Learning on the Rings: Self-Supervised 3D Finger Motion Tracking Using Wearable Sensors

Hao Zhou, Taiting Lu*, Yilin Liu, Shijia Zhang, Mahanth Gowda*

The Pennsylvania State University

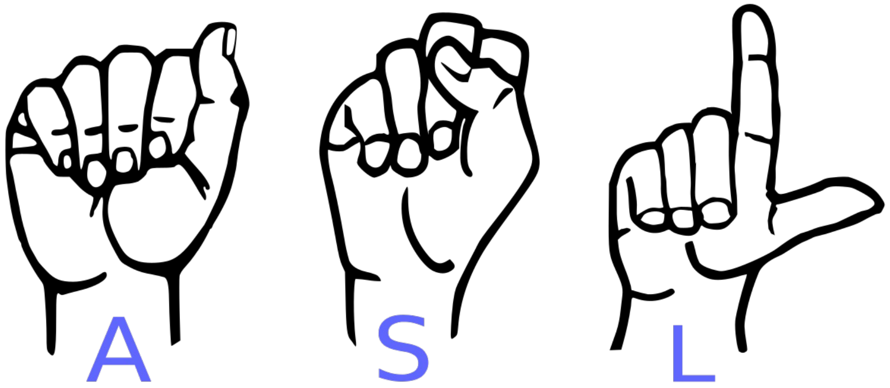
September 2022

PennState

2022UBICOMP

Motivation - Application Study

- VR/AR
  - Enabling finer-grained hand control
- Sign Language Recognition and Translation
  - Bridging deaf and hearing community



Potential Solutions & Problems

Cameras (Vision)

- Well-built & high-quality datasets
- Sensitive to occlusions/lights
- Privacy issues
- Portability Issue

Wearables (e.g., IMUs, Wrist bands)

- Lack of datasets
- Not limited to occlusions/lights
- Less concerned on privacy issues
- Ubiquitous

Credits: https://hackmag.com/coding/lets-code-for-leap-motion/ and https://www.wired.com/review/tap-strap-2/

ssLOTR

self-supervised
Learning On The Rings

A self-supervised learning finger motion tracking aided framework using IMU sensors

Main challenges & solutions

- Labeled IMU data expensive to collect

→Self-supervised learning for effective representation learning

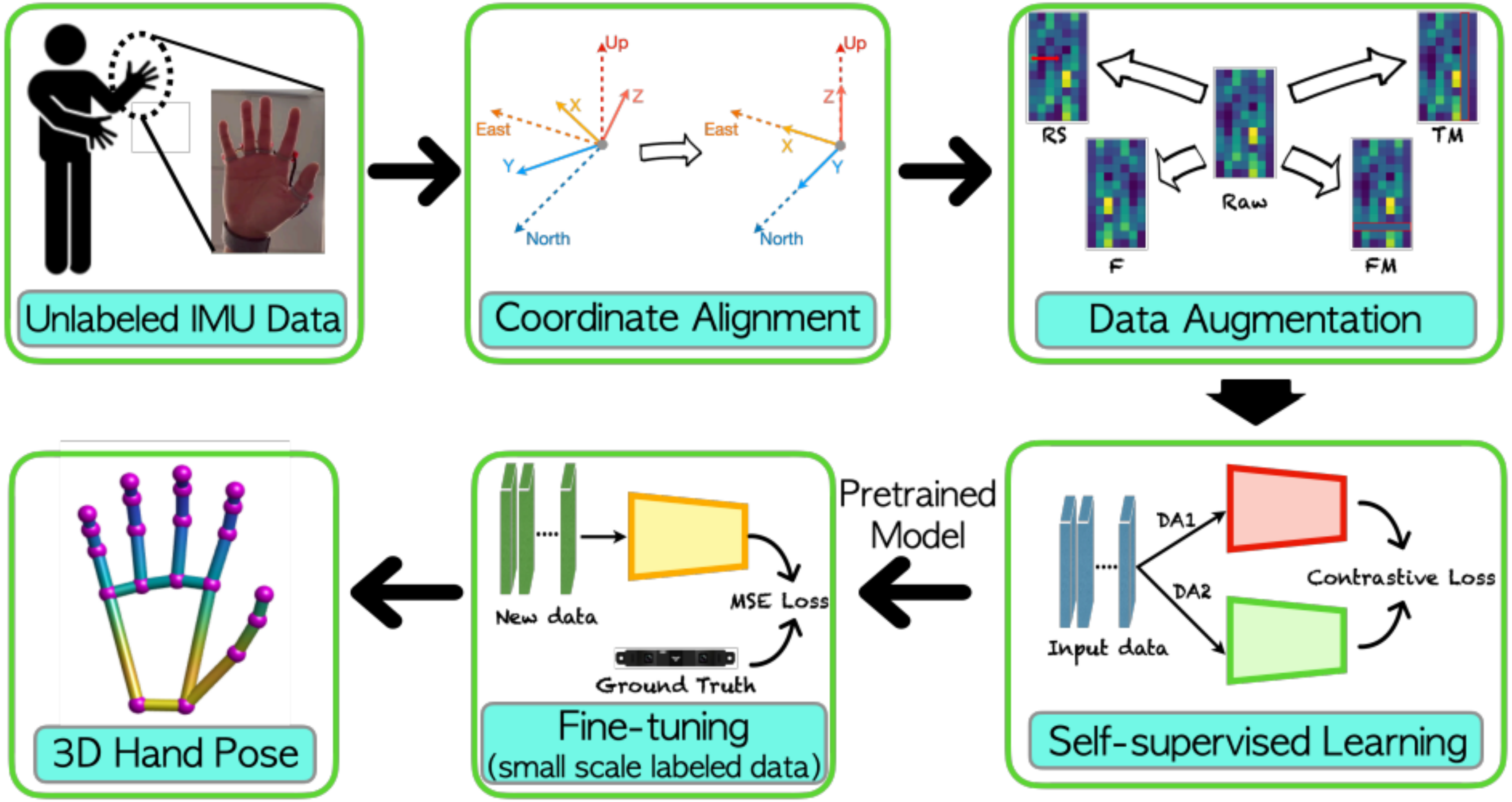
- Sensor data diversity across users, wearing locations, etc.

