emg2tendon: From sEMG Signals to Tendon Control in Musculoskeletal Hands

Sagar Verma

Micropilot

Cambridge, MA 02138

Email: sagar@micropilot.org
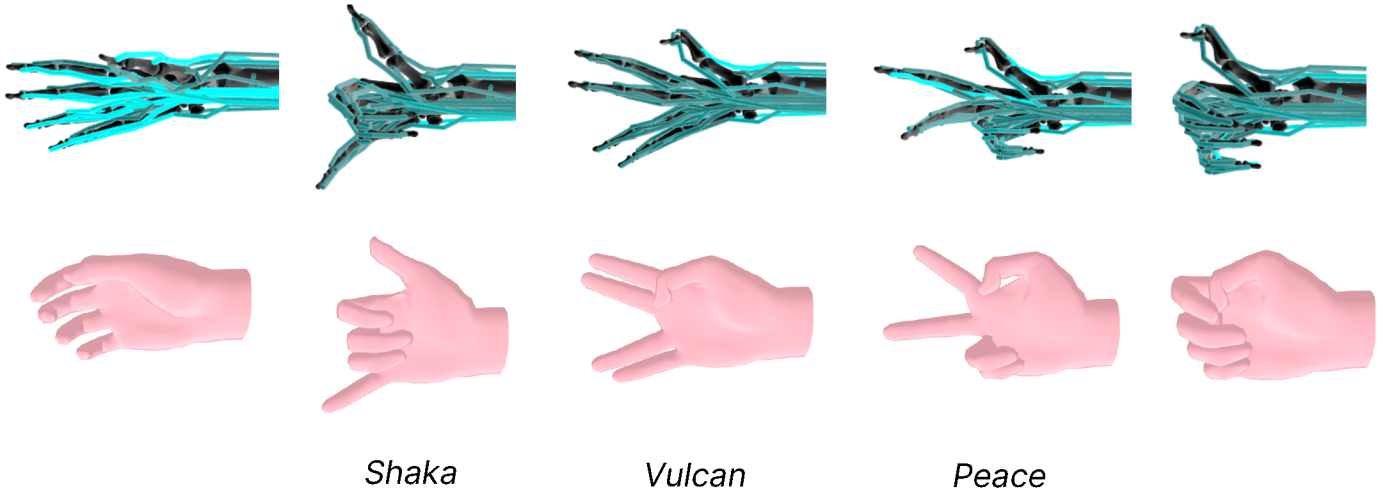
Figure 1. We introduce the *emg2tendon* dataset, an extension of *emg2pose*, for learning tendon control signals from sEMG signals. The bottom row presents hand poses for the gestures ('Shaka,' 'Vulcan,' and 'Peace') sourced from the *emg2pose* dataset, while the top row depicts the corresponding musculoskeletal hand poses generated using the proposed pipeline.

Abstract—Tendon-driven robotic hands offer unparalleled dexterity for manipulation tasks, but learning control policies for such systems presents unique challenges. Unlike joint-actuated robotic hands, tendon-driven systems lack a direct one-to-one mapping between motion capture (mocap) data and tendon controls, making the learning process complex and expensive. Additionally, visual tracking methods for real-world applications are prone to occlusions and inaccuracies, further complicating joint tracking. Wrist-wearable surface electromyography (sEMG) sensors present an inexpensive, robust alternative to capture hand motion. However, mapping sEMG signals to tendon control remains a significant challenge despite the availability of EMG-to-pose data sets and regression-based models in the existing literature.

We introduce the first large-scale EMG-to-Tendon Control dataset for robotic hands, extending the *emg2pose* dataset, which includes recordings from 193 subjects, spanning 370 hours and 29 stages with diverse gestures. This dataset incorporates tendon control signals derived using the MyoSuite MyoHand model, addressing limitations such as invalid poses in prior methods. We provide three baseline regression models to demonstrate *emg2tendon* utility and propose a novel diffusion-based regression model for predicting tendon control from sEMG recordings. This dataset and modeling framework marks a significant step forward for tendon-driven dexterous robotic manipulation, laying the groundwork for scalable and accurate tendon control in robotic hands. https://emg2tendon.github.io/

