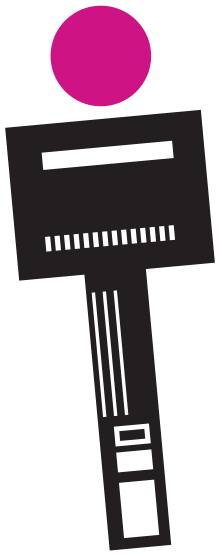

RAM ROBOTICS AND MECHATRONICS

MUSCLE-DRIVEN ROBOTIC CONTROL FOR HUMAN-ROBOT INTERACTION

S. (Shixun) Liu

MSC ASSIGNMENT

Committee:
dr. F.J. Siepel
B. Lan, MSc
dr. ir. K. Niu
dr. J. Reenalda

November, 2025

084RaM2025
Robotics and Mechatronics
EEMCS
University of Twente
P.O. Box 217
7500 AE Enschede
The Netherlands

UNIVERSITY OF TWENTE. | TECHMED CENTRE

UNIVERSITY OF TWENTE. | DIGITAL SOCIETY INSTITUTE

Muscle-Driven Robotic Control for Human-Robot Interaction*

Shixun Liu[1]

***Abstract*— Human-robot interaction (HRI) faces challenges in achieving seamless and intuitive communication, particularly due to limitations of vision-based methods such as occlusions and privacy concerns. Surface electromyography (sEMG) provides a wearable alternative for intent recognition, but existing research predominantly focuses on gesture classification rather than continuous motion estimation. This thesis proposes a novel muscle-driven control framework that uses dual-channel sEMG signals for real-time hand trajectory estimation and robotic following tasks. A Transformer-based deep learning model is trained to decode raw sEMG data into spatial hand positions, while a machine learning mapping directly translates these positions into robotic joint angles for low-latency control. Experimental validation demonstrates the system's capability in dynamic interactions such as high-fives and object following, with performance benchmarked against a vision-based pose estimation system. The results highlight the feasibility of sEMG-based continuous motion decoding for camera-free HRI, offering a privacy-preserving and physiologically grounded approach for assistive and collaborative robotics.**

Keywords— Human-robot interaction, surface electromyography, motion trajectory estimation, real-time control.

I. INTRODUCTION

Human-robot interaction (HRI) has emerged as a critical domain in assistive robotics, rehabilitation, and collaborative systems, where enabling seamless and intuitive communication between humans and machines remains a fundamental challenge [1]. Effective HRI is central to creating robotic systems that can cooperate with humans in daily life, medical assistance, or collaborative workplaces. A core requirement is accurate and timely recognition of human intent, allowing robots to adapt their actions to user needs in a natural and responsive manner.

Traditional intent recognition methods have heavily relied on vision-based approaches [2], which use cameras and advanced computer vision algorithms—such as OpenPose or MediaPipe—to track human body movements [3]. These systems benefit from rich spatial information and have achieved notable success in controlled environments. Deep learning has further boosted the performance of such vision-based models, enabling increasingly accurate pose estimation and motion analysis. However, these methods face significant limitations in real-world use, including susceptibility to occlusions, variable lighting conditions, and inherent privacy concerns associated with continuous visual monitoring [4].

As an alternative, wearable biosignal interfaces, particularly surface electromyography (sEMG), have been studied to capture electrical activity from muscle contractions and infer user intent in a direct and non-visual manner. In recent years, sEMG has been widely explored for predicting continuous motion intention [5], classifying motion gestures [6] , and improving overall human-robot interaction [7].While these approaches demonstrate the potential of sEMG to provide a physiologically grounded communication channel, most existing studies have primarily focused on gesture recognition or discrete motion intention classification [8]–[10]. A smaller body of work has attempted to use sEMG for continuous estimation of joint kinematics or limb trajectories [11]–[17], but these efforts remain relatively fragmented. Recent reviews further highlight that the vast majority of sEMG-based human–machine interface research still emphasizes classification tasks, with only limited attention to regression-based approaches for continuous motion decoding, and no mature solution yet established for real-time, continuous limb trajectory estimation [18].

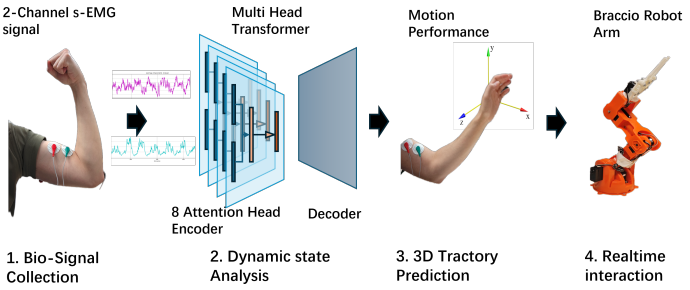

Fig. 1. Conceptual workflow of the proposed sEMG-based robotic arm control system. Dual-channel sEMG signals are processed and fed into a multi-head Transformer network for **Dynamic State Analysis**, enabling the prediction of the hand's 3D motion trajectory in real time. The predicted trajectory is subsequently transmitted to the Braccio robotic arm, allowing **real-time interaction** between human motion intention and robotic execution.

To address this gap, this thesis explores a muscle-driven robotic control paradigm that leverages dual-channel sEMG signals for continuous hand motion trajectory estimation and real-time interaction. The workflow is showed as Fig. 1

Compared to vision-based methods, sEMG offers the advantage of direct, continuous, and privacy-preserving access to human motor activity. If sEMG signals can be effectively decoded into precise hand trajectories and mapped onto robot kinematics [19]–[24], robots can be controlled to follow human movements in real time. Such capabilities would enable highly intuitive forms of interaction—for example, a robotic arm that seamlessly mirrors a user's motion to perform a high-five or cooperative following task—without dependence on cameras or external markers. In this work,

[1]Shixun Liu is with the Robotics and Mechatronics group under Faculty of Electrical Engineering, Mathematics and Computer Science, University of Twente, 7500 AE Enschede, The Netherlands

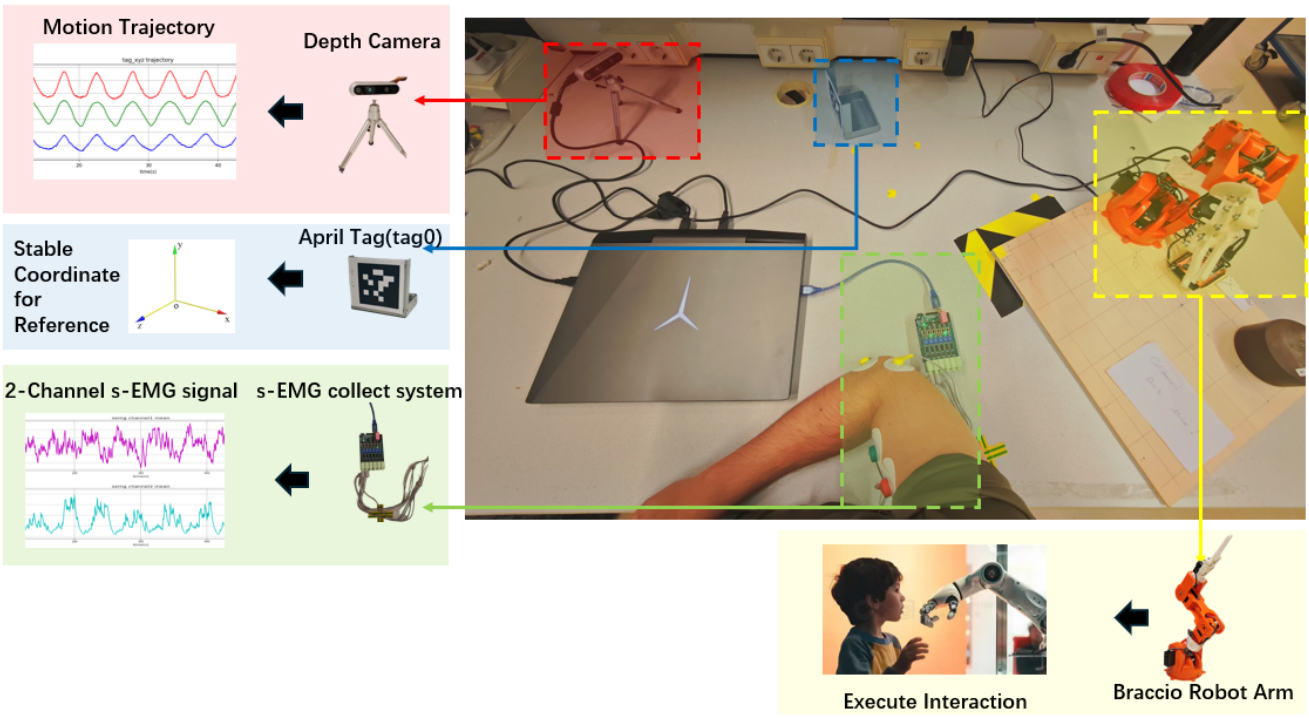

Fig. 2. Overall experimental device setup, including the RealSense D435i Depth Camera to record Motion Trajectory, AprilTag(Tag0) to provide Stable Coordinate for Reference, Multi-channel sEMG acquisition system to get 2-Channel s-EMG signal and Braccio Robot Arm to Execute Interaction.