→ contrastive learning along with data augmentations

- Commercial products are close-source & no access to raw sensor data

→ Develop our own evaluation platform/prototype

ssLOTR: Overall Workflow



ssLOTR: Coordinate Alignment

1. Preliminary study shows same motion from different wrist orientations results in different sensor readings.
   a. Sensor readings from local frames

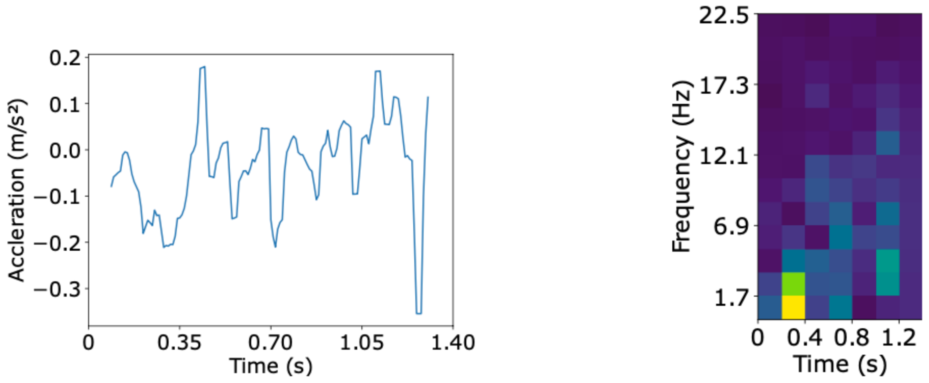
2. Transforming data to a consistent frame, e.g., from Local CF to Global CF, and finally to Wrist CF

$$\begin{bmatrix} X_{wcf} & Y_{wcf} & Z_{wcf} \end{bmatrix} = \begin{bmatrix} X_l & Y_l & Z_l \end{bmatrix} R_{finger} R_{wrist}^{T}$$

