# Upper-Limb Pose Prediction using Six-Axis Wrist-Worn IMUs

Abhay, Jeongah, Patrick

CS690R May 6th, 2025

# Motivation

- Growing demand for wearable-based human pose estimation
- Wrist-worn IMUs are lightweight, affordable, and practical
- Challenge: Inferring full arm motion from limited IMU input
- Goal: Predict upper-limb skeletal motion using only 6-axis wrist IMU

# Project Overview

- Predict 3D joint coordinates from wrist-worn IMU data
- Inputs: Accelerometer + Gyroscope (6-axis)

| Sensor Type | Axis X | Axis Y | Axis Z |
|---|---|---|---|
| Accelerometer | *ax* | *ay* | *az* |
| Gyroscope | *gx* | *gy* | *gz* |

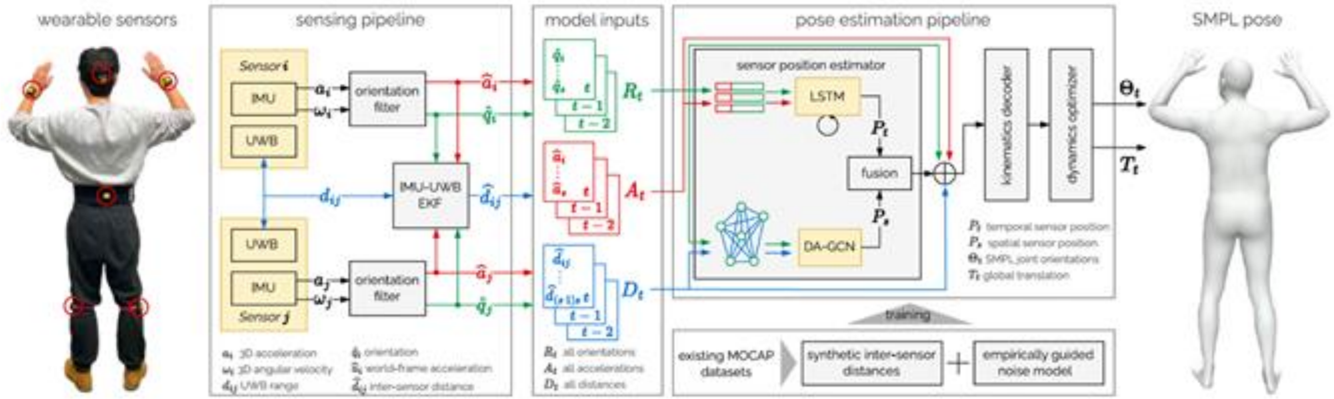
- Outputs: 3D skeletal positions of upper limb

  - Shoulder, Elbow, Wrist, Finger (Left/Right)

- Evaluation: Compare predicted vs ground-truth MoCap positions

# Related works

## Ultra Inertial Poser: Scalable Motion Capture and Tracking from Sparse IMU and UWB

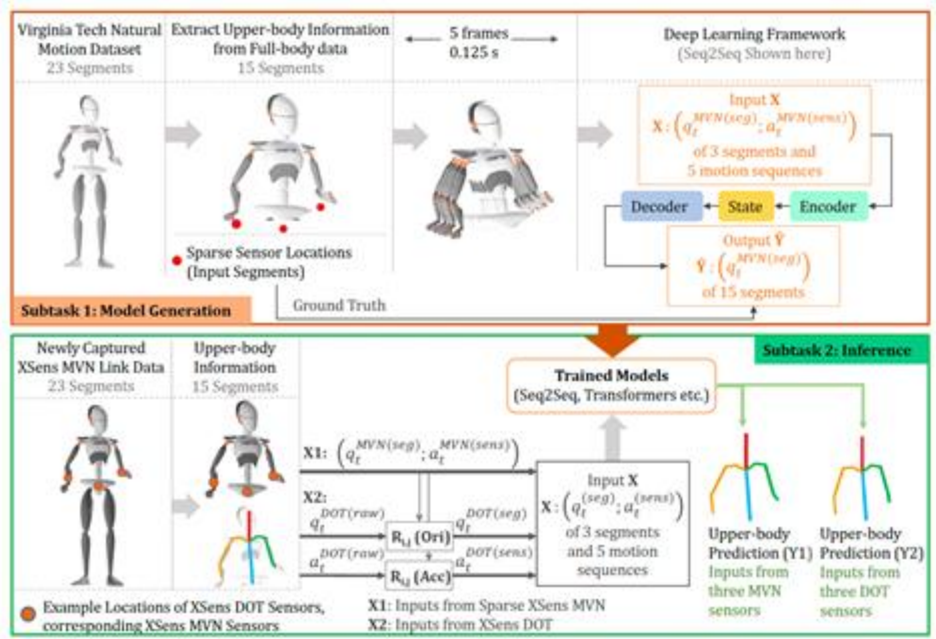



- This paper uses **sparse IMU + UWB sensors** to predict **full-body pose**.
- Our project is different: we use **only two wrist-worn IMUs** to predict **only the upper-limb** pose.
- **Takeway**: They showed that even with sparse sensors, accurate pose prediction is possible by using **graph-based fusion of signals. ->** We can apply **temporal sequence models (LSTM)** to capture motion over time.

Rayan A. et al. Ultra Inertial Poser: Scalable Motion Capture and Tracking from Sparse Inertial Sensors and Ultra-Wideband Ranging. https://arxiv.org/abs/2404.19541

# Related works

## Capturing Upper Body Kinematics and Localization with Low-Cost Wearable Sensors

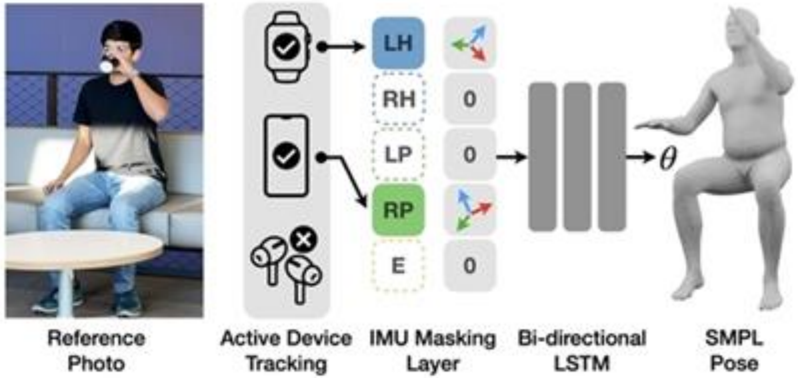



- This paper uses **three IMUs (wrists + pelvis)** to predict **upper-body kinematics**.
- Our project is even more challenging: **two wrist IMU**, no pelvis sensor.
- **Takeaway**: Their use of a **Seq2Seq deep learning model** showed that even sparse IMUs can predict upper-limb motion well. -> We again saw the importance of **temporal sequence modeling** with LSTM and similar models

Sarker, A. et al. Capturing Upper Body Kinematics and Localization with Low-Cost Sensors for Rehabilitation Applications. Sensors 2022, 22, 2300.

# Related works

IMUPoser: Full-Body Pose Estimation using IMUs in Phones, Watches, and Earbuds




- IMUPoser demonstrates **real-time full-body pose estimation** from only consumer IMUs (phones, watches, earbuds) using **a two-layer bi-LSTM** and a brief inverse-kinematics **(IK)** refinement.
- IMUPoser relies on sensor-provided **global orientations** with an **LSTM+IK pipeline** and active device-tracking, while our approach sensor data in **local orientations**, then uses a **Bi-LSTM** for end-to-end pose regression with **biomechanical constraints**.

Vimal Mollyn et al. 2023. IMUPoser: Full-Body Pose Estimation using IMUs in Phones, Watches, and Earbuds. In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23). Association for Computing Machinery, New York, NY, USA, Article 529, 1–12

# Project Objectives

**Goal:**
Apply and experiment with methodologies learned in class to predict upper-limb skeletal motion using only wrist-worn 6-axis IMUs.

**Key Investigation Points:**

- What types of **models** (e.g., LSTM, BiLSTM) best capture upper-limb motion from IMU data?
- How much impact does **IMU preprocessing** (e.g., filtering, orientation estimation) have on model performance?
- Can introducing **biomechanical constraints** improve an accuracy of predicted poses?

# Dataset

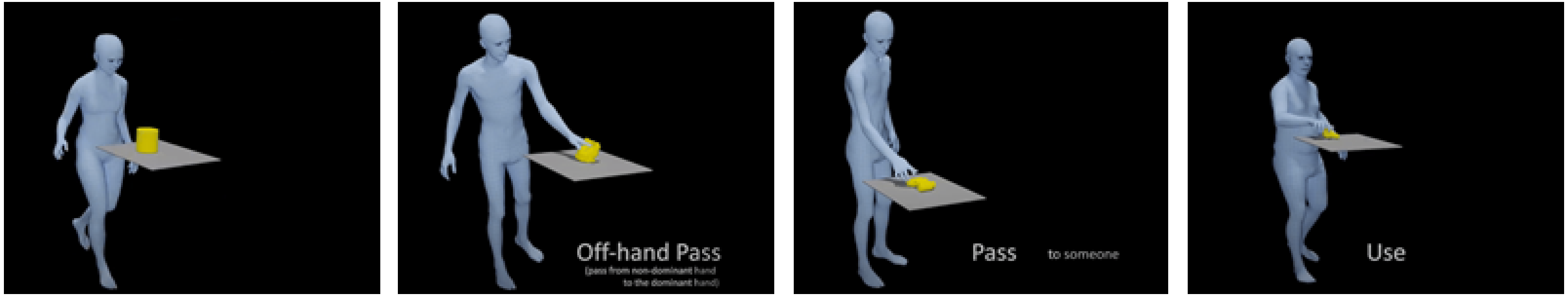
AMASS: 300 subjects, 40+ hours of motion capture

GRAB: Subset of AMASS. Detailed hand-object interaction dataset

- 10 subjects interact with 51 everyday objects
- Activities: Lift, pass, off-hand pass, and use

Data details:

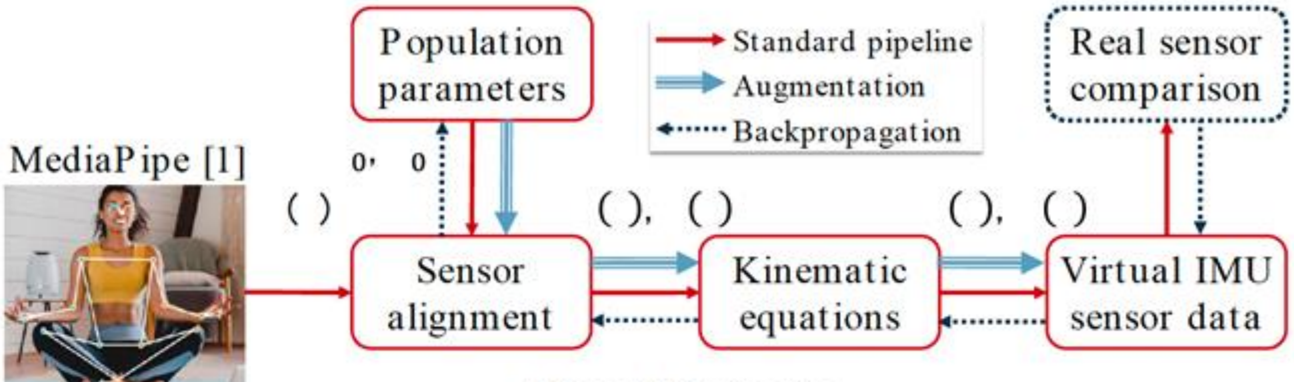
- Inertial Data (Accel, gyro): Two IMUs, which each is located on a forearm
- MoCap: shoulders, elbows, wrists, and thumbs




Taheri, Omid, et al. "GRAB: A dataset of whole-body human grasping of objects." *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part IV 16*. Springer International Publishing, 2020.

# IMU Data Generation

- Use Virtual IMU to synthesize IMU data from MoCap
- Bypassed MediaPipe (red path in pipeline)
- Backprop-based sensor alignment (optional)




VirtualIMU pipeline

# Data preprocessing

- Filtering:
  - Reduce high-level noises and drift in IMU acceleration and gyroscope data
  - Using low-pass and band-pass filters (0.1 - 20 Hz)

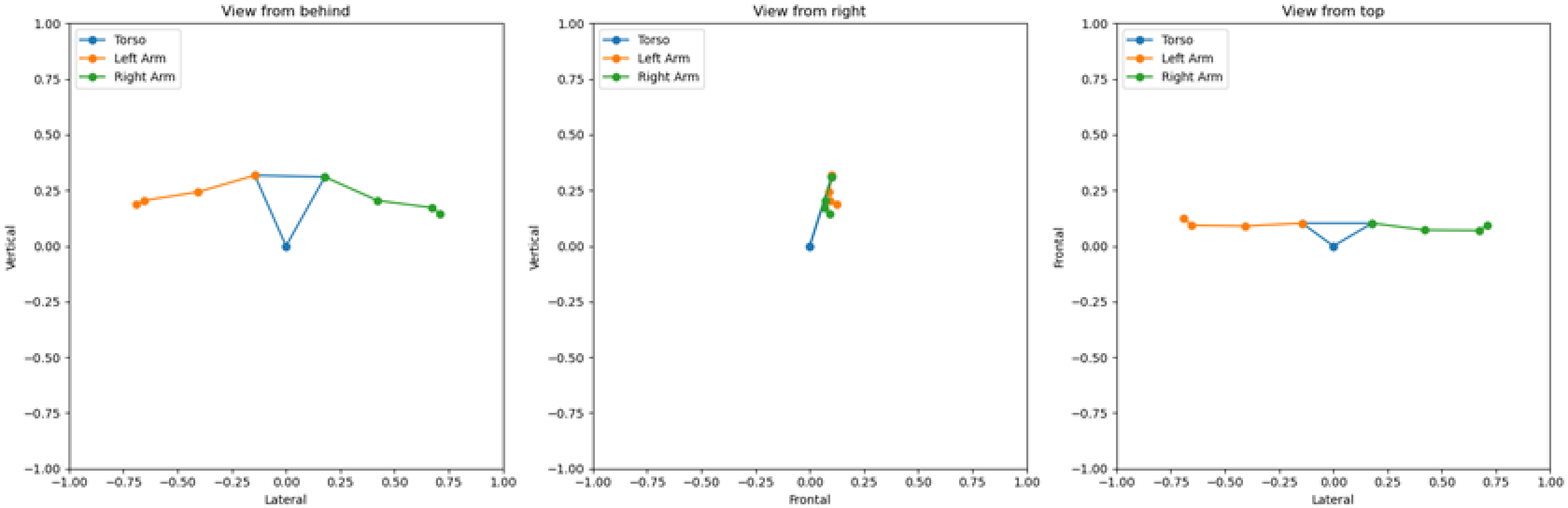
- Estimate sensor orientation and transform acceleration and angular velocity into the global coordinate system (Task 2.6 Assignment 1)
  - Steps:
    - Estimate orientation using AQUA algorithm.
    - Rotate local IMU data to global coordinates.
    - Remove gravity from the Z-axis.

- Data Segmentation:
  - Sliding window approach with window size of 1 second and overlap of 75%

# Example

Ground-truth Pose (MoCap)

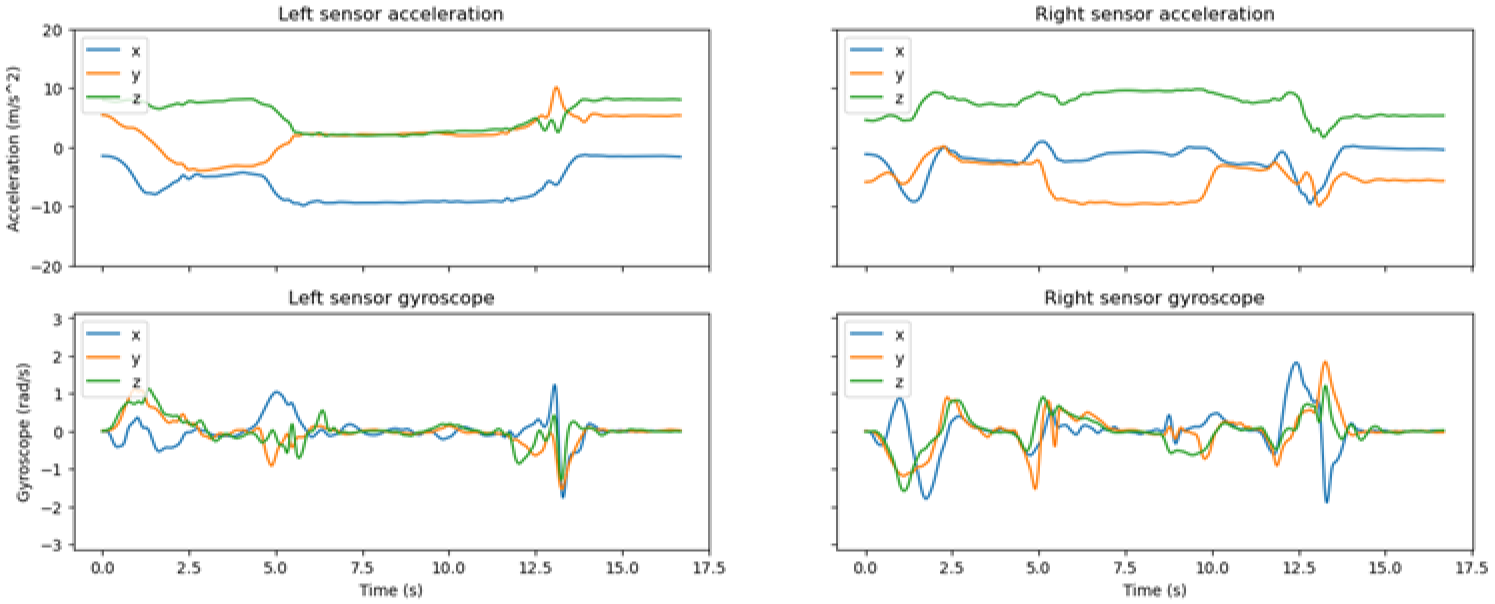
- Directly from GRAB dataset



# Example

Input IMU (Sensor data)

- Generated from GRAB Mocap via Virtual-IMU



# Pipeline

Joint relative statistical dependencies

MoCap

IMU

Input: 6-ax IMU

IMU to upper-limb pose estimator

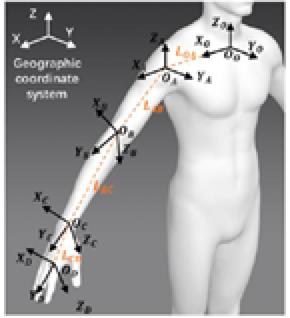
Biomechanical constraints

(1) Joint angle constraint loss

(2) Bone length consistency loss

Output: 3D joint coord.

Ground truth: 3D joint coord.




Losses examples:

- Mean per joint position error
- Mean per joint velocity error

# Biomechanical constraints

Vetrice, Georgiana & Deaconescu, Andrea. (2017). Development of elbow rehabilitation equipment using pneumatic muscles. MATEC Web of Conferences

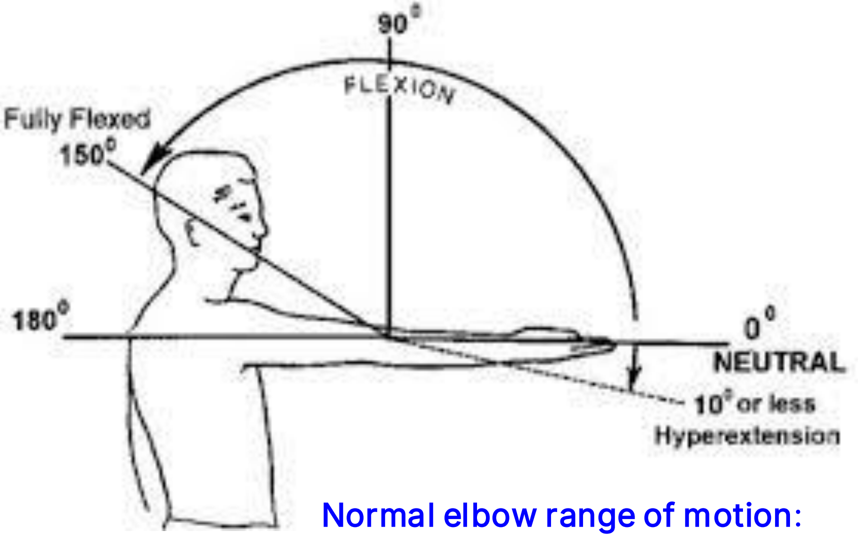
## (1) Joint angle constraint loss

### What does it check?

- **elbow joint flexion angle** — the angle formed at the elbow between the upper arm and forearm

### How is it calculated?

- Use shoulder → elbow → wrist to compute the joint angle
- Apply penalty if angle is outside the valid range
- Average penalties across batch and time steps




**Normal elbow range of motion:**

- Full extension: ~0°
- Full flexion: ~150°

# Biomechanical constraints

Image source: https://basicmedicalkey.com/upper-limb/

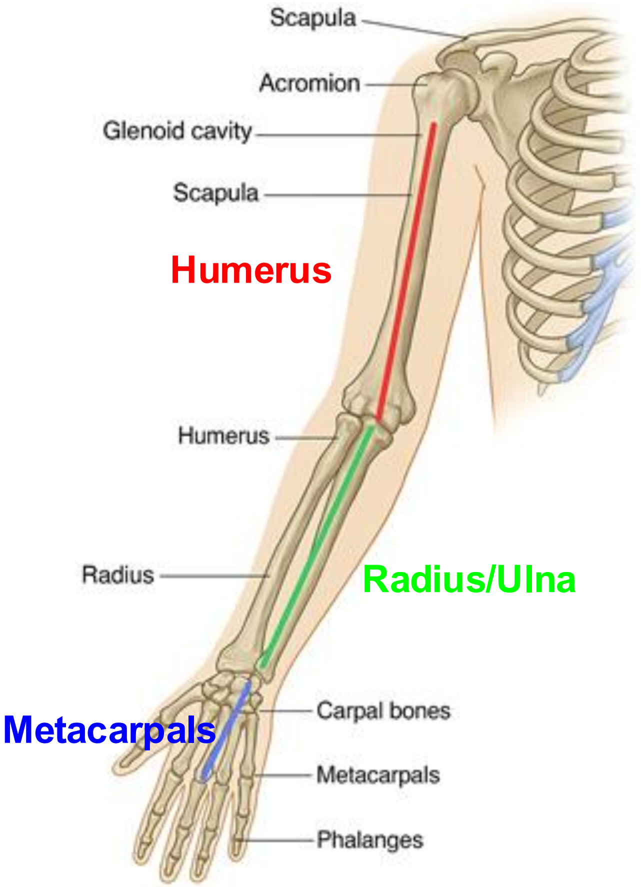
## (2) Bone length consistency loss

📌 **What does it check?**

- **lengths of the bones**

❓**How is it calculated?**

- For each bone, the length is computed using the formula `length = sqrt(sum(vector^2))`
- Measure variance over time
- Apply separately to upper arm, forearm, and hand, for both the left and right arms
- Average all variances as the final loss



# Biomechanical constraints

## (1) Joint angle constraint loss

- Joint-Angle Coordination Patterns Ensure Stabilization of a Body-Tool System (2016, Frontiers in Psychology, Van der Steen et al)
- 3D human pose estimation based on 2D–3D consistency with synchronized adversarial training (2024, Robotics and Autonomous Systems, Yicheng Deng et al)
- Human Joint Angle Estimation Using Deep Learning-Based Three-Dimensional Human Pose Estimation for Application in a Real Environment (2024, MDPI, Jin-Young Choi et al)

## (2) Bone length consistency loss

- Motion Projection Consistency Based 3D Human Pose Estimation with Virtual Bones from Monocular Videos (2022, IEEE transaction on cognitive and developmental systems, Guangming Wang et al)
- A Geometry Loss Combination for 3D Human Pose Estimation (2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), Ai Matsune et al)
- BLAPose: Enhancing 3D Human Pose Estimation with Bone Length Adjustment (2024, Chih-Hsiang Hsu et al, 2024)

# Loss Functions

Without biomedical constraints:

$$Loss = MSE$$

With biomedical constraints:

$$Loss = \lambda_1 MSE + \lambda_2 Loss_{angle} + \lambda_3 Loss_{bone}$$

$MSE$: Mean Square Error
$Loss_{angle}$: Joint angle constraint loss
$Loss_{bone}$: Bone length consistency loss
$\lambda_1, \lambda_2, \lambda_3$: Weights

# Model Architectures

**LSTM** (Baseline)

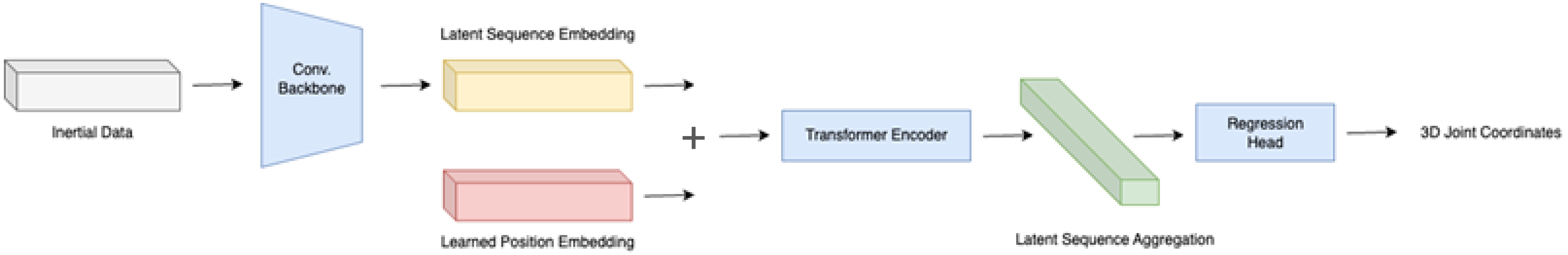
- 2-layer, unidirectional
- Models short-range dynamics with low latency and modest compute

**Bidirectional LSTM**

- 2-layer, processes sequences forward + backward
- Captures both past and future temporal dependencies

**ConvTransformer**

- Inspired by Shavit and Klein's 2021 work [1]
- Combines convolutional feature extraction with a Transformer encoder to model spatial-temporal patterns for 3D joint prediction




[1] Shavit, Y., & Klein, I. (2021). Boosting inertial-based human activity recognition with transformers. *IEEE Access*, *9*, 53540-53547.

# Evaluation Metrics

We evaluate the model using quantitative metrics:

**Mean Per Joint Position Error (MPJPE):**

- Measures the average Euclidean distance between predicted and ground truth joint positions.
- Lower MPJPE indicates higher pose accuracy.

$$MPJPE = \frac{1}{N_F} \cdot \frac{1}{N_J} \sum_{f,j} \|p_{f,j} - \hat{p}_{f,j}\|_2$$

**Where:**

- $N_F$ is the number of frames
- $N_J$ is the number of joints
- $p_{f,j}$ is the ground truth position of joint $j$ at frame $f$
- $\hat{p}_{f,j}$ is the predicted position

# Evaluation Metrics

We evaluate the model using quantitative metrics:

**Mean Per Joint Velocity Error (MPJVE):**

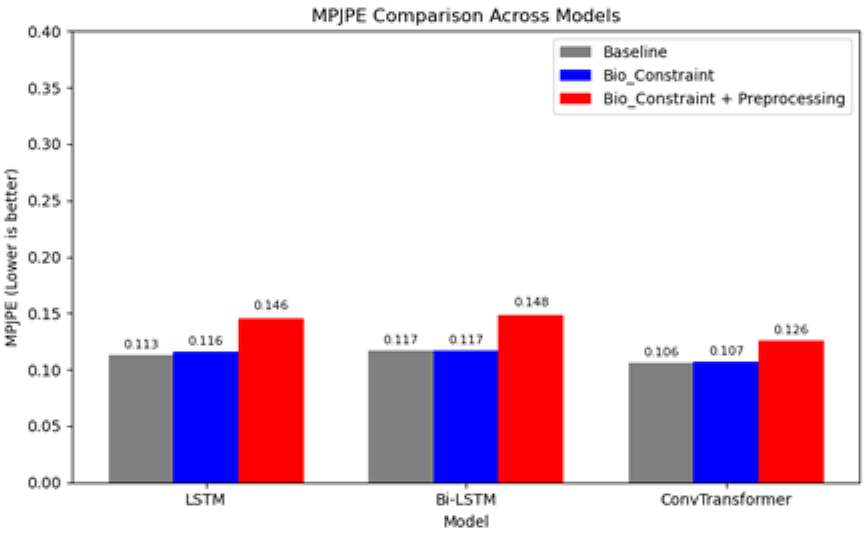
- Measures the average difference in joint velocities between predicted and ground truth over time.
- Captures temporal smoothness and motion consistency.

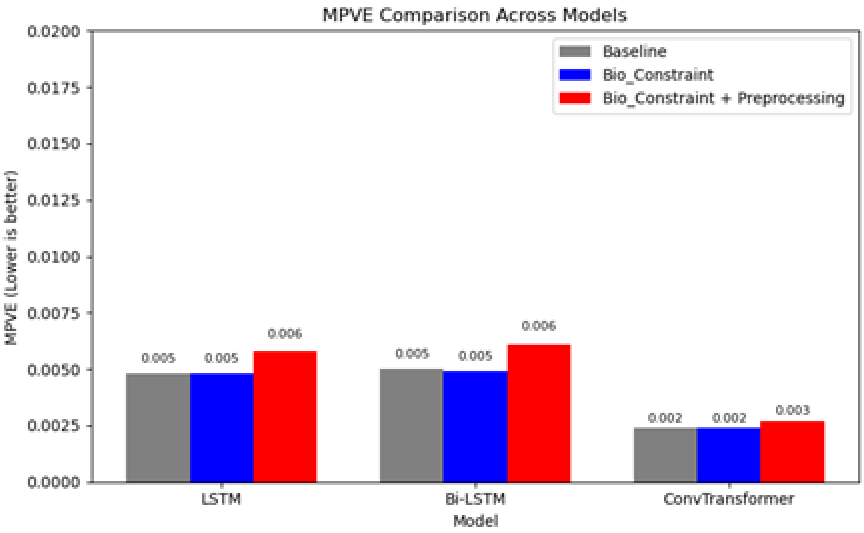
$$MPJVE = \frac{1}{N_F} \cdot \frac{1}{N_J} \sum_{f,j} \left\| v_{f,j} - \hat{v}_{f,j} \right\|_2$$

**Where:**

- $N_F$ is the number of frames
- $N_J$ is the number of joints
- $v_{f,j}$ is the ground truth velocity of joint $j$ at frame $f$
- $\hat{v}_{f,j}$ is the predicted velocity

# Result

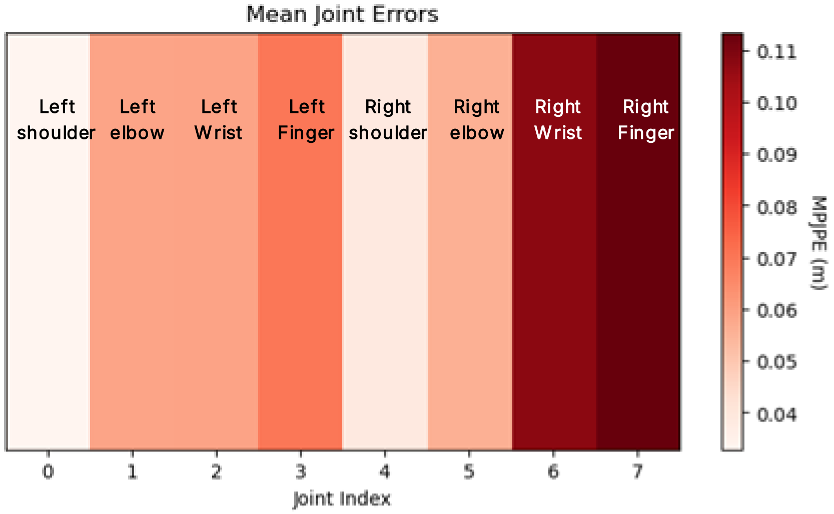


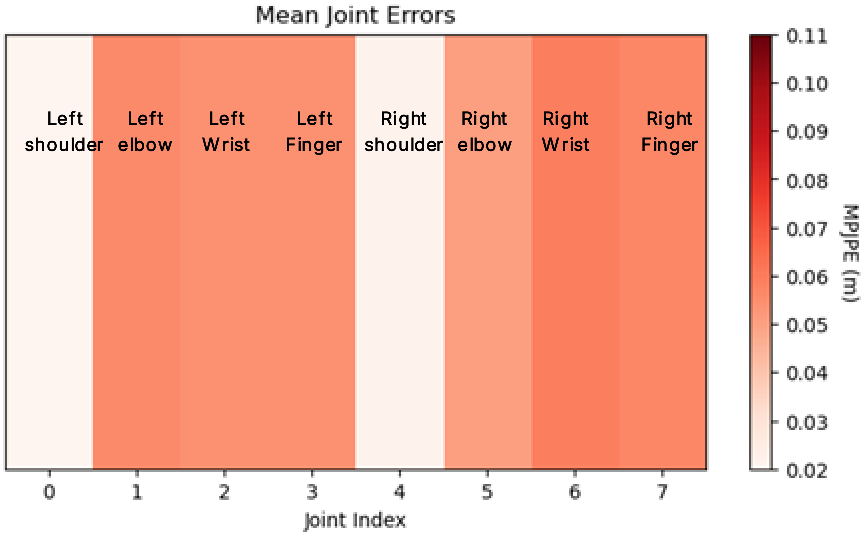

- **ConvTransformer (baseline)** delivers the best overall performance.
- **LSTM** shows the most significant performance degradation under the *Bio Constraint + Preprocessing* condition, followed by **Bi-LSTM.**
- **ConvTransformer** stayed **stable and robust** across all experiments.

# Joint-wise difficulty

LSTM (baseline)



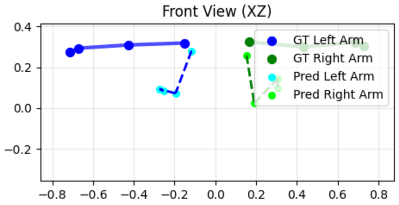

ConvTransformer (baseline)

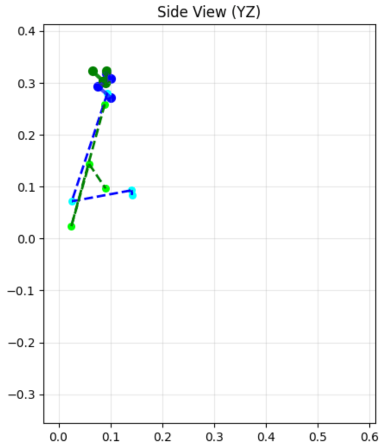


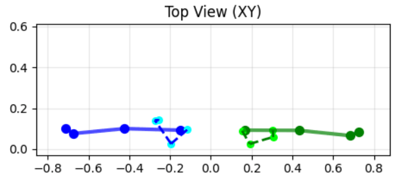

- **Overall error drop**: ConvTransformer roughly halves the worst-case errors (joints 6 & 7) and reduces average MPJPE across all joints.
- **Variance shrinks**: LSTM errors span ~0.03–0.115 m, whereas ConvTransformer compresses that range to ~0.02–0.057 m.

# Drinking water

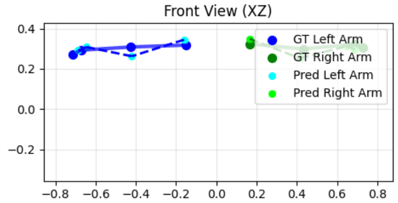
## LSTM (Baseline)

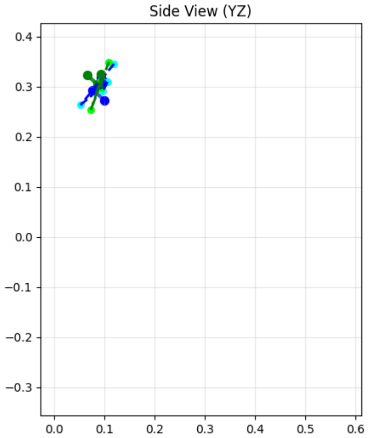


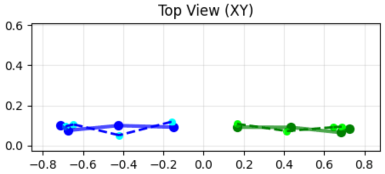




## ConvTransformer (with bioconstraint)

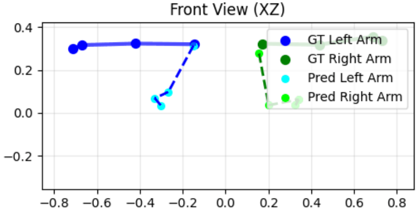


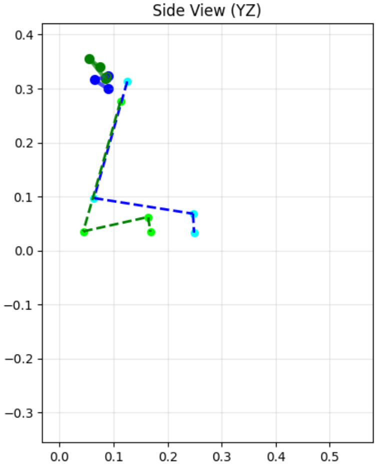



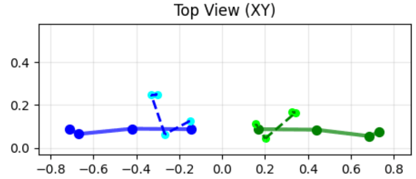



# Peel with a knife

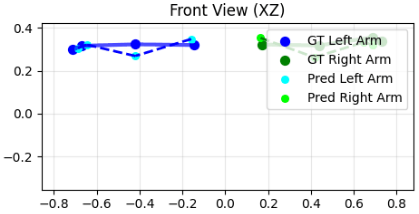
## LSTM (Baseline)

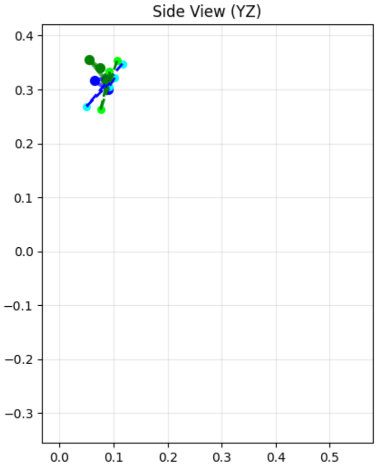


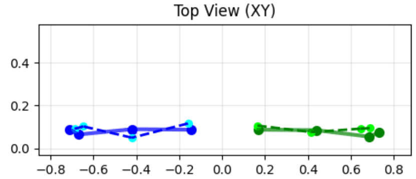




## ConvTransformer (with bioconstraint)

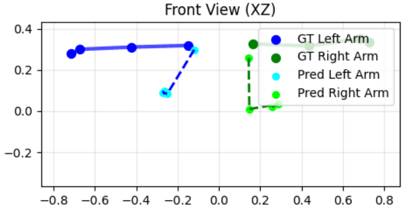


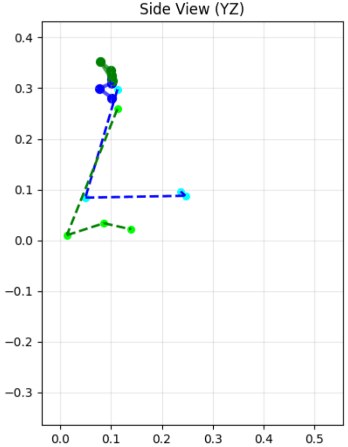



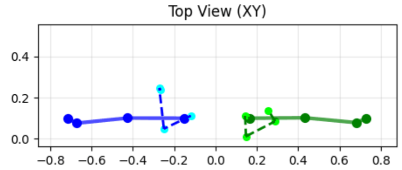



# Picking up a cup

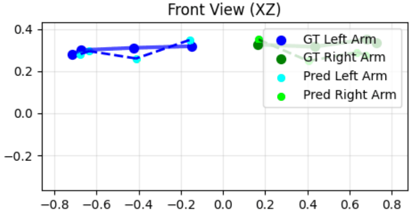
## LSTM (Baseline)

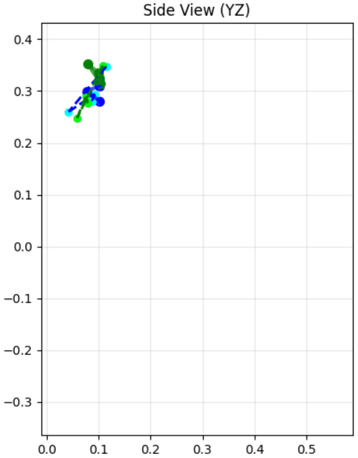


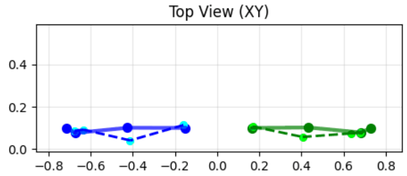




## ConvTransformer (with bioconstraint)



# Conclusion

- Model-Wise Performance Insights:
  - **LSTM** and **ConvTransformer** showed the best overall performance, achieving the lowest MPJPE and MPVE under the baseline (no constraint) condition.
  - The **LSTM** delivers nearly transformer-level accuracy with fewer parameters and much lower inference latency, making it the optimal choice for resource-constrained, real-time deployment.

- **Preprocessing steps mostly degraded** performance, likely due to **information loss**, or **sensitivity to transformed inputs**.

# Future work

- Domain gap between synthetic IMU and real IMU / Real-world testing with actual IMU
- Hyperparameter tuning
- Extending Joint Constraints to Multi-Axial Joints

Thank you!