a transformer-based deep learning model is trained to infer hand trajectories from raw sEMG signals, while a direct mapping from estimated spatial positions to robotic motor angles is established through machine learning to achieve smooth and low-latency control. A vision-based benchmark system using pose estimation is also implemented, allowing systematic evaluation of the sEMG-driven approach in terms of accuracy, latency, and robustness. Experiments with interactive robotic tasks demonstrate the feasibility of using sEMG for trajectory-level control in HRI, while also highlighting both the strengths and the current limitations of this modality compared to vision.

The significance of this study lies in demonstrating the practicality of wearable-sensor-driven HRI for scenarios where cameras are unreliable or intrusive. By enabling robots to follow human motion with physiological grounding and without visual dependence, the proposed approach provides a foundation for developing more accessible, adaptive, and socially compatible robotic systems.

II. METHOD AND MATERIALS

A. Braccio Robot Control System

This section describes the integrated control system developed for the Braccio robotic arm, including the hardware setup, data collection methodology, and the joint angle prediction model. The system enables precise end-effector positioning through vision-based tracking and machine learning-based inverse kinematics.

1) Device Setup: The experimental setup consists of three main components that work in concert to enable accurate motion tracking and control:

Braccio Robot Arm: The Braccio Tinkerkit is a 6-DOF educational robotic arm manufactured by Arduino. It features five servo motors for joint control (base rotation, shoulder, elbow, wrist vertical, and wrist rotation) plus a gripper, with a maximum payload of 150g. The arm is controlled through an Arduino Uno board that receives joint angle commands via serial communication, making it suitable for research applications requiring precise positional control.

AprilTags: Fiducial markers based on the AprilTag library were utilized for visual tracking. These square, high-contrast markers consist of a black border surrounding an inner binary pattern that encodes a unique identifier. The employed dictionary, `DICT_APRILTAG_36h10`, offers a large set of unique tag IDs with strong robustness against perspective distortion and lighting variation. AprilTags enable reliable detection and precise 6-DOF pose estimation, achieving sub-centimeter accuracy under well-calibrated and properly illuminated conditions.

RealSense D435i Depth Camera: An Intel RealSense D435i depth sensing camera was used as the primary vision sensor. This device combines a stereo depth module with a color sensor and an inertial measurement unit (IMU). It provides RGB video at 1280×720 resolution and depth information at up to 90 frames per second, with a depth field of view of 87°×58°. The camera was calibrated prior to experiments to ensure accurate 3D measurements.

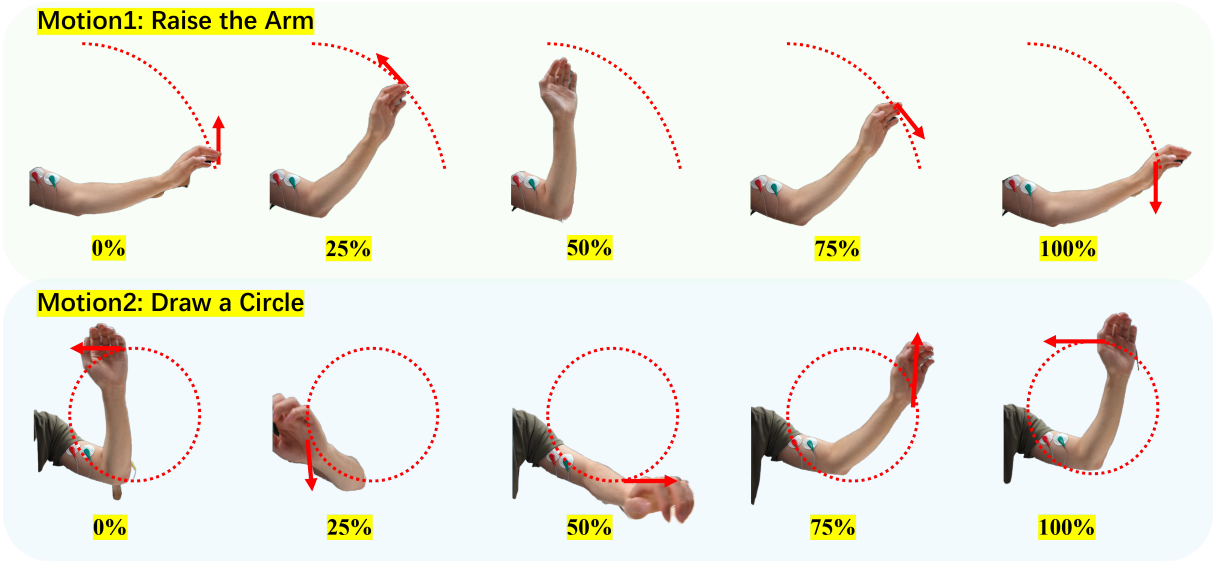

Fig. 3. Two types of motions performed during the experiment. In **Motion 1**, participants lifted and bent their right arm before returning to the initial position near the robot arm. In **Motion 2**, participants drew a clockwise circle in the air with their hand. Each cycle lasted about 4–6 seconds, and motion execution was voluntary and unconstrained.

2) Data Collection and Processing: A systematic data collection protocol was established to generate training data for the joint angle prediction model. Two April tags were deployed in the workspace: Tag1 was fixed to the laboratory table surface, establishing a stable reference coordinate system relative to the Braccio base, while Tag0 was precisely attached to the end-effector (gripper tip) to track its 3D position.

Prior to data collection, joint angle limits were empirically adjusted to ensure physically feasible and collision-free configurations within the robot's reachable workspace. The sampled ranges were defined as follows:

- Base rotation: 0° to 90°
- Shoulder: 45° to 135°
- Elbow: dynamically adjusted based on shoulder angle:
  - 45°–155° when shoulder $< 95°$
  - 15°–45° when shoulder $= 95°$
  - 0°–50° when shoulder $> 95°$
- Wrist vertical: adapted according to shoulder and elbow configuration:
  - 60°–90° for shoulder $< 95°$ and elbow $< 90°$
  - 15°–45° for shoulder $< 95°$ and elbow $\geq 90°$
  - 10°–60° for shoulder $> 95°$
- Wrist rotation: 45°, 90°, and 135°

This hierarchical sampling strategy was designed to ensure sufficient workspace coverage while preventing self-collision and avoiding extreme joint configurations that could compromise mechanical stability.

A Python script was developed to systematically traverse all viable joint angle combinations within these constraints, generating diverse end-effector positions throughout the workspace. During each movement, the RealSense camera simultaneously tracked both tags at 30 Hz. The 3D position of Tag0 (end-effector) was transformed into the Tag1 coordinate system using homogeneous transformation matrices:

$$P_{\text{Tag1}} = T_{\text{Tag1}}^{\text{Camera}} \cdot (T_{\text{Tag0}}^{\text{Camera}})^{-1} \cdot P_{\text{Tag0}} \tag{1}$$

where $T_{\text{Tag}}^{\text{Camera}}$ represents the transformation from camera coordinates to tag coordinates. Joint angle configurations were synchronized with the corresponding end-effector positions, resulting in a comprehensive dataset of 12,500 samples, each containing joint angles $\theta = [\theta_1, \theta_2, \theta_3, \theta_4, \theta_5]$ and the corresponding 3D position $p = [x, y, z]$ in Tag1 coordinates.

3) Joint Angle Predictor: A Multi-Layer Perceptron (MLP) was implemented to predict joint angles from image-space end-effector observations (u, v, depth). The network input is a 3-dimensional vector and the network outputs 4 joint angles. The architecture is an improved MLP with three hidden layers and built-in regularization components:

- Input: 3 features (u, v, depth).
- Hidden layers: 256, 128, 64 neurons.
- Nonlinearities and regularization: Batch Normalization after each linear block, LeakyReLU activations (negative slope $= 0.01$) and Dropout ($p = 0.2$).
- Output: 4 linear outputs (one per predicted joint angle).