PennState

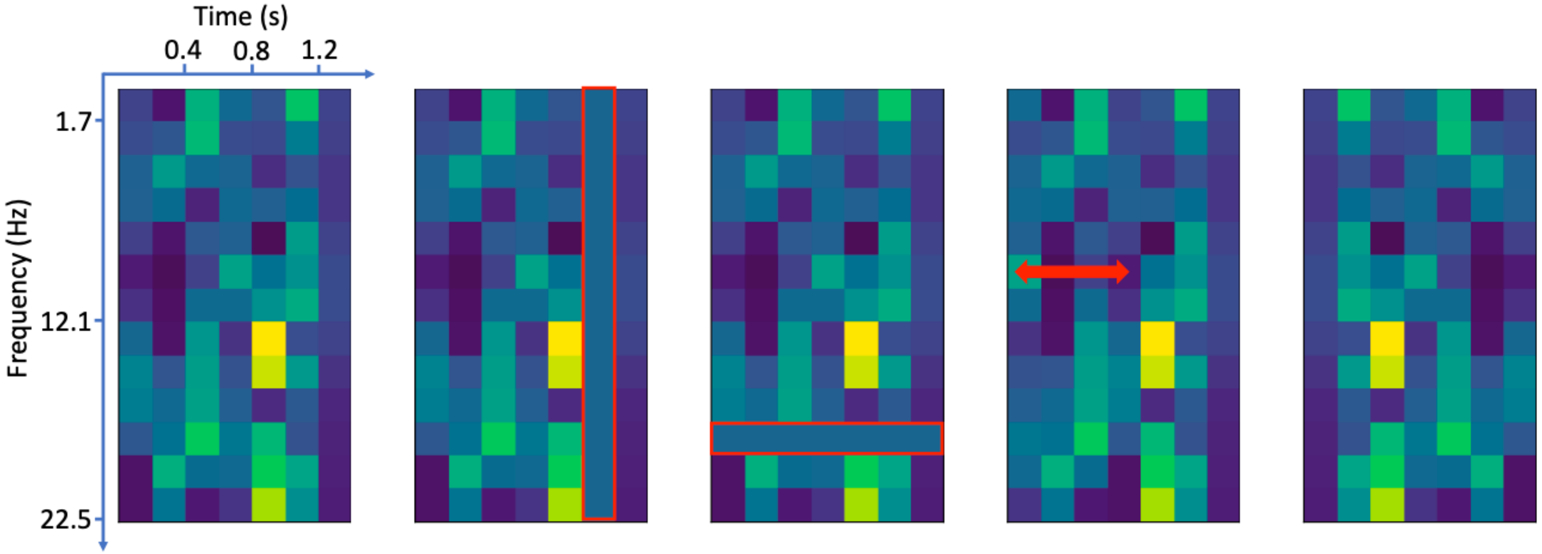
ssLOTR: STFT and Data Augmentation

- Short-Time Fourier Transform for capturing both time and frequency domain information.



