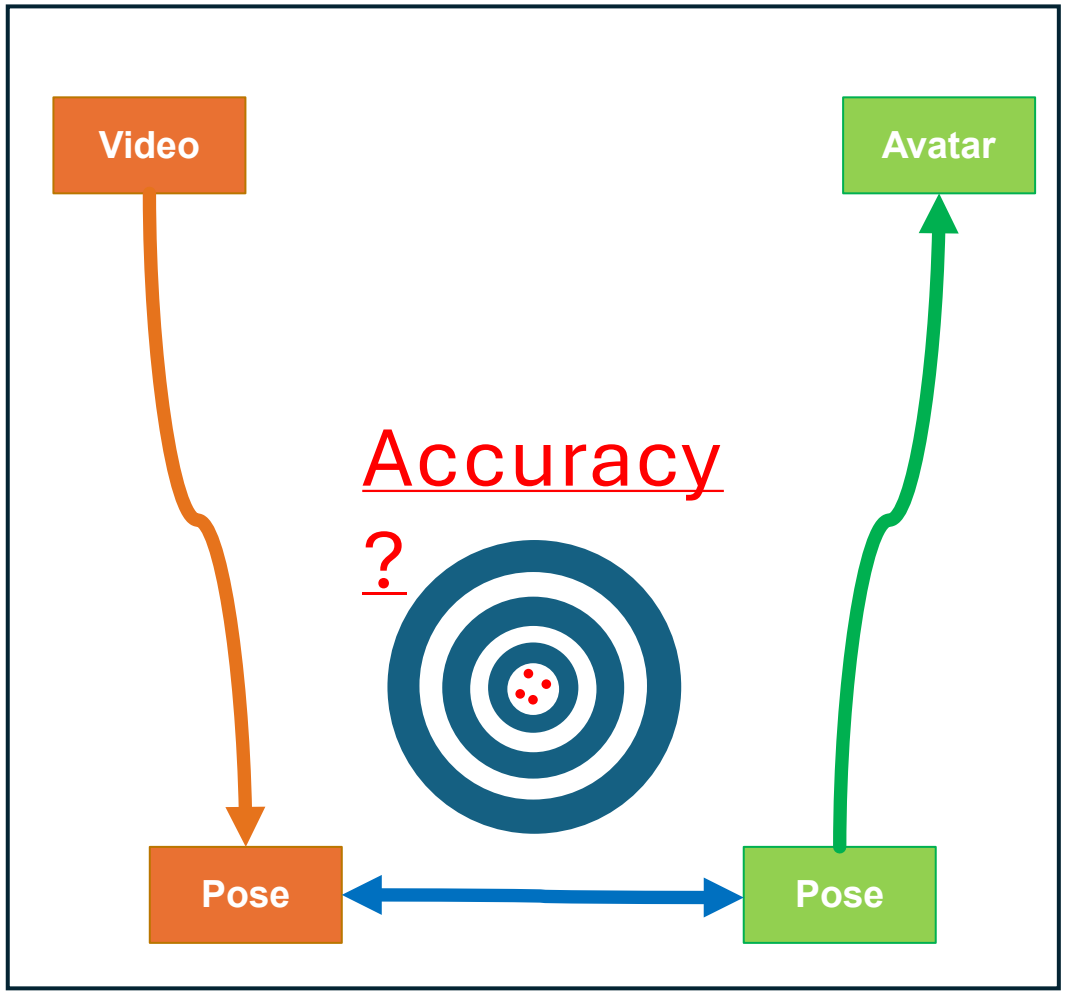


How to translate from one sign language to the other?



Research Statement

- How accurate is real-time sign2sign (pose-to-pose) translation?



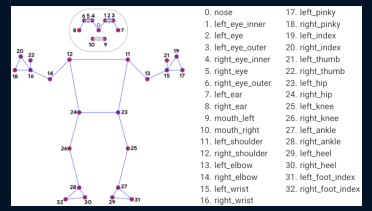
Research Values

- Low-cost system for sign language translation.
- Direct translation, no text involved.
- Provide immersive XR experience for direct communication for Deaf and Muted communities.