I. INTRODUCTION

The human hand, with its exceptional dexterity and compliance, serves as an ideal model for robotic hands designed for complex manipulation tasks. From assembling small components in manufacturing to performing intricate surgical procedures, fully anthropomorphic, tendon-driven robotic hands offer the highest level of control and adaptability [49]. However, traditional visual pose-tracking methods, such as motion capture and depth sensors [24, 21, 18], suffer from occlusions, limited field of view, and environmental constraints, making them impractical for continuous real-world use. Wrist-worn surface electromyography (sEMG) sensors provide a viable alternative by capturing neuromuscular signals that drive hand movement [40]. Mapping sEMG to tendon control signals enables natural and responsive robotic hand operation, with significant implications for prosthetics, teleoperation, and human-robot interaction [22].

Anthropomorphic hands require intricate joint mechanisms, including hinges, linkages, and tendon-driven actuators, to achieve human-like movement [44, 48, 52, 20]. Unlike non-anthropomorphic robotic designs [27, 29, 55], which lack a direct mapping to human motion, tendon-driven hands offer

greater biomechanical accuracy and compliance. The human hand's evolved structure is optimized for dexterous interactions, and much of the artificial world is designed to accommodate it. Ensuring a one-to-one mapping between human and robotic hands in mechanical design and control enhances teleoperation, imitation learning, and adaptive control strategies. Advances in imitation learning, particularly diffusion-based models [8, 35, 53], enable more efficient policy learning compared to traditional reinforcement learning [34, 28], allowing robotic hands to perform nuanced, real-time interactions.

sEMG-based sensing eliminates the limitations of vision-based tracking by directly capturing the electrical potentials generated by muscle activations [12, 33]. Rather than mapping sEMG signals directly to joint angles, an alternative and more robust approach is to infer movement through a musculoskeletal hand model that operates based on tendon-driven actuation [49]. This method ensures biologically consistent pose estimation and provides a stable framework for learning and predicting control signals. By integrating sEMG signals with tendon-driven control, robotic hands can achieve a level of dexterity and responsiveness closer to that of the human hand, offering new possibilities for prosthetics, rehabilitation, and teleoperated systems.

In this work, our objective is to advance research on EMG-to-tendon control models for anthropomorphic hand control by extending the existing *emg2pose* dataset. Our contributions are as follows:

- **First large-scale dataset:** We introduce the first large-scale dataset specifically designed for EMG-to-tendon control signals, extending the *emg2pose* dataset with recordings from 193 subjects, 370 hours of data, and 29 stages of diverse gestures.
- **Comprehensive benchmarking:** We provide an extensive benchmarking framework to learn EMG-to-tendon control, tendon control-to-pose and direct EMG-to-pose mappings.
- **Conditional Latent Diffusion Models (CLDM):** We present a novel approach using conditional latent diffusion model to achieve highly accurate musculoskeletal hand control signal prediction from sEMG recordings.

The paper is organized as follows: Section II discusses related work, covering sensors (optical and sEMG) for pose tracking, tendon-based actuation, and control policy learning methods, including diffusion models and reinforcement learning. Section III provides background on musculoskeletal hand models and sEMG, establishing the relationship between sEMG data and muscle movements. Section IV details the extension of the *emg2pose* dataset to *emg2tendon*. Section V introduces the benchmark experiments and explains the conditional latent diffusion model for modeling the EMG-to-tendon relationship. Section VI presents an extensive ablation study comparing EMG-to-tendon-to-pose and direct EMG-to-pose approaches. Section VII discusses the limitations of the proposed dataset and methods. Finally, Section VIII concludes the paper and discusses future research directions.

II. RELATED WORK

A. Pose from Vision

Vision-based pose estimation primarily uses depth, RGB, or both as inputs alongside large open-sourced datasets [14, 36, 2, 3, 47, 50, 1]. Labeling approaches typically rely on marker-based motion capture systems [54], which provide high-quality labels but create distributional shifts during deployment due to the absence of markers in real-world applications. Alternate labeling methods, such as multi-view camera systems [9, 15], synthetic datasets [10], or magnetic sensors [45], often result in lower-quality labels. While motion capture markers are well-suited for obtaining high-quality labels for surface electromyography (sEMG), they do not interfere with the data used for predictions. Vision-based datasets largely focus on static hand poses, interactions with objects, or hand-to-hand interactions.