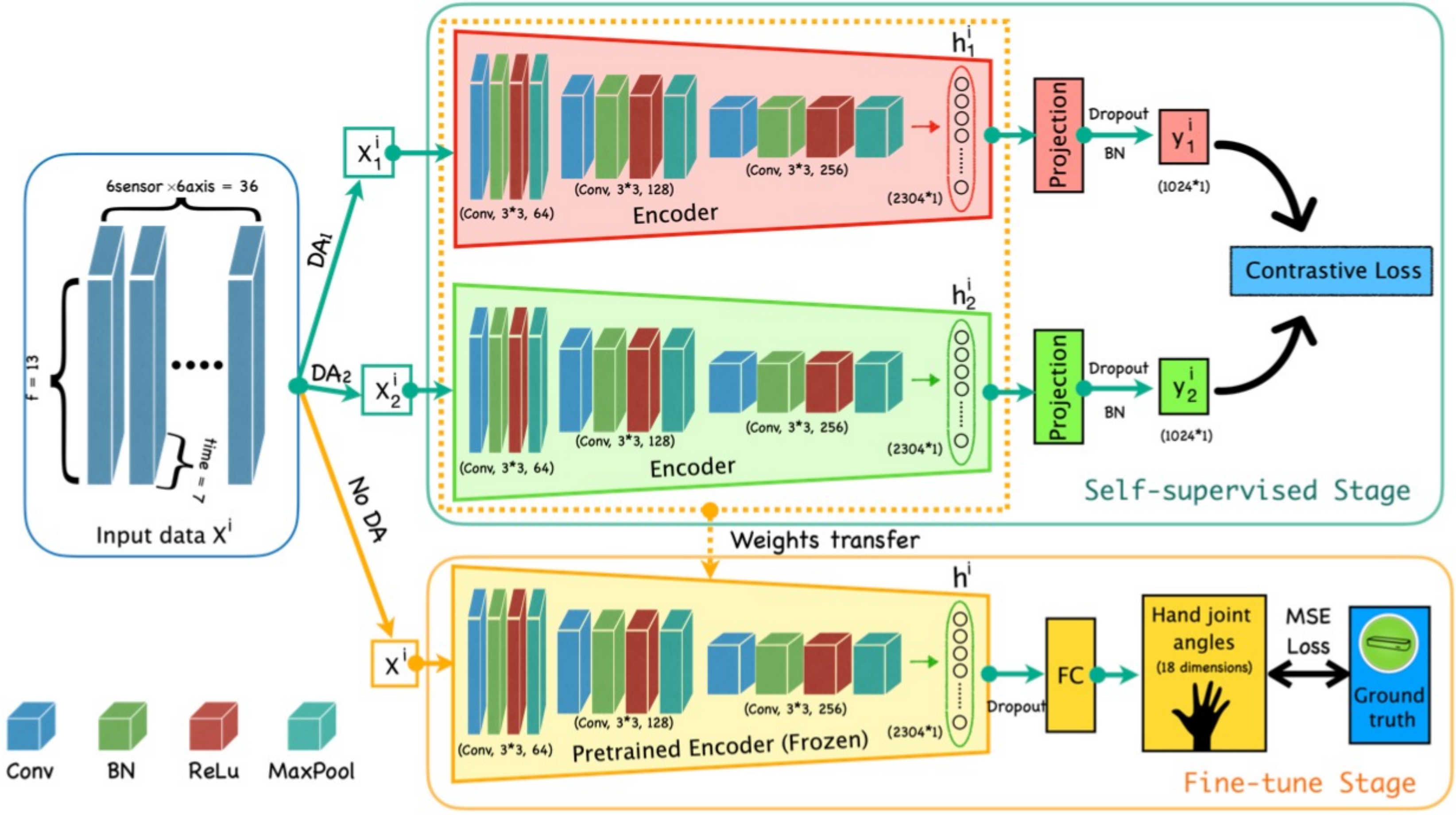
(a) Time domain data (b) Spectrogram

- Four data augmentations designed for contrastive learning.

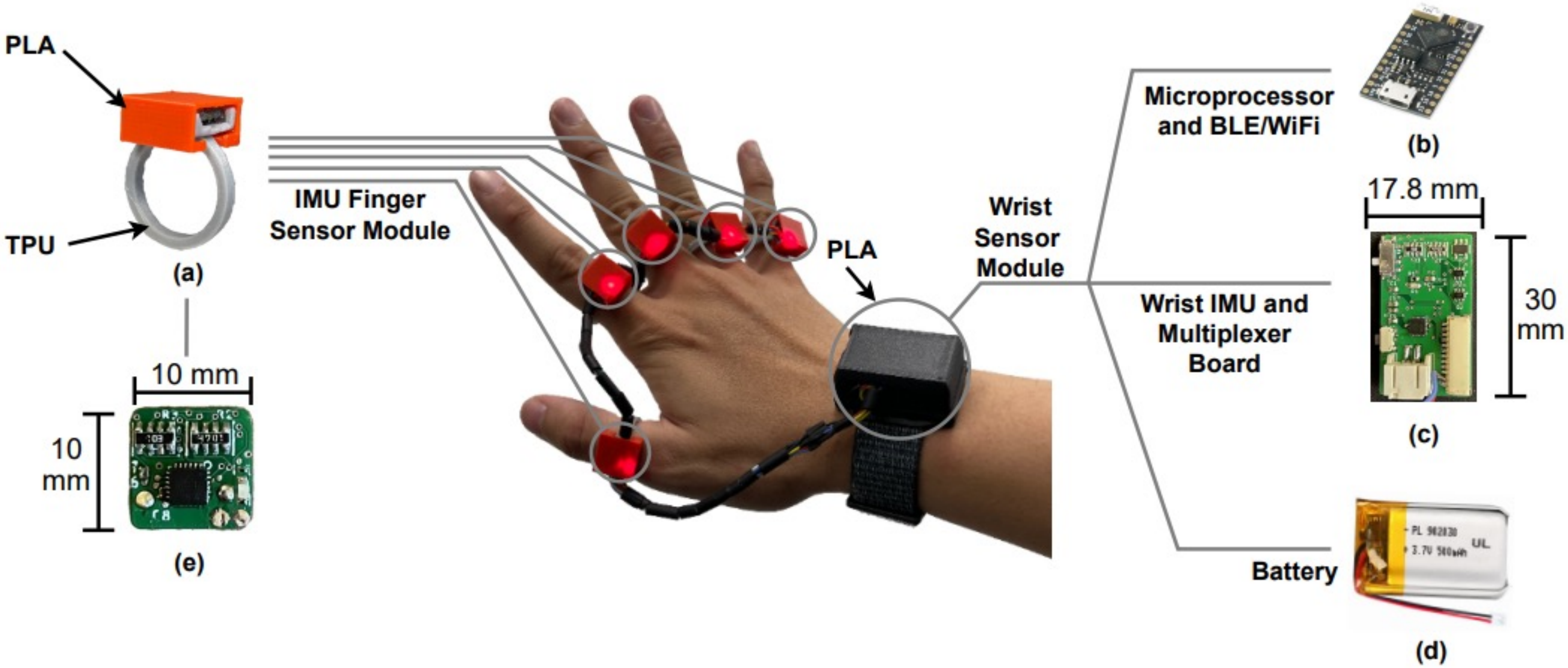


(a) Raw (b) Time Masking (c) Frequency Masking (d) Random Swapping (e) Flipping

ssLOTR: Self-supervised learning framework at the different stages.