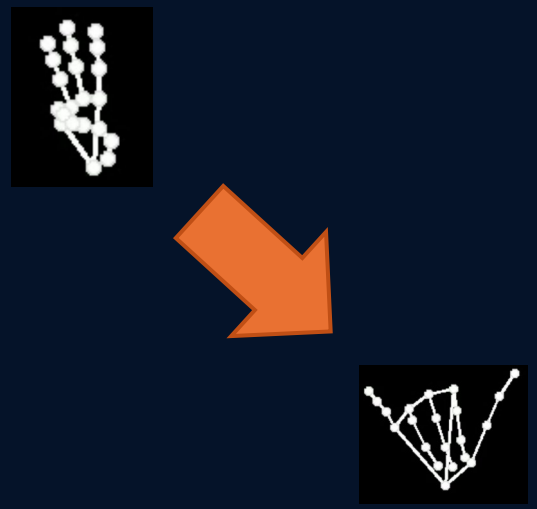
Image illustration by AI

Our Approach

MediaPipe Holistic [Grishchenko and Bazarevsky 2020]



Video to Pose

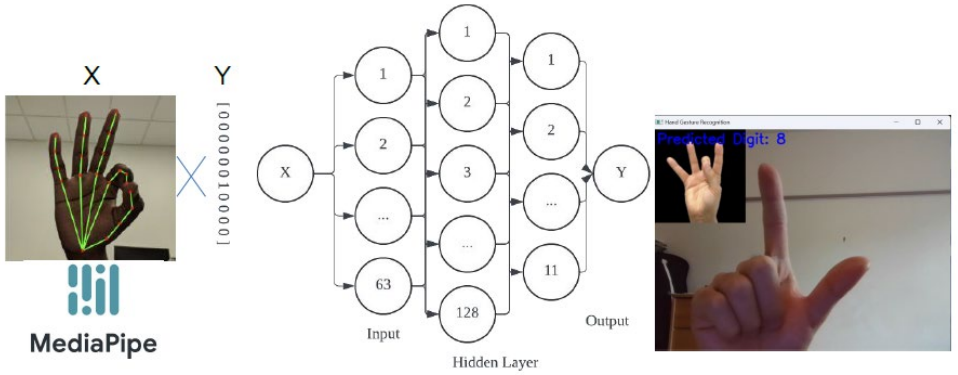


Pose to Pose

SiMAX - SignTime GmbH

Pose to Avatar

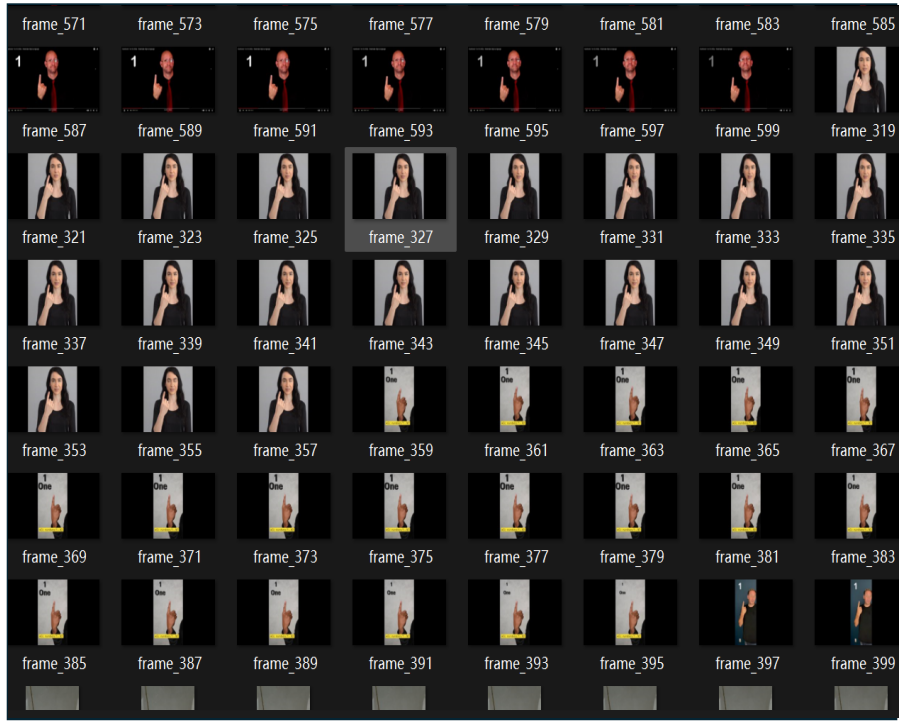
Our First Attempt



- American Sign Language ⇔ Chinese Sign Language
- 10 hand gestures (from number 1 to 10)
- 21 keypoints of a hand
- Covolution Neural Network

Training

- 14K total images (7K each)
- 80% used for training and 20% for testing
- Ground-truth: a probability vector of size 11.
  - One of each gesture and an additional one for non-digit. e.g. [1, 0, 0, …, 0] for gesture 1.

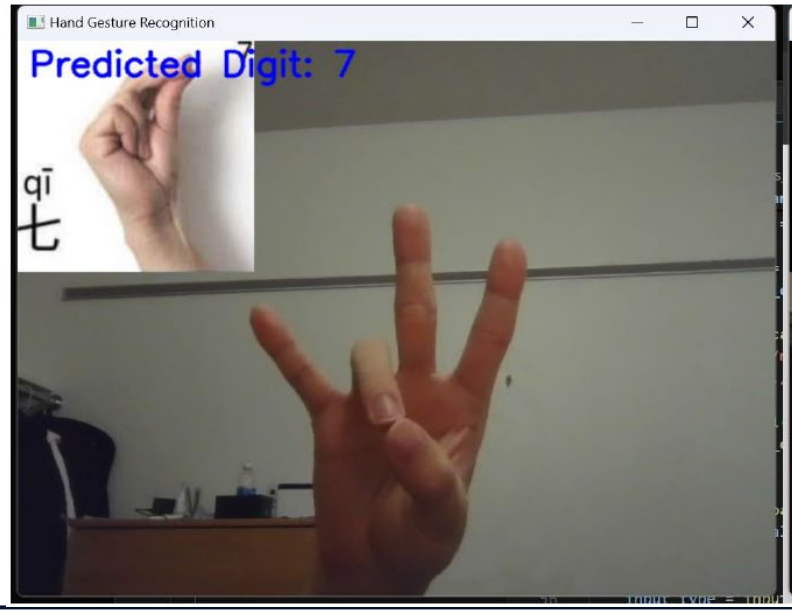


Data from public domain

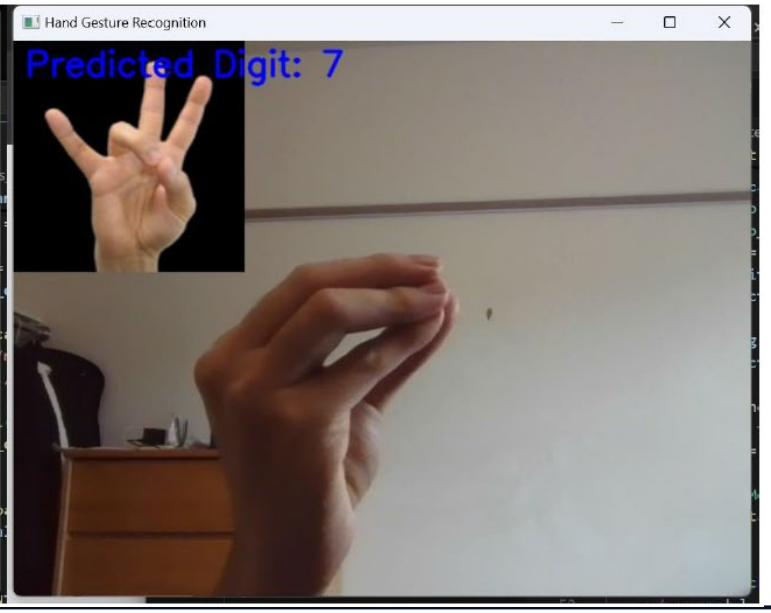
Testing

- Given a prediction vector from the model, the predicted gesture has the highest probability.