During training, only a position-based loss function was employed. Specifically, the predicted joint angles θ_{pred} were passed through the forward kinematics model $FK(\theta)$, which computes the corresponding end-effector position based on the Braccio arm's Denavit–Hartenberg parameters. The training objective minimized the mean squared error between the predicted end-effector position and the ground-truth position, formulated as:

$$\mathscr{L}_{\text{pos}} = \text{MSE}\big(FK(\theta_{\text{pred}}), p_{\text{true}}\big) \tag{2}$$

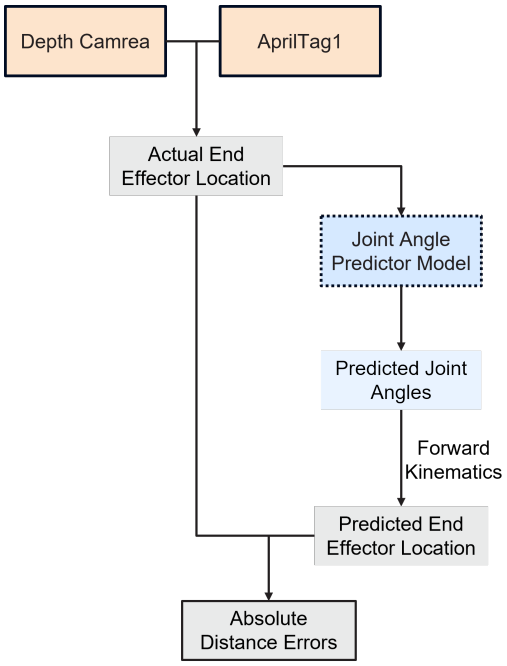

Fig. 4. Conceptual workflow for obtaining the **Joint Angle Predictor Model Absolute Distance Errors**. The *Actual End-Effector Location* is measured using a depth camera and an AprilTag (ID 1) rigidly attached to the robot's end effector. This actual position is fed into the *Joint Angle Predictor Model* to estimate the corresponding joint angles. Through *Forward Kinematics*, the predicted joint angles are transformed into a *Predicted End-Effector Location*. Finally, the *Absolute Distance Errors* between the predicted and actual end-effector positions are computed to evaluate the performance of the robotic execution system.

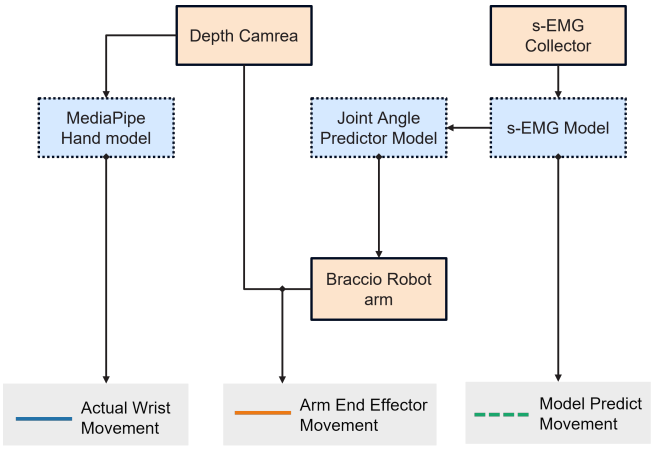

Fig. 5. Conceptual workflow illustrating how the three movement trajectories are obtained. The **Actual Wrist Movement** (blue solid line) is captured by the depth camera and processed through the *MediaPipe Hand* model. The **Model Predicted Movement** (green dashed line) is generated from the s-EMG signals collected by the *s-EMG Collector* and processed using the *s-EMG Model*. The predicted movement is then fed into the *Joint Angle Predictor Model*, executed by the *Braccio Robot Arm*, and monitored via the depth camera to obtain the **Arm End-Effector Movement** (orange solid line).

This approach focuses on ensuring that the predicted joint configurations yield accurate end-effector positions in Cartesian space, without directly penalizing individual joint angle deviations.

Training was performed using the Adam optimizer with a learning rate of 0.001, batch size of 32, weight decay of 1×10^{-5}, and 1000 epochs. The final model contained approximately 43k trainable parameters.

Model performance was evaluated using position-based mean absolute error (MAE) between the predicted and ground-truth end-effector positions The results were:

- Position MAE:
  - x: 13.83mm, y: 12.07mm, z: 20.22mm, overall: 27.31mm

These results demonstrate that the MLP model can approximate the inverse kinematics mapping with reasonable accuracy in Cartesian space, though residual errors in the z direction indicate limited precision in depth estimation from image-based inputs.

B. Surface electromyography predict system

1) Device Setup: **Surface electromyography** (sEMG) is a non-invasive technique that measures electrical signals generated by muscle fibers during contraction and relaxation. These bioelectrical signals are captured through electrodes placed on the skin surface, providing valuable insights into neuromuscular activity and movement intention. sEMG signals typically range from 0 to 500 Hz in frequency and 0 to 10 mV in amplitude, requiring specialized amplification and filtering for accurate acquisition.

Multi-channel sEMG acquisition system (SizhiRui Technology) enables simultaneous recording from multiple muscle sites, offering comprehensive monitoring of coordinated muscle activation patterns [25]–[28]. The two-channel sEMG device employed in this study incorporates differential amplification, band-pass filtering (20-450 Hz), and analog-to-digital conversion at 1000 Hz sampling rate to ensure high-quality signal acquisition while minimizing motion artifacts and environmental noise.

For motion tracking,**Mediapipe** [29] provides a real-time hand landmark detection framework that identifies 21 anatomical keypoints in the hand, including wrist joint, finger joints, and fingertip positions. The model utilizes a convolutional neural network architecture that processes RGB input frames and outputs normalized 2D coordinates of hand landmarks with sub-pixel accuracy, enabling precise tracking of wrist movements and hand gestures.

Fig. 2 illustrates the complete experimental setup. The two-channel sEMG device was configured to record muscle activities specifically from the biceps and triceps muscles of the right upper arm. Arm movement trajectories were captured using an Intel RealSense D435i spatial depth camera, which provided synchronized RGB and depth streams. The Mediapipe hand landmark detection model extracted 2D wrist positions from RGB frames, and these coordinates were mapped to corresponding depth values to reconstruct 3D motion trajectories in camera coordinates. A fixed Apriltag (Tag1) [30], [31] was employed as a reference marker to transform wrist 3D locations from camera coordinates into a

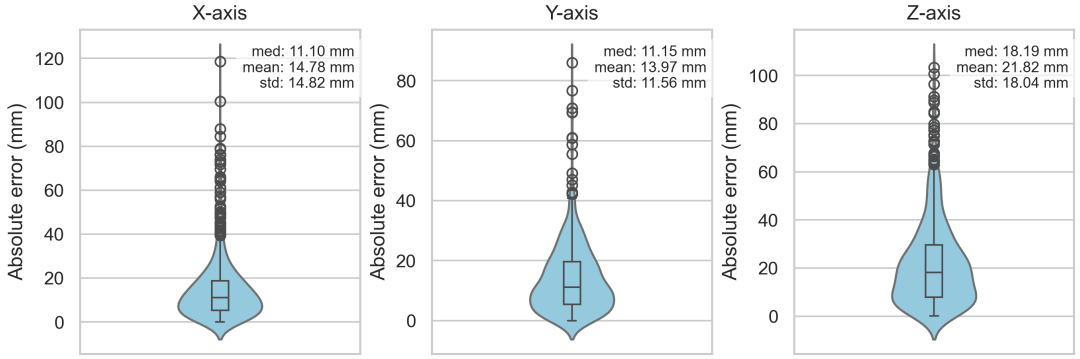

Fig. 6. Box and violin plots of absolute distance errors between the forward-kinematics predicted end-effector locations and ground-truth end-effector locations. The absolute distance errors were computed along the x, y, and z axes, with the following results: x-axis: 14.78 ± 14.82 mm (median: 11.10 mm), y-axis: 13.97 ± 11.56 mm (median: 11.15 mm), and z-axis: 21.82 ± 18.04 mm (median: 18.19 mm).

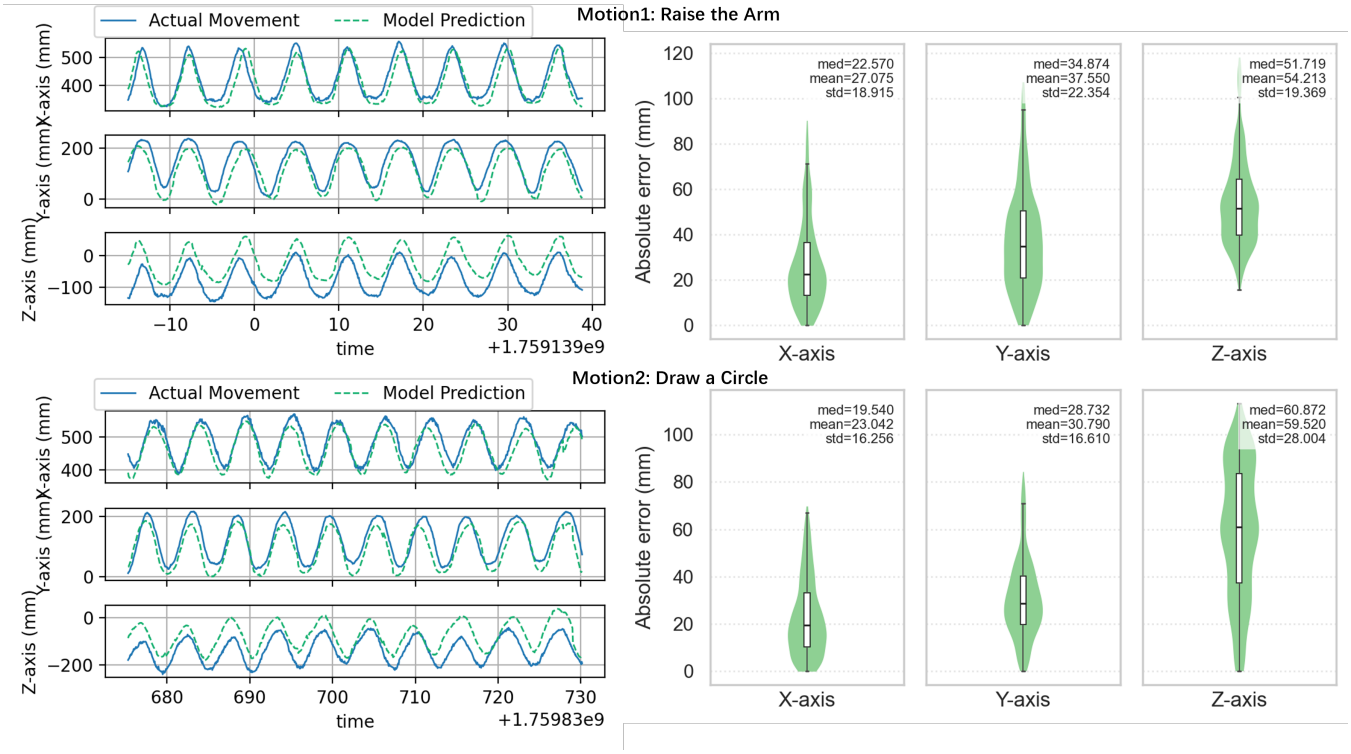

Fig. 7. Visualization and quantitative evaluation of sEMG-based wrist trajectory prediction performance for two motion types. The left panels illustrate the **Model Predicted Movement** (green dashed line) and the **Actual Wrist Movement** (blue solid line), obtained from the sEMG model and MediaPipe Hand tracking, respectively. The right panels present box and violin plots of the absolute positional errors along the x, y, and z axes. For **Motion 1**, the absolute errors were: x-axis: 27.08 ± 18.91 mm (median: 22.57 mm), y-axis: 37.55 ± 22.35 mm (median: 34.87 mm), z-axis: 54.21 ± 19.37 mm (median: 51.72 mm). For **Motion 2**, the corresponding errors were: x-axis: 23.04 ± 16.26 mm (median: 19.54 mm), y-axis: 30.79 ± 16.61 mm (median: 28.73 mm), and z-axis: 59.52 ± 28.00 mm (median: 60.87 mm).

stable Apriltag coordinate system, ensuring consistent spatial referencing across sessions.

Prior to experiments, sEMG electrodes were carefully attached to the right upper arm of each subject following standard protocols (Fig. 2). The subjects' skin was thoroughly cleaned with alcohol wipes to remove dirt, oil, and dead skin cells [32], [33], optimizing electrode-skin contact impedance and signal quality. Electrode positions were adjusted during preliminary testing to maximize signal-to-noise ratio and ensure clear visualization of sEMG waveforms. The spatial camera was positioned approximately 1.5 meters in front of the subject and Apriltag to maintain optimal field of view for recording movements and wrist locations throughout the experimental workspace.

2) Description of the Performed Motions: The performed motions are illustrated in Fig. 3 . In Motion 1, subjects lifted their right arm away from the robot arm, bent the elbow freely, and then returned to the starting position, where the forearm was parallel to the table and close to the robot arm. In Motion 2, subjects drew a clockwise circle in the air using their hands. These two motions represent two modes of

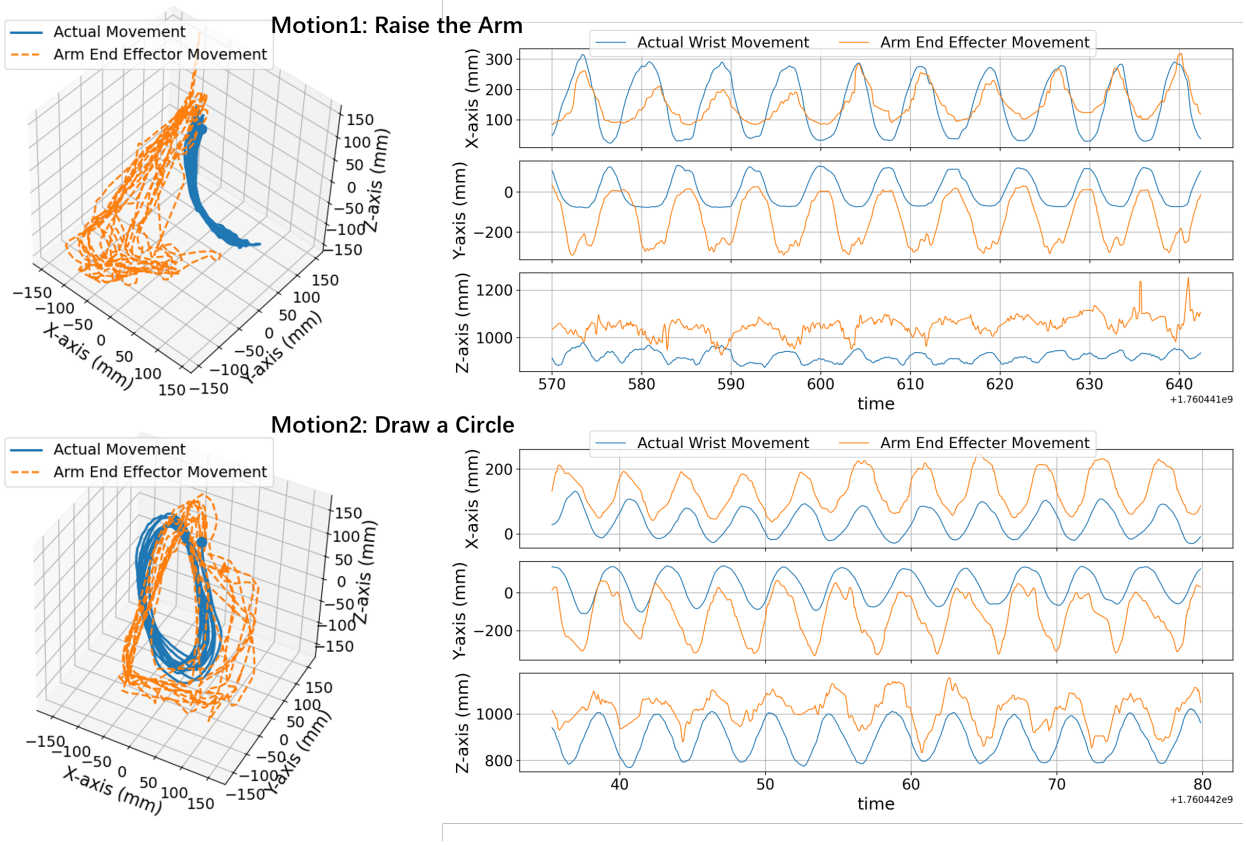

Fig. 8. Reconstructed 3D trajectories and corresponding per-axis displacements for two motion types. The left panels show the 3D trajectories of the **Actual Wrist Movement** (blue solid line) obtained via MediaPipe tracking and the **Arm End-Effector Movement** (orange solid line) captured using the AprilTag-based localization system. The right panels depict their respective displacements along the x, y, and z axes.

interaction with the robot arm, such as high-fiving (Motion 1) or having the robot arm follow (Motion 2). For both motions, no specific instructions were given to restrict the range of motion or pace. The duration of each motion cycle was approximately 4 to 6 seconds, and each motion session lasted about 10 minutes. The motions were highly voluntary and exhibited variability in each cycle.

C. Signal Processing

During the experiments, the sEMG device operated at a sampling rate of approximately 900 Hz, while the visual tracking system captured data at about 17 Hz. The sEMG signals underwent fundamental processing at the acquisition stage but were not subjected to additional alterations to preserve the integrity of the original information.

Synchronization of the signals was implemented based on the frames acquired by the binocular camera, which had a frame rate of 17 Hz. To maximize the retention of sEMG data, for each camera frame timestamp, the 30 closest sEMG samples were selected. In experiments where the camera was not utilized, the system's internal clock emulated a 17 Hz sampling rate. Ultimately, all signals were synchronized to a unified 17 Hz timeline.

D. Transformer-based Network Structure

A Dual-Channel sEMG Transformer network was implemented to perform continuous 3D coordinate regression from dual-channel surface electromyography (sEMG) signals. The network combines convolutional feature extraction with Transformer-based sequence modeling to capture both short-term temporal patterns and long-range dependencies inherent in sEMG dynamics.

1) Architecture Overview: The model processes input sEMG sequences of shape $(batch\_size, 2, 30, 85)$ and outputs corresponding 3D coordinate sequences $(batch\_size, 85, 3)$. The architecture consists of the following components:

- **Channel-specific encoding:** Each of the two sEMG channels is processed independently through a 1D convolutional block (kernel size 3, padding 1) to extract temporal features while preserving channel-specific activation patterns.
- **Channel fusion:** The outputs from both channels are concatenated and linearly projected to form a unified representation, enhancing inter-channel feature interaction:

$$H_{\text{fused}} = W_f \cdot \text{Concat}(H_1, H_2) + b_f \tag{3}$$

- **Positional encoding:** Sinusoidal positional encodings

are added to the fused sequence to preserve temporal order.

- **Transformer encoder:** A 4-layer Transformer encoder with 8 attention heads and a hidden dimension of 256 models long-range temporal dependencies. Each encoder block includes multi-head self-attention and position-wise feed-forward layers with residual connections and layer normalization.
- **Regression head:** A linear projection maps the Transformer's output features to 3D coordinates:

$$\hat{Y} = H_{\text{transformer}} W_r + b_r \tag{4}$$

where $W_r \in \mathbb{R}^{128\times 3}$ and $b_r \in \mathbb{R}^3$.

2) Training and Evaluation: Training was conducted using the Adam optimizer with a learning rate of 1×10^{-4}, batch size of 32, and a mean squared error (MSE) loss between predicted and ground-truth 3D coordinates. Each training sequence consisted of a window size of 40 and a stride of 1, ensuring sufficient temporal continuity for motion prediction. The model was trained for 50 epochs, and the final checkpoint was selected based on the lowest validation loss.

The Transformer encoder was configured with a model dimension of 128, 8 attention heads, 4 encoder layers, and a feed-forward hidden dimension of 256. Dropout layers and layer normalization were employed throughout the architecture to stabilize training and mitigate overfitting.

Due to variations in window size, stride, and data composition across different motion categories, the detailed quantitative evaluation is not presented here. Instead, the focus is placed on the model's capacity to generalize across movement types. Qualitatively, the Transformer-based model demonstrated stable convergence and effective learning of temporal dependencies within dual-channel sEMG sequences, enabling consistent and smooth 3D motion reconstruction from raw neural activation signals.

III. EXPERIMENTAL RESULTS

Before evaluating the full sEMG-based robotic following system, the Joint Angle Predictor (JAP) Model was first quantitatively tested to assess its accuracy in reconstructing spatial positions from inverse-predicted joint angles.The workflow is showed in Fig. 4. The end-effector positions measured in the tag1 coordinate system were used as model inputs, and the predicted joint angles were subsequently processed through the same forward kinematics computation used during training to obtain the model-estimated end-effector positions. The absolute distance errors between the predicted and ground-truth positions along the x, y, and z axes were then calculated, with the resulting distributions shown as combined box and violin plots in Figure 6. The results were as follows:

- x-axis: 14.78 ± 14.82 mm (median: 11.10 mm)
- y-axis: 13.97 ± 11.56 mm (median: 11.15 mm)
- z-axis: 21.82 ± 18.04 mm (median: 18.19 mm)

After completing model training, the sEMG-based robotic arm following system was experimentally validated using two distinct motion patterns. An AprilTag (ID 1) was rigidly mounted on the end-effector of the robotic arm to provide accurate pose measurements. The human wrist motion was simultaneously tracked, enabling quantitative comparison between the predicted robotic trajectory and the true wrist movement.

In addition to the robotic experiments, the predictive performance of the sEMG deep learning model was first evaluated quantitatively. Since the model was designed to anticipate future wrist positions to compensate for system latency, the predicted trajectories were temporally aligned with the recorded wrist trajectories for comparison. The overall data acquisition and processing workflow for obtaining the *Actual Wrist Movement*, *Model Predicted Movement*, and *Arm End-Effector Movement* is illustrated in Figure 5, which conceptually demonstrates how these three trajectories were generated and compared in subsequent analyses.

For both motion types, the predicted and actual wrist trajectories were visualized, and the absolute positional differences were computed. The overall distributions were plotted as box and violin plots, as shown in Figure 7.

For **Motion 1**, the absolute errors were:

- x-axis: 27.08 ± 18.91 mm (median: 22.57 mm)
- y-axis: 37.55 ± 22.35 mm (median: 34.87 mm)
- z-axis: 54.21 ± 19.37 mm (median: 51.72 mm)

For **Motion 2**, the corresponding errors were:

- x-axis: 23.04 ± 16.26 mm (median: 19.54 mm)
- y-axis: 30.79 ± 16.61 mm (median: 28.73 mm)
- z-axis: 59.52 ± 28.00 mm (median: 60.87 mm)

Figure 8 illustrates the reconstructed 3D trajectories and corresponding per-axis displacements (x, y, and z) for both motion types. To quantitatively assess tracking accuracy, absolute distance errors between the robotic end-effector and wrist trajectories were computed for each axis. The overall error distributions are visualized as box and violin plots in Figure 9.

For **Motion 1**, the absolute displacement errors were:

- x-axis: 53.56 ± 30.75 mm (median: 52.48 mm)
- y-axis: 149.19 ± 53.26 mm (median: 148.22 mm)
- z-axis: 130.86 ± 51.43 mm (median: 143.87 mm)

For **Motion 2**, the corresponding errors were:

- x-axis: 95.17 ± 28.42 mm (median: 91.25 mm)
- y-axis: 157.08 ± 64.50 mm (median: 157.85 mm)
- z-axis: 119.73 ± 38.46 mm (median: 116.93 mm)

To further analyze motion repeatability, each continuous movement sequence was segmented into individual motion cycles and temporally aligned. The overlaid trajectories across multiple repetitions are presented in Figure 10. Solid lines represent the mean trajectory, while shaded regions denote one standard deviation across cycles. These results show consistent motion reproduction across trials, with the smallest variability observed along the x and z axes.

To evaluate the temporal performance of the system after predictive compensation, a residual motion delay analysis was conducted along the y-axis, as shown in Figure 11, which exhibited the most distinct cyclic motion. For each

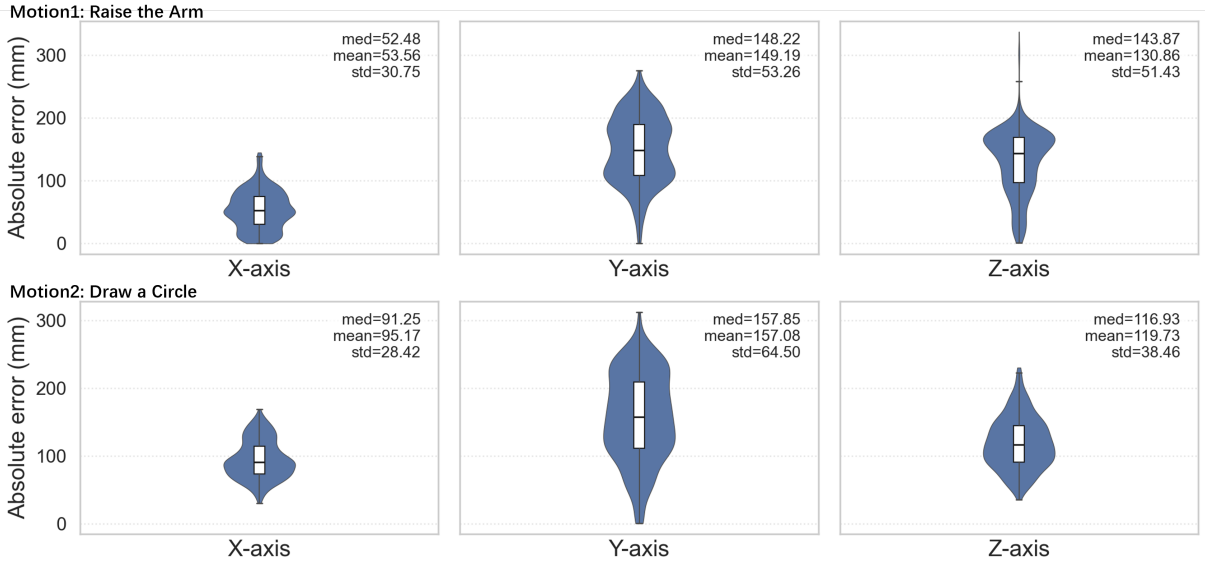

Fig. 9. Quantitative evaluation of tracking accuracy between the **Actual Wrist Movement** and the **Arm End-Effector Movement**. The box and violin plots illustrate the distributions of absolute distance errors along the x, y, and z axes for both motion types. For **Motion 1**, the absolute displacement errors were 53.56 ± 30.75 mm (median: 52.48 mm) along the x-axis, 149.19 ± 53.26 mm (median: 148.22 mm) along the y-axis, and 130.86 ± 51.43 mm (median: 143.87 mm) along the z-axis. For **Motion 2**, the errors were 95.17 ± 28.42 mm (median: 91.25 mm), 157.08 ± 64.50 mm (median: 157.85 mm), and 119.73 ± 38.46 mm (median: 116.93 mm) for the x, y, and z axes, respectively.

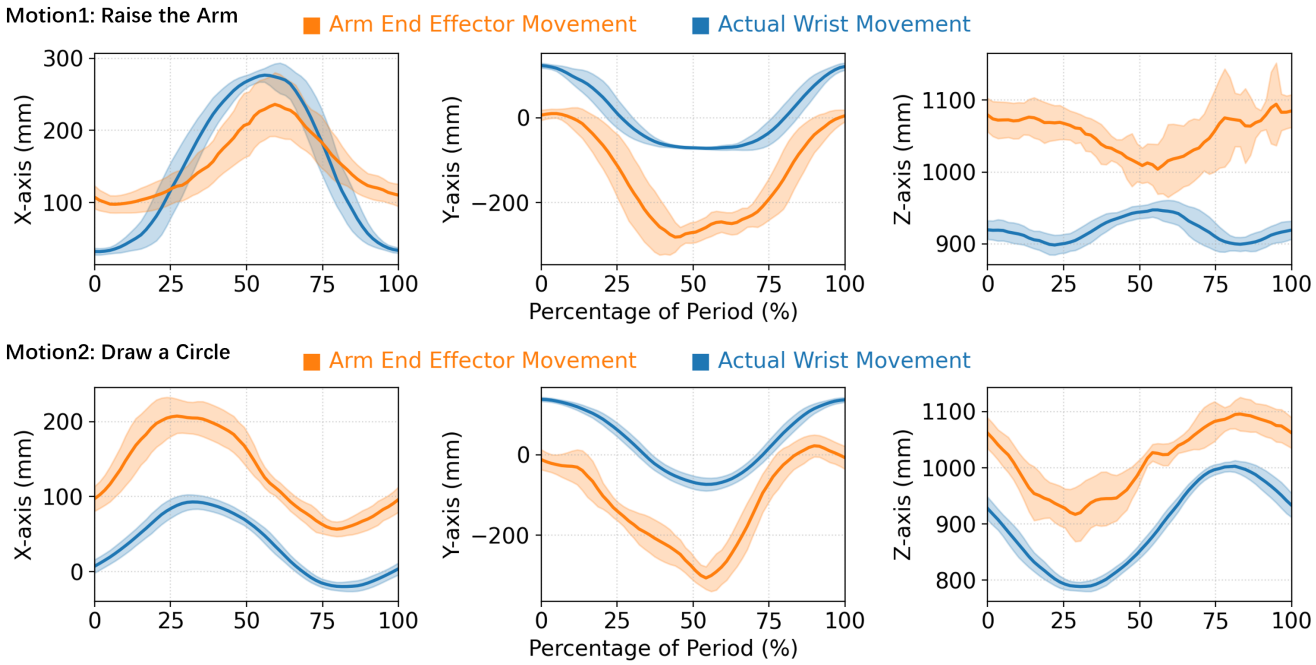

Fig. 10. Analysis of motion repeatability across multiple cycles for both motion types. Each continuous sequence of the **Actual Wrist Movement** and the **Arm End-Effector Movement** was segmented into individual cycles and temporally aligned. Solid lines represent the mean trajectories across repetitions, while shaded areas indicate one standard deviation, reflecting the variability within each motion cycle.

motion type, the peaks and troughs of the y-axis trajectories from both the wrist and the robotic end-effector were identified. The nearest peak-to-peak and trough-to-trough absolute time differences were then computed to estimate the residual response delay distribution of the sEMG-driven robotic following system.

For **Motion 1**, the absolute time differences were:

- Peak–peak: 0.516 ± 0.352 s
- Trough–trough: 1.188 ± 0.379 s

For **Motion 2**, the corresponding results were:

- Peak–peak: 0.502 ± 0.178 s
- Trough–trough: 0.181 ± 0.115 s

These temporal analyses provide a detailed characterization of the residual response behavior of the real-time sEMG-

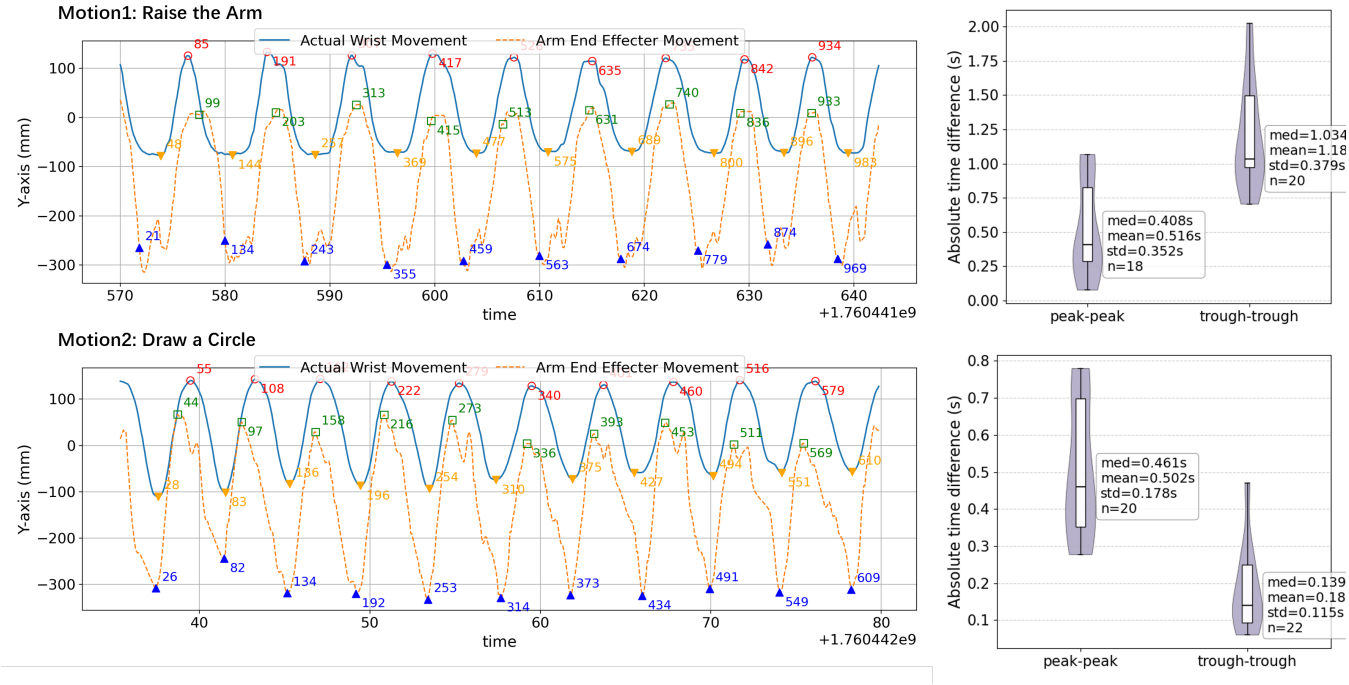

Fig. 11. Residual motion delay analysis along the y-axis for the two motion types. Peaks and troughs of the y-axis trajectories from both the **Actual Wrist Movement** (blue solid line) and the **Arm End-Effector Movement** (orange solid line) were identified (left panel). The nearest peak-to-peak and trough-to-trough absolute time differences were computed and visualized using combined violin and box plots (right panel), illustrating the distribution of residual response delays in the sEMG-driven robotic following system after predictive compensation. For **Motion 1**, peak-to-peak and trough-to-trough delays were 0.516 ± 0.352 s and 1.188 ± 0.379 s, respectively; for **Motion 2**, the corresponding delays were 0.502 ± 0.178 s and 0.181 ± 0.115 s.

driven robotic following system, illustrating how closely the robotic execution aligns with human wrist motion after predictive compensation.

IV. DISCUSSION

This study presents a novel framework for real-time hand trajectory estimation and robotic following using surface electromyography (sEMG) signals. Unlike traditional motion tracking systems that rely on full-limb sensor arrays [14], [34], our approach demonstrates the feasibility of tracking forearm movements using a single upper-arm-mounted device. This is particularly advantageous for wearable prosthetic systems, where only partial limb measurements are available.

The proposed system employs a Transformer-based deep learning model to decode dual-channel sEMG signals into continuous spatial positions for low-latency control. This method captures the biomechanical dynamics of the entire arm, enabling the robotic arm to interact with the user in a manner that reflects natural human motion. Notably, this approach does not require biomechanical parameters such as muscle fiber length or pennation angle, distinguishing it from model-based motion estimation techniques [35]. This characteristic enhances the system's adaptability, allowing it to generalize across different users and motion types with minimal training data.

Despite its advantages, several limitations were identified. The Joint Angle Predictor model achieved moderate accuracy in position reconstruction, with median errors of 11.10 mm, 11.15 mm, and 18.19 mm along the x-, y-, and z-axes, respectively. However, the sEMG-driven robotic system exhibited higher tracking errors in dynamic tasks, with median errors up to 157.85 mm in the y-axis for Motion 2. Additionally, residual delays were observed, including peak-to-peak differences of approximately 0.5 seconds, highlighting the challenges in achieving real-time synchronization between human movements and robotic responses.

These discrepancies can be attributed to several factors. Kinematic mismatches arise due to inherent differences between robotic and human motion patterns, as robotic kinematics often cannot fully replicate human biomechanics due to joint constraints [36]. Model instability in the Transformer-based network caused prediction fluctuations and occasional irregular trajectories, a common issue in deep learning for sEMG where noise and variability can lead to unreliable outputs [37] [38]. Joint angle constraints of the robotic arm further limited motion trajectories, exacerbating mismatches as hardware design affects imitation learning [36]. Moreover, tag occlusion from suboptimal angles or occluded frames may have influenced assessment accuracy, a known problem in fiducial marker systems like AprilTag [39].

Nevertheless, this work contributes by developing a real-time robotic interaction system driven solely by sEMG signals, extending beyond traditional gesture classification to continuous motion estimation. The use of Transformer architectures allows for the capture of temporal dependencies in bio-signals, demonstrating the potential of muscle-intent-

based control as a natural and privacy-friendly alternative to vision-based interaction. This is particularly relevant in scenarios where cameras are impractical or pose privacy concerns.

For future directions, multi-modal integration combining sEMG with complementary sensors such as inertial measurement units or depth cameras could enhance robustness and accuracy by compensating for kinematic and model issues, as advocated in sensor fusion approaches for EMG-based systems [40]. Application expansion to rehabilitation training and collaborative robotics could unlock broader utility, building on existing robotic devices for upper limb therapy and adaptive cooperation [41]. In conclusion, this study validates sEMG as a viable modality for continuous motion decoding in human-robot interaction, offering a privacy-aware and physiologically grounded solution, though further refinements in modeling and hardware integration are needed to achieve higher precision and broader applicability in real-world settings.

V. CONCLUSIONS

This study demonstrates the feasibility of using surface electromyography (sEMG) for real-time hand trajectory estimation and robotic following in human-robot interaction. By leveraging a Transformer-based deep learning model, dual-channel sEMG signals were decoded into continuous 3D wrist positions, enabling the Braccio robotic arm to track human movements without requiring full-limb sensor arrays or explicit biomechanical parameters. The system successfully captured the dynamic relationships of forearm and upper-arm motion, offering a privacy-friendly, vision-free alternative to traditional camera-based tracking.

Experimental results showed that the Joint Angle Predictor model achieved moderate accuracy, while the sEMG-driven robotic system exhibited higher tracking errors and residual delays in dynamic tasks, highlighting challenges in real-time synchronization and robotic imitation of human biomechanics. These limitations stem from kinematic mismatches, model prediction variability, robotic joint constraints, and occasional marker occlusion.

Despite these challenges, the framework establishes a practical pathway for continuous, low-latency control of wearable or assistive robotic devices. Its minimal sensor requirements and adaptability to new users or motion types suggest strong potential for prosthetic applications. Future work integrating complementary sensors, improving model robustness, and refining robotic hardware could further enhance accuracy and reliability, extending applicability to rehabilitation, collaborative robotics, and broader human-robot interaction scenarios.

REFERENCES

[1] L. Zongxing, Z. Jie, Y. Ligang, C. Jinshui, and L. Hongbin, "The human–machine interaction methods and strategies for upper and lower extremity rehabilitation robots: A review," *IEEE Sensors Journal*, vol. 24, no. 9, pp. 13773–13787, 2024.

[2] Y. Chen, Y. Tian, and M. He, "Monocular human pose estimation: A survey of deep learning-based methods," *Computer Vision and Image Understanding*, vol. 192, p. 102897, Mar. 2020.

[3] N. Robinson, B. Tidd, D. Campbell, D. Kulić, and P. Corke, "Robotic vision for human-robot interaction and collaboration: A survey and systematic review," *ACM Transactions on Human-Robot Interaction*, vol. 12, p. 1–66, Feb. 2023.

[4] S. Edriss, C. Romagnoli, L. Caprioli, V. Bonaiuto, E. Padua, and G. Annino, "Commercial vision sensors and ai-based pose estimation frameworks for markerless motion analysis in sports and exercises: a mini review," *Frontiers in Physiology*, vol. Volume 16 - 2025, 2025.

[5] Z. Wei *et al.*, "Continuous motion intention prediction using semg for upper-limb rehabilitation: A systematic review of model-based and model-free approaches," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 32, pp. 1487–1504, 2024.

[6] A. Jaramillo-Yánez *et al.*, "Real-time hand gesture recognition using surface electromyography and machine learning: A systematic literature review," *Sensors (Basel, Switzerland)*, vol. 20, no. 9, p. 2467, 2020.

[7] Y. Liu, X. Li, L. Yang, and H. Yu, "A transformer-based gesture prediction model via semg sensor for human–robot interaction," *IEEE Transactions on Instrumentation and Measurement*, vol. 73, pp. 1–15, 2024.

[8] A. Jaramillo-Yáñez, M. E. Benalcázar, E. Mena-Maldonado, *et al.*, "Real-time hand gesture recognition using surface electromyography and machine learning: A systematic literature review," *Sensors (Basel)*, vol. 20, no. 9, p. 2467, 2020. Systematic review showing many real-time sEMG works are gesture-recognition oriented.

[9] B. Lan, S. Stramigioli, and K. Niu, "Anatomical region recognition and real-time bone tracking methods by dynamically decoding a-mode ultrasound signals," in *2024 10th IEEE RAS/EMBS International Conference for Biomedical Robotics and Biomechatronics (BioRob)*, pp. 327–332, 2024.

[10] W. Geng, Y. Du, W. Jin, *et al.*, "Gesture recognition by instantaneous surface emg images," *Scientific Reports*, vol. 6, p. 36571, 2016. Demonstrates high-accuracy sEMG-based gesture classification (NinaPro etc.).

[11] J. Liu, S. H. Kang, D. Xu, Y. Ren, S. J. Lee, and L.-Q. Zhang, "Emg-based continuous and simultaneous estimation of arm kinematics in able-bodied individuals and stroke survivors," *Frontiers in Neuroscience*, vol. 11, p. 480, 2017. Example of continuous, simultaneous estimation of arm kinematics from EMG.

[12] F. Xiao *et al.*, "Continuous estimation of joint angle from electromyography using multiple time-delayed features and random forests," *Biomedical Signal Processing and Control*, vol. 43, pp. 236–244, 2018. Shows continuous elbow joint estimation using MTDF features and Random Forests.

[13] D. Sun, A. Cappellari, B. Lan, M. Abayazid, S. Stramigioli, and K. Niu, "Automatic robotic ultrasound for 3d musculoskeletal reconstruction: A comprehensive framework," *Technologies*, vol. 13, no. 2, 2025.

[14] K. Niu, V. Sluiter, B. Lan, J. Homminga, A. Sprengers, and N. Verdonschot, "A method to track 3d knee kinematics by multi-channel 3d-tracked a-mode ultrasound," *Sensors*, vol. 24, no. 8, 2024.

[15] B. Lan, M. Abayazid, N. Verdonschot, S. Stramigioli, and K. Niu, "Deep learning based acoustic measurement approach for robotic applications on orthopedics," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 921–927, 2024.

[16] B. Lan, M. Abayazid, N. Verdonschot, S. Stramigioli, and K. Niu, "Sirc-unet: Improving bone tracking precision of a-mode ultrasound by decoding hierarchical resolution features," *IEEE Sensors Journal*, vol. 24, no. 22, pp. 38174–38184, 2024.

[17] X. Ma *et al.*, "Continuous estimation of knee joint angle based on surface electromyography using a long short-term memory neural network and time-advanced feature," *Sensors*, vol. 20, no. 17, p. 4966, 2020. LSTM-based continuous joint angle estimation for lower limb — another example of regression-style EMG decoding.

[18] S. P. Sitole and F. C. Sup IV, "Continuous prediction of human joint mechanics using emg signals: A review of model-based and model-free approaches," *IEEE Transactions on Medical Robotics and Bionics*, 2023. Review focusing on continuous (regression) decoding from EMG; reports most studies concentrate on classification (88.4% vs 11.6% regression in earlier review) and highlights knowledge gaps.

[19] B. Lan, S. Stramigioli, and K. Niu, "Hierarchical transformer fusion of gaze attention and muscle activity for forearm movement estimation," *IEEE Transactions on Biomedical Engineering*, pp. 1–8, 2025.

[20] B. Lan, E. Juffermans, I. Tamadon, and K. Niu, "A preliminary study of the pulse oximetry for early breast cancer detection," Feb. 2025.

10th Dutch Biomedical Engineering Conference, BME 2025, BME 2025 ; Conference date: 30-01-2025 Through 31-01-2025.

[21] "The study of relationship between muscle activation and deformation during arm motion," Feb. 2025. 10th Dutch Biomedical Engineering Conference, BME 2025, BME 2025 ; Conference date: 30-01-2025 Through 31-01-2025.

[22] M. Vannucci, M. R. Rodríguez-Luna, B. Lan, K. Niu, A. Lapergola, E. Reitano, N. Zorzetti, M. Goglia, D. S. Keller, and S. Perretta, "Sex-based differences in musculoskeletal pain among surgeons: an international survey," *Surgical Endoscopy*, pp. 1–8, 2025.

[23] B. Lan, G. Krijnen, S. Stramigioli, and K. Niu, "Deciphering muscular dynamics: A dual-attention framework for predicting muscle contraction from activation patterns," *IEEE Journal of Biomedical and Health Informatics*, vol. 29, no. 9, pp. 6510–6523, 2025.

[24] D. Sun, S. Stramigioli, and K. Niu, "Hcce-cunet based multi-class musculoskeletal segmentation for robotic ultrasound system," *IEEE Transactions on Medical Robotics and Bionics*, pp. 1–1, 2025.

[25] B. Lan and K. Niu, "Muscle activation–deformation correlation in dynamic arm movements," *J: multidisciplinary scientific journal*, vol. 8, Feb. 2025.

[26] M. Vannucci, D. Sun, B. Lan, K. Niu, A. Riba, N. Heikens, M. Fehervari, M. R. Rodríguez-Luna, and S. Perretta, "Endoscopic sleeve gastroplasty video assessment: do technical features influence esg integrity and weight loss at 6 and 12 months follow-up?," *Surgical Endoscopy*, pp. 1–9, 2025.

[27] G. Y. Ward, D. Sun, and K. Niu, "An autonomous fluoroscopic imaging system for catheter insertions by bilateral control scheme: A numerical simulation study," *Machines*, vol. 13, no. 6, 2025.

[28] B. Lan and K. Niu, "Predicting muscle thickness deformation from muscle activation patterns: A dual-attention framework," *arXiv preprint arXiv:2409.18266*, 2024.

[29] C. Lugaresi, J. Tang, H. Nash, C. McClanahan, E. Uboweja, M. Hays, F. Zhang, C.-L. Chang, M. G. Yong, J. Lee, W.-T. Chang, W. Hua, M. Georg, and M. Grundmann, "Mediapipe: A framework for building perception pipelines," 2019.

[30] E. Olson, "AprilTag: A robust and flexible visual fiducial system," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pp. 3400–3407, IEEE, May 2011.

[31] J. Wang and E. Olson, "AprilTag 2: Efficient and robust fiducial detection," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, October 2016.

[32] H. J. Hermens, B. Freriks, C. Disselhorst-Klug, and G. Rau, "Development of recommendations for semg sensors and sensor placement procedures," *Journal of Electromyography and Kinesiology*, vol. 10, no. 5, pp. 361–374, 2000.

[33] D. Sun, A. Cappellari, B. Lan, and K. Niu, "Automatic robotic ultrasound scanning for muscle segmentation and reconstruction," Feb. 2025. 10th Dutch Biomedical Engineering Conference, BME 2025, BME 2025 ; Conference date: 30-01-2025 Through 31-01-2025.

[34] B. Lan, M. Abayazid, N. Verdonschot, S. Stramigioli, and K. Niu, "Sirc-unet: Improving bone tracking precision of a-mode ultrasound by decoding hierarchical resolution features," *IEEE sensors journal*, vol. 24, pp. 38174–38184, Nov. 2024.

[35] M. Pang, S. Guo, Q. Huang, and et al., "Electromyography-based quantitative representation method for upper-limb elbow joint angle in sagittal plane," *J. Med. Biol. Eng.*, vol. 35, pp. 165–177, 2015.

[36] C. Lin, X. Zhang, and C. Zhao, "A parallel and efficient transformer deep learning network for continuous estimation of hand kinematics from electromyographic signals," *Scientific Reports*, vol. 15, p. 36150, 2025.

[37] Z. Chen, H. Wang, H. Chen, and T. Wei, "Continuous motion finger joint angle estimation utilizing hybrid semg-fmg modality driven transformer-based deep learning model," *Biomedical Signal Processing and Control*, vol. 85, p. 105030, 2023.

[38] C. Lin and X. Zhang, "Fusion inception and transformer network for continuous estimation of finger kinematics from surface electromyography," *Frontiers in Neurorobotics*, vol. Volume 18 - 2024, 2024.

[39] S. Abbas, S. Aslam, K. Berns, and A. Muhammad, "Analysis and improvements in apriltag based state estimation," *Sensors*, vol. 19, no. 24, p. 5480, 2019.

[40] H. Zhang, S. M. Sid'El Moctar, S. Boudaoud, and I. Rida, "A comprehensive review of semg-imu sensor fusion for upper limb movements pattern recognition," *Information Fusion*, vol. 125, p. 103422, 2026.

[41] J. H. Lim, K. He, Z. Yi, C. Hou, C. Zhang, Y. Sui, and L. Li, "Adaptive learning based upper-limb rehabilitation training system with collaborative robot," in *2023 45th Annual International Conference of the IEEE Engineering in Medicine amp; Biology Society (EMBC)*, p. 1–5, IEEE, July 2023.