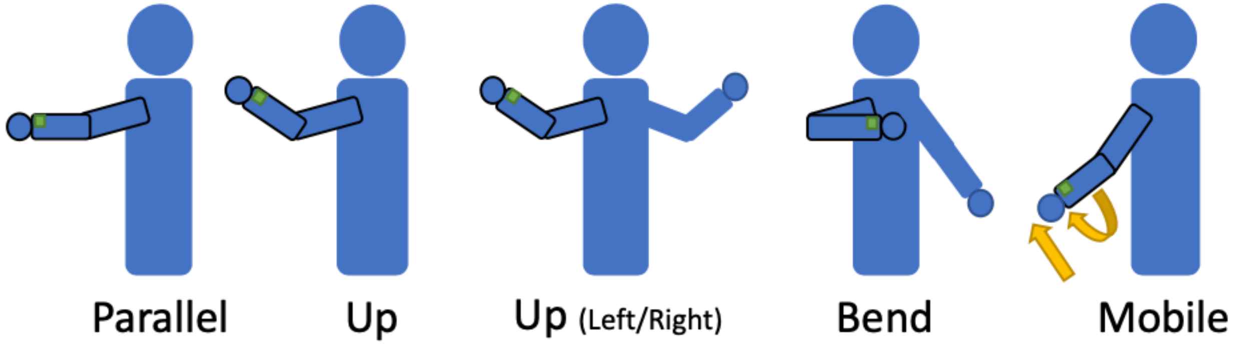


ssLOTR: Hardware



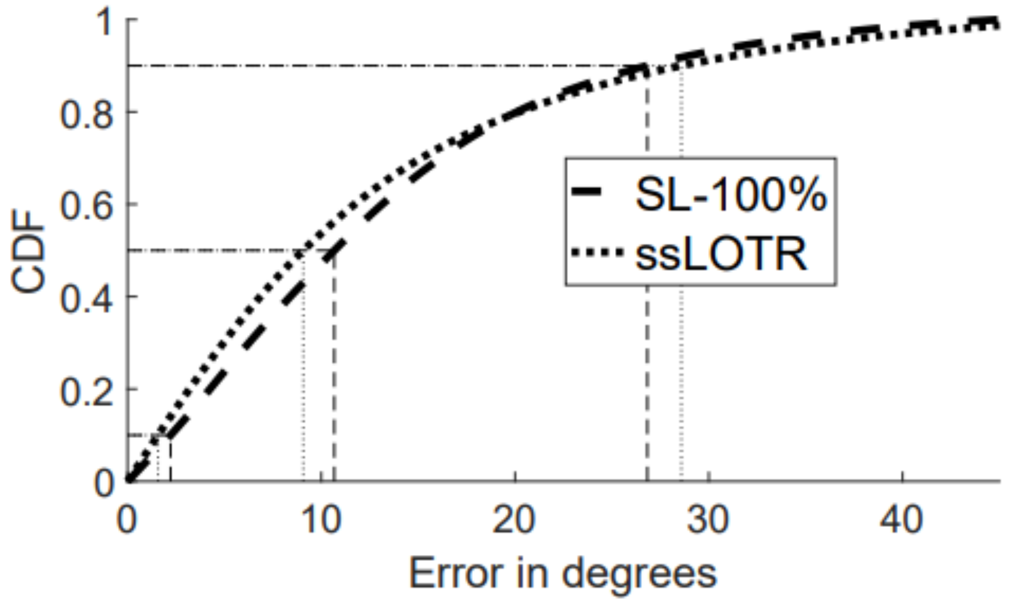
Data Collection

- 12 users (8 males and 4 females)
- Leap motion as Ground Truth
- 5 sessions, 2 minutes each