B. Pose from sEMG

Several studies have investigated hand pose regression using sEMG signals. Liu et al. [26], for example, used the MyoBand to estimate hand poses during various movements, collecting data from 11 participants. Their study assessed sEMG decoding models for hand pose across users and sessions using convolutional [26] and LSTM architectures. Similarly, SensingDynamics [46] utilized a clinic-grade system to collect datasets from 13 participants, employing custom 3D convolutional architectures to predict hand joint angles, landmark positions, and grip forces. Both studies reported low error rates on held-out test sets within participants. However, these datasets are limited in scale, featuring only 11 to 13 participants with short recording durations of 15 to 20 minutes per participant, which restricts generalization across users. In contrast, the *emg2pose* dataset [40], collected using the sEMG-RD wristband from CTRL-labs at Reality Labs et al. [11], stands as the largest dataset available, providing a comprehensive one-to-one mapping between sEMG recordings and hand joint poses.

C. Tendon based Actuation

Musculoskeletal models [30, 25, 41, 13, 43] have been developed to simulate the kinematics of muscles and physiological joints. While these models provide valuable insights, their high computational demands and limited support for contact forces restrict their utility for studying complex hand-object interactions, confining them to optimization-based methods. Recently, the MyoHand model [5, 51] has addressed these limitations. MyoHand supports contact-rich interactions and is computationally efficient for data-driven exploration. Several methods have been developed for MyoHand to perform in-hand manipulation of simple objects such as balls, pens, and cubes [5, 49], demonstrating its capability for tendon-driven manipulation tasks.

D. Control Policies

Data-driven control approaches, particularly Reinforcement Learning (RL), have demonstrated success in solving complex dexterous manipulation tasks using joint-based

control [38, 23, 37, 7]. RL techniques have been applied to enable naturalistic movements, often utilizing motion capture data to learn complex behaviors [32, 31, 16]. For biomechanical models, control policies have been adapted to muscle actuators to produce more naturalistic behaviors. Unlike robotic joint-based control, biomechanical systems face challenges due to their over-actuated control spaces, which hinder exploration efficiency [42]. Machine learning techniques, including direct optimization [4, 39] and deep reinforcement learning [19, 42, 17, 5, 51, 49], have been used to synthesize behaviors such as walking, running, reaching movements, and biped locomotion. While significant progress has been made in in-hand object rotations and stylistic movements [49, 6], complex dexterous hand-object manipulations remain a challenging frontier in this domain.

III. BACKGROUND

We begin with a brief overview of: (i) the physiologically accurate musculoskeletal hand model (ii) the neuro-muscular interactions during finger muscle activations and how it manifests as EMG sensor data.

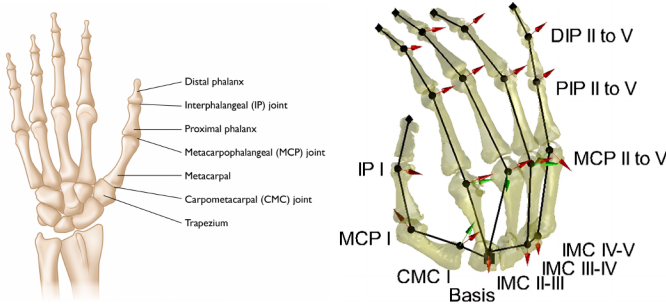### A. Musculoskeletal Hand Model

Figure 2. Anatomical structure of the human hand, showing key joints.

The human hand consists of four fingers and a thumb, each with multiple joints that enable fine motor control and dexterous manipulation. Figure 2 illustrates the skeletal structure of the human hand, highlighting key joints responsible for complex articulation. The four fingers comprise three primary joints: the metacarpophalangeal (MCP), proximal interphalangeal (PIP), and distal interphalangeal (DIP) joints. The PIP and DIP joints each have a single degree of freedom (DoF), allowing flexion and extension, while the MCP joint possesses two DoFs: flexion/extension and adduction/abduction. This results in a total of four DoFs per finger.

The thumb has a distinct anatomical structure compared to the fingers. It includes an interