Robot Arm Motion Generation by Representation Learning of IMU-captured Human Arm Motion

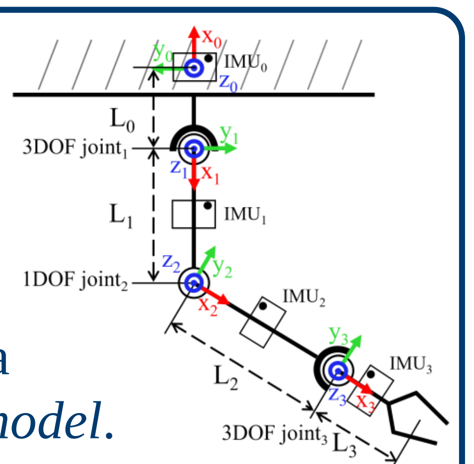
Jayjun Lee, Imperial College London (jayjun.lee19@imperial.ac.uk)

1. Objective

- Enable robot arm with the capability to generate human-like motion for a *drinking task*.
- Drinking task:
  1) Reaching; 2) Drinking; 3) Returning.
- A system for **capturing human arm motion**.
- A system for **representation learning the dynamics** of the human arm drinking motion.
- A system that utilises the learned representation to **generate human-like robot arm motion**.

2. Background

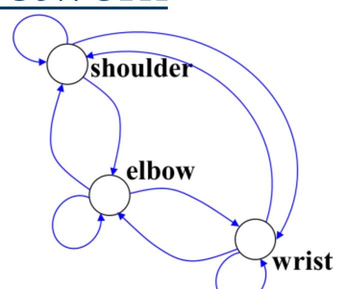
Modelling Human Arm



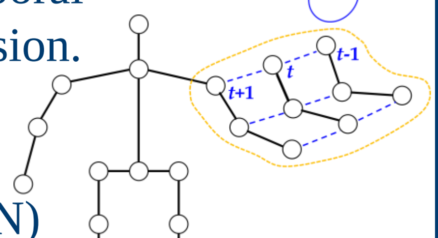
- A human arm has a complex structure and behaviours.
- A human arm is modelled as a *simplified 7 DOF kinematic model*.

Spatio-Temporal Graph Neural Network

- A human arm pose as spatial graphs.
  - Joints = nodes & bones = edges.
  - 3D coordinates as feature vector.

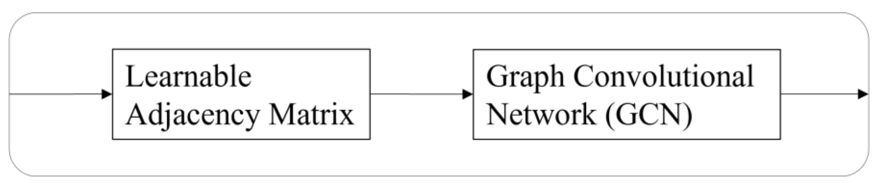


- Human arm motion as spatio-temporal graphs stacked in temporal dimension.



- Space-Time-Separable Graph Convolutional Network (STS-GCN)
  - Learnable adjacency matrix factorises the input motion graph's adjacency matrix to separate spatial and temporal adjacency matrices.
  - Graph convolutional network applies graph convolution per node as defined by the edge weights of the learned adjacency matrix.

Space-Time-Separable Graph Convolutional Network (STS-GCN)



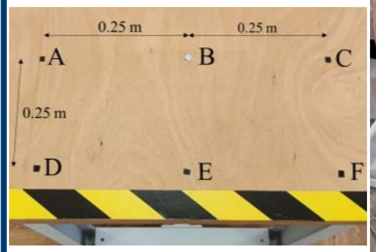
- Temporal Convolutional Network (TCN)
  - Graph convolution across time.

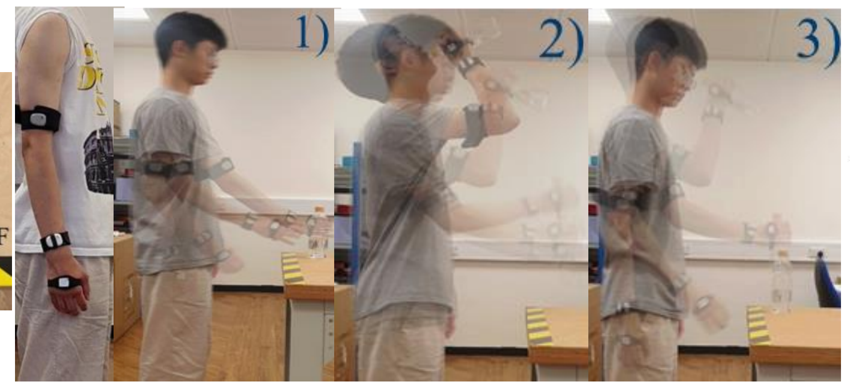
UR5e Robot Arm

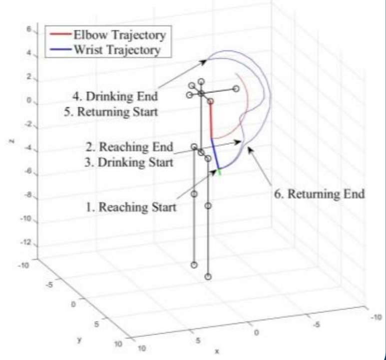
- An industrial robotic arm designed to execute repetitive manual tasks.

3. Human Arm Motion

- Three *commercial, wearable IMUs* are attached to log the Euler angles of each arm segment to reconstruct the original motion.
- The IMU-based motion capture system can capture a range of human arm motions.
- A *human arm drinking motion dataset* of 300 samples is collected.



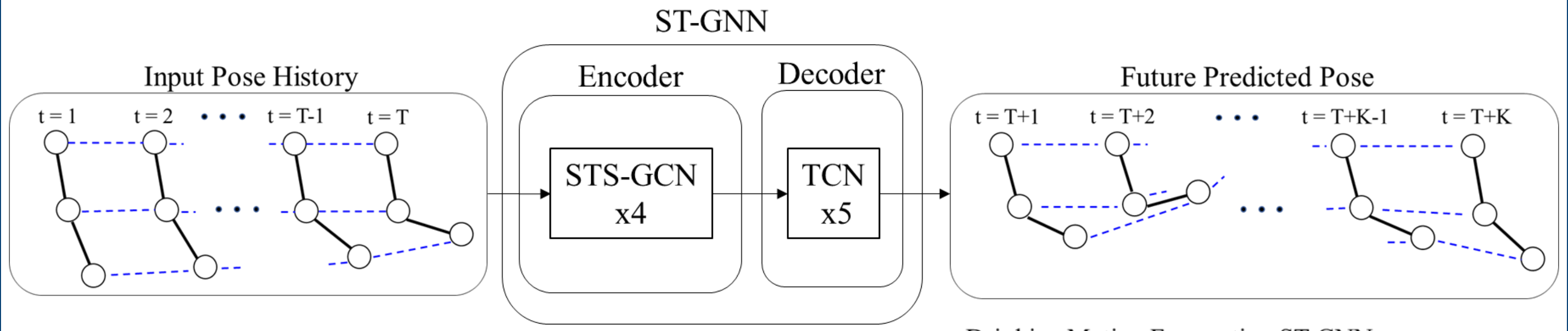




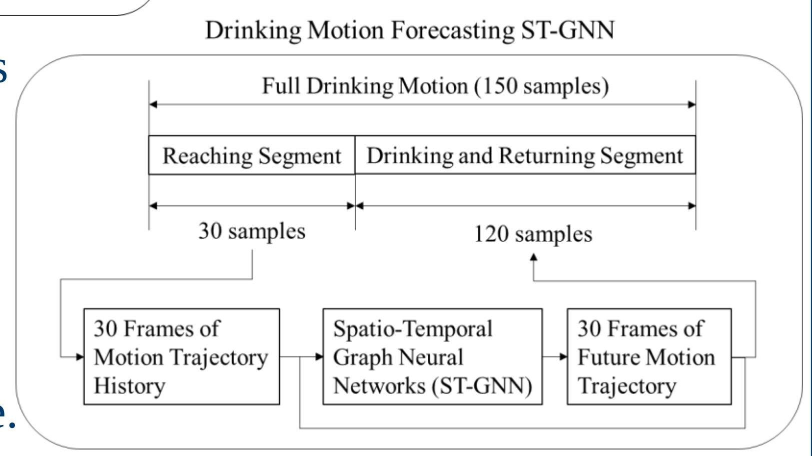
4. Implementation

Motion Forecasting Problem

- *Given an input pose history of an arm motion, can we predict the future motion by learning the dynamics of the motion?*
- Spatio-Temporal Graph Neural Network (ST-GNN) is used to learn the dynamics of the human arm drinking motion using the dataset.
- The reconstructed drinking motion is each 150 sample long.
  - First 30 sample (reaching) : Input pose history.
  - Rest 120 sample (drinking + returning) : Future predicted pose.

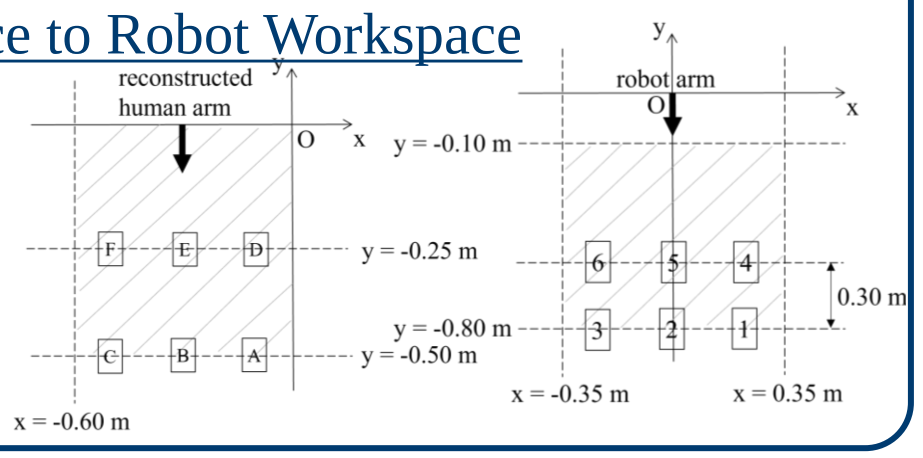


- Encoder-decoder structure generates the future poses by decoding the learned representation.
- 30-frame motion input, 30-frame motion output ST-GNN.
- Recursive connection of the output as new input to generate 120 sample.



Mapping Human Workspace to Robot Workspace

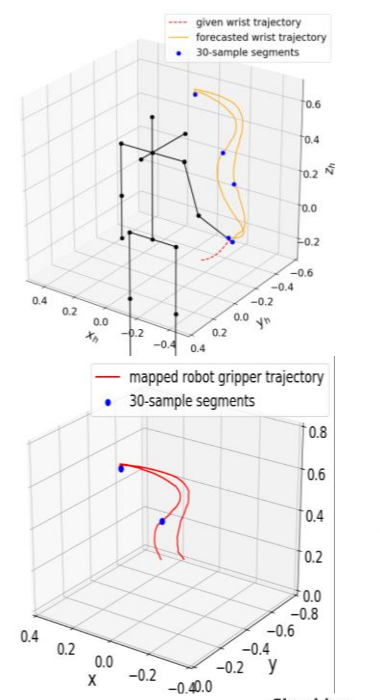
- The ST-GNN generates future predicted human arm motion.
- The human motion needs to be mapped to the robot arm space.

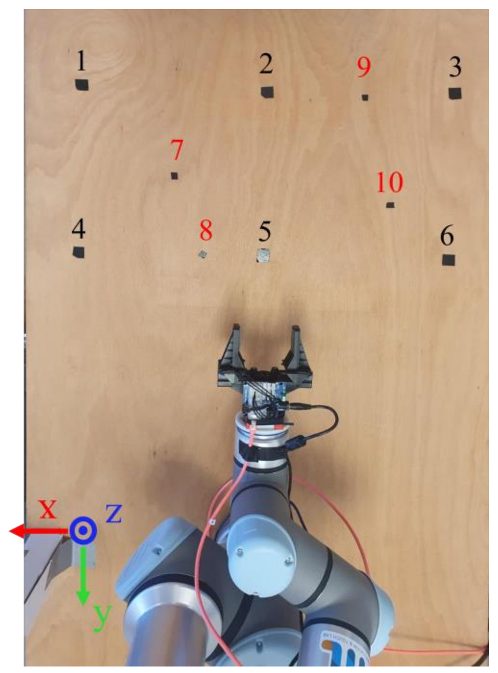


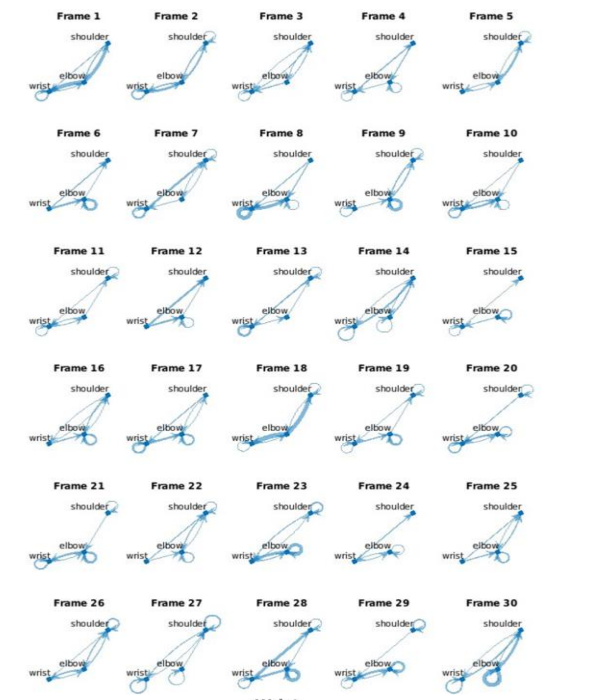
5. Results and Analysis

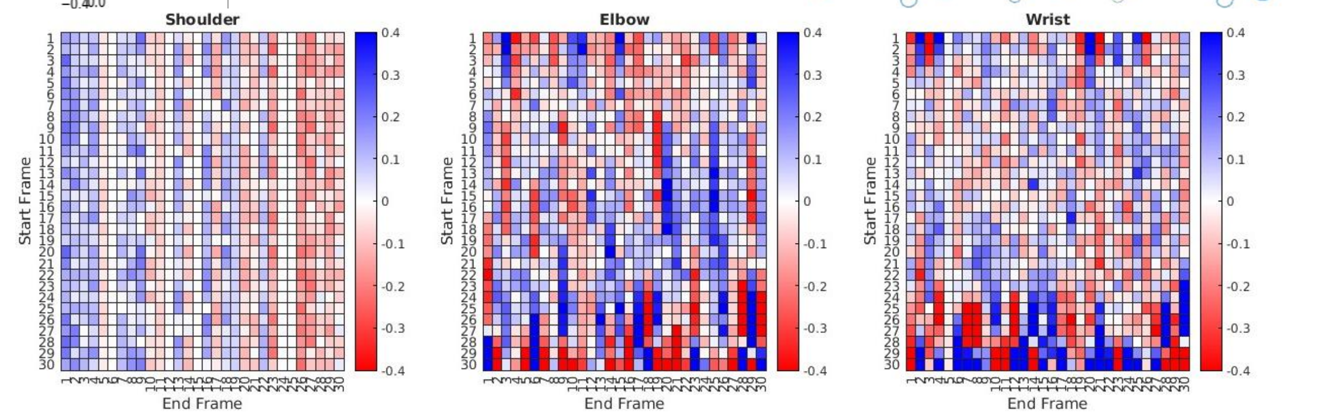
- 6 bottle positions in the drinking motion dataset. (9:1 train/val split)
- 4 new unseen bottle positions to test generalisability of ST-GNN.
- Evaluation metric: Mean Per Joint Position Error (MPJPE).

Data	Mean 120-sample MPJPE Loss	Mean 30-sample MPJPE Loss	Max 30-sample MPJPE Loss	Min 30-sample MPJPE Loss
Training	1.197	0.299	0.307	0.047
Validation / Test	1.301	0.325	0.235	0.050
Unseen Bottle Position	1.407	0.352	0.498	0.206









7. Conclusion

- IMU-based human arm motion capture system is created.
- Human arm drinking motion dataset is collated.
- Full ML pipeline for representation learning the human arm motion dataset by ST-GNN model is created.
- UR5e interface for real-time communication/control is established.
- Mapping from human to robot arm workspace is established.
- Learned representation of the motion dynamics can be visualised.