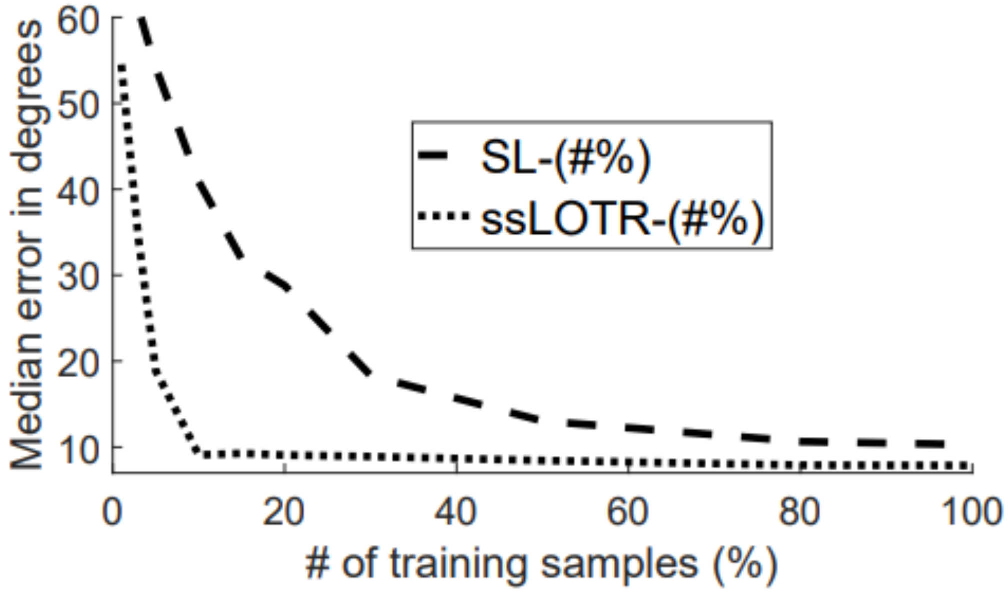


Results

- Tracking errors: 9.07 degrees and 6.55 mm
- Only 15% real data needed to finetune the model



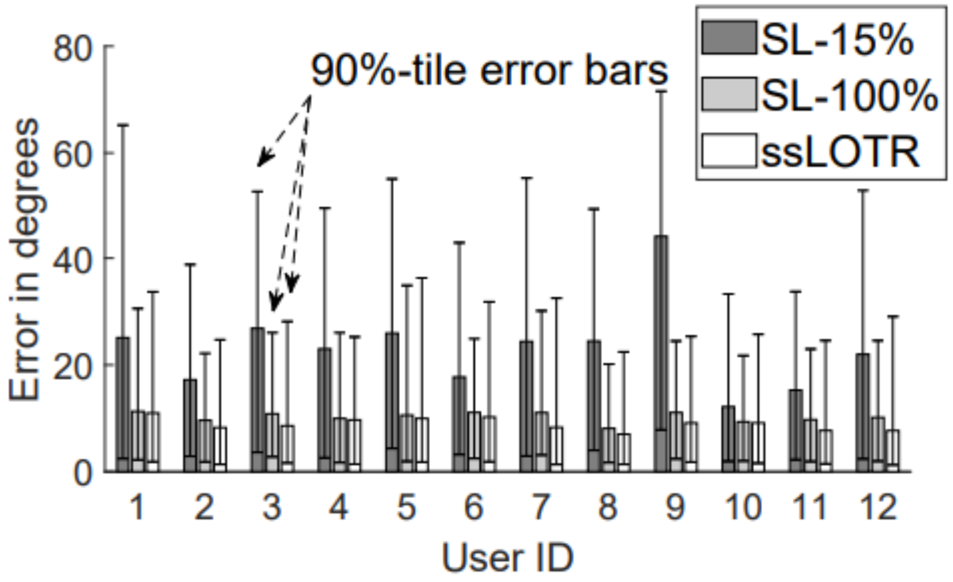
(a)



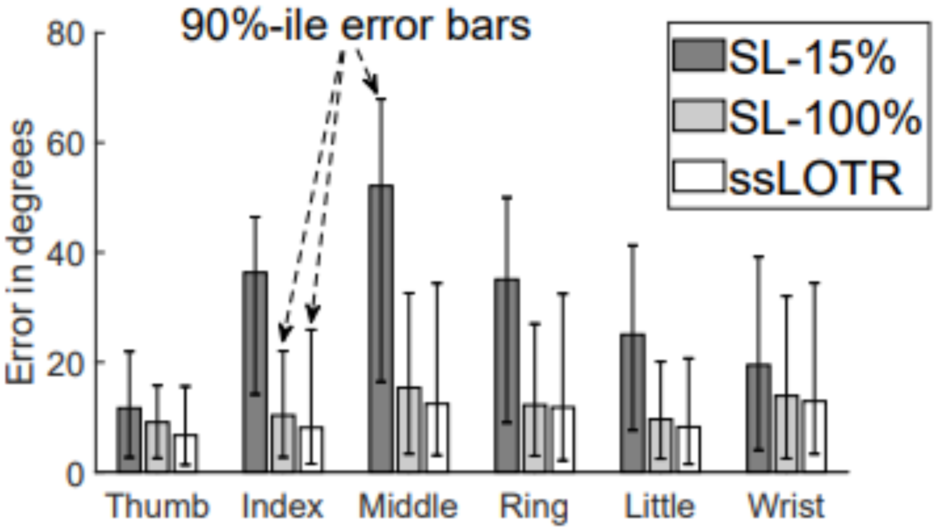
(b)

Results

- ssLOTR is stable across users and fingers (wrist)



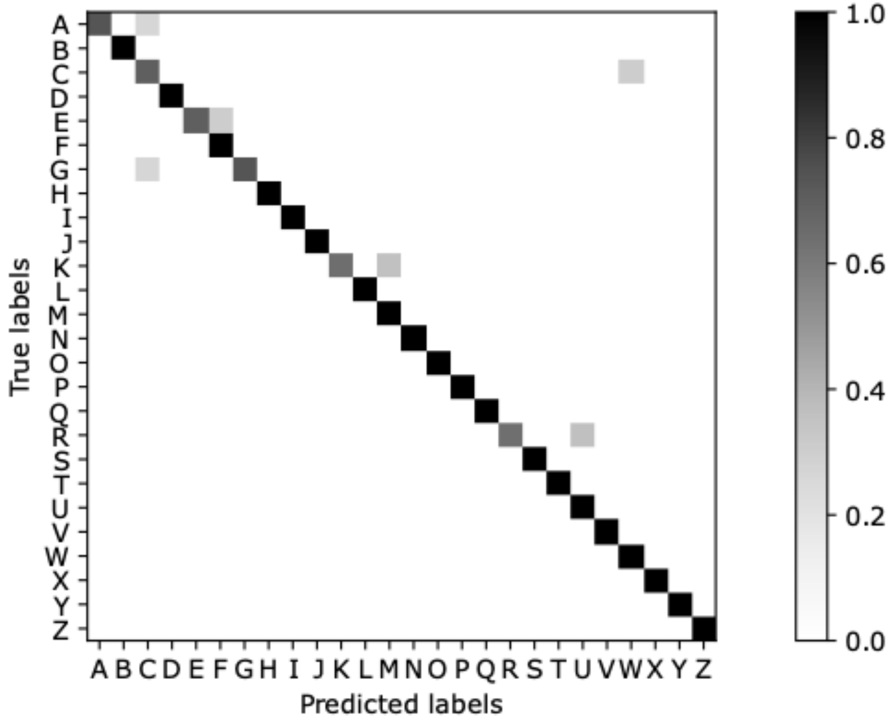
(a)

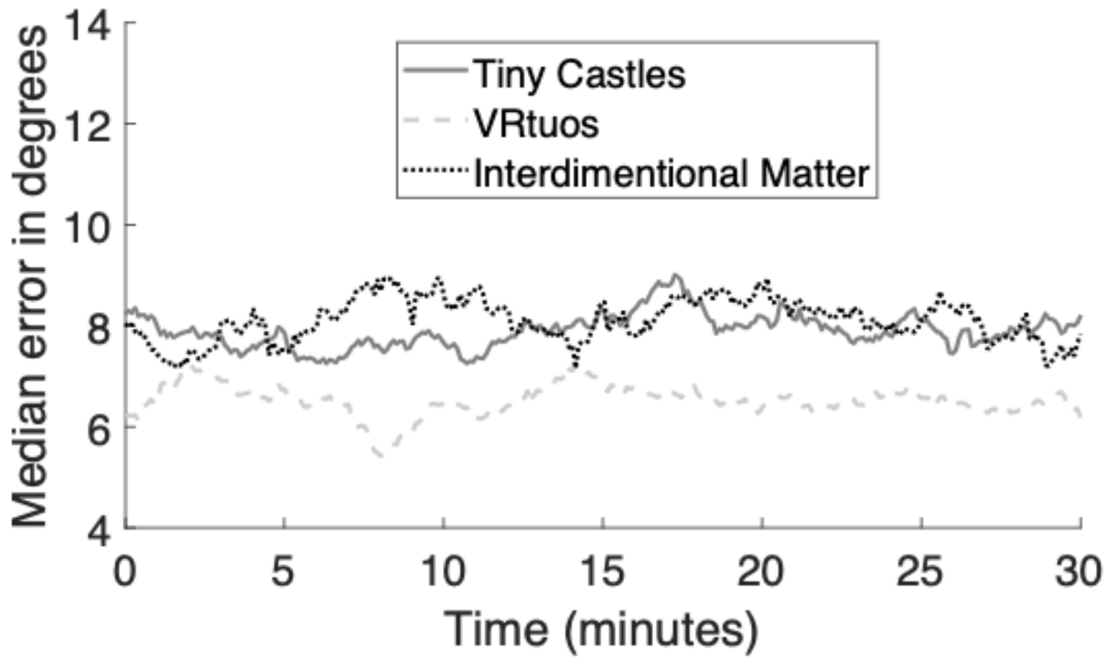


(c)