ASL to CSL



CSL to ASL

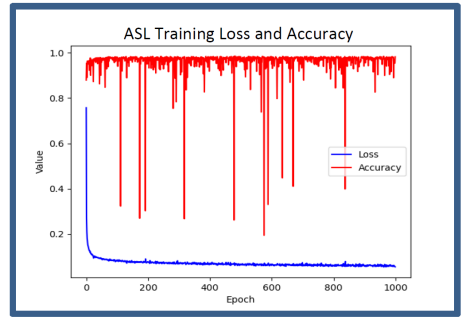


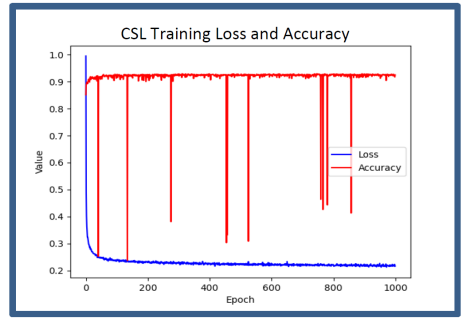
Discussion

- Accuracy fluctuation
  - Likely caused by insufficient data
  - The model may have been trapped into local minimum
- The first attempt was a success (high accuracy)
  - However, we did not achieve pose-to-pose yet
  - And the results are limited to only ten gestures

Average Training Loss and Accuracy

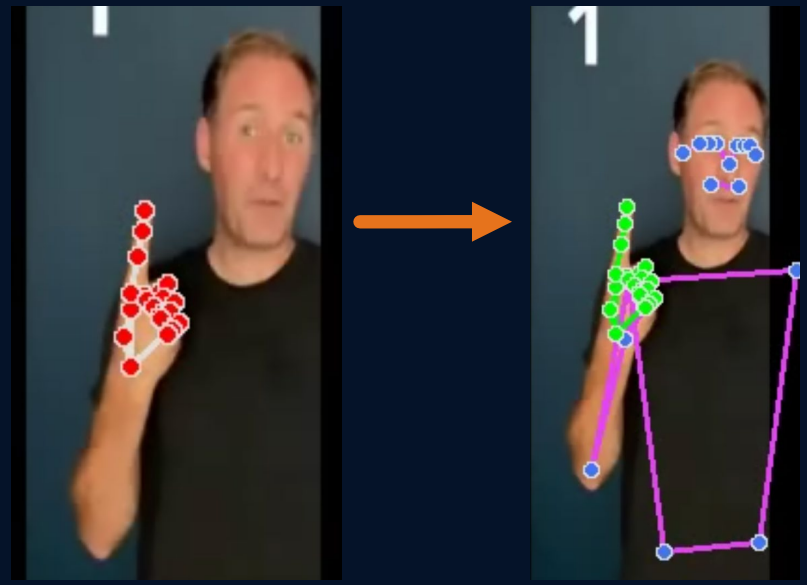
	ASL	CSL
Loss	0.070	0.228
Accuracy	96.08%	91.76%



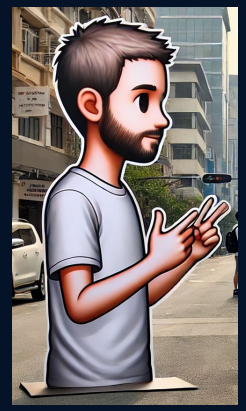


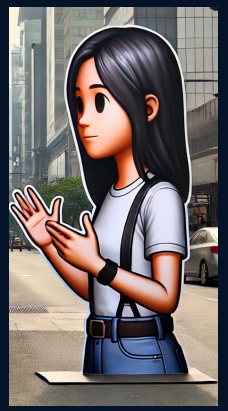
Work-in-progess

- From one hand to the entire upper body
- Change from CNN to Transformer Model
- Implement an AR Sign Language Avatar



Our Vision





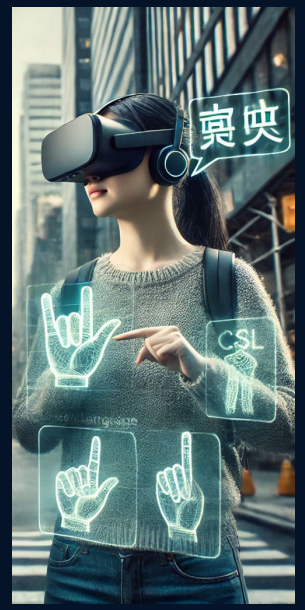


Image illustration by AI