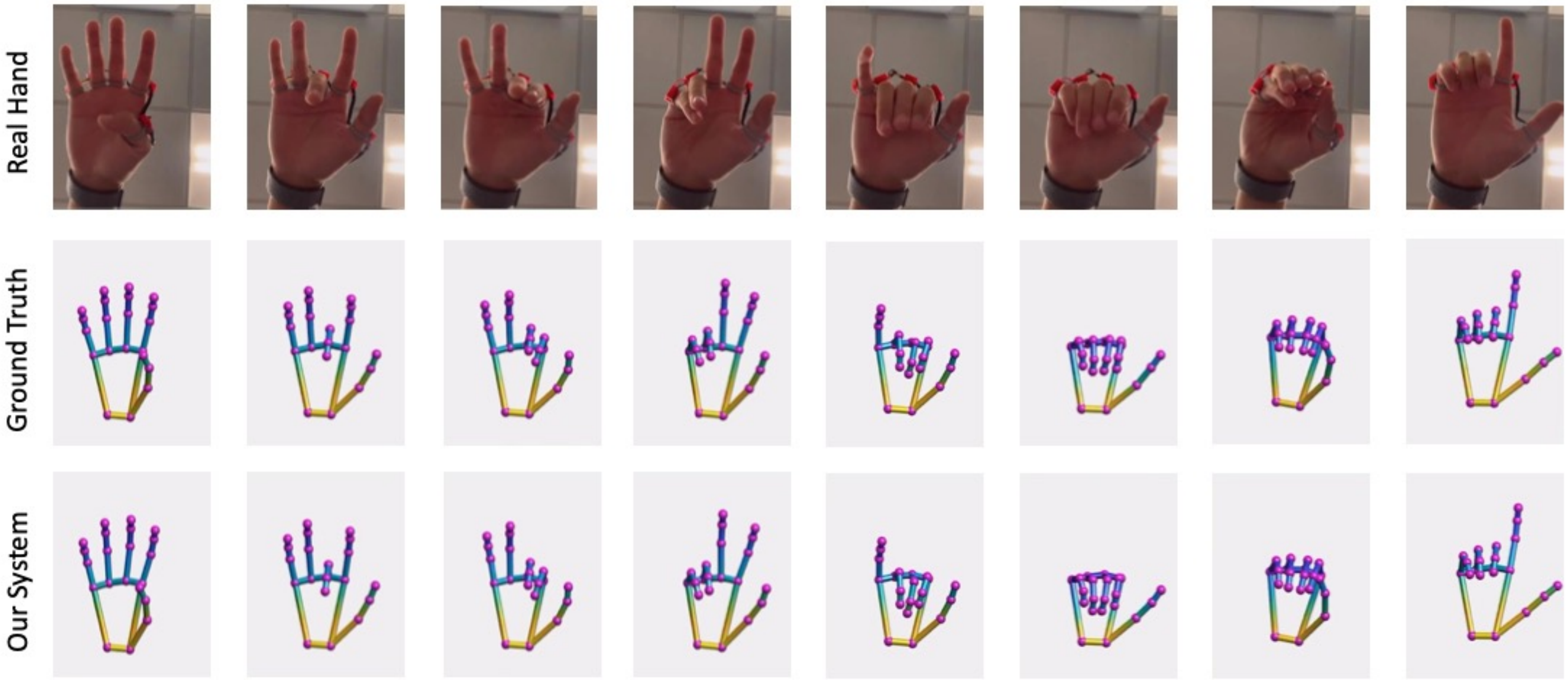
Results

- ssLOTR for real-world applications
  - ASL characters recognition (left) and VR games (right)





Qualitative Results



Discussion, Limitation and Future Work

- Form-factor improvement
- Lacking the ability to automatically learn data distribution from
  - Finger motion speed
  - Sensor noisy inputs
- Necessary Preprocessing
  - Preprocess needed (e.g., WCF transformation, STFT, etc.)
- Human body pose detection using wearables

Conclusion

- We present ssLOTR as the first self-supervised learning framework 3D finger motion tracking using IMUs.
- We also design an evaluation platform for efficient sensing and comfortable wearing that enables dexterous motion of fingers.

Thank you

&

Questions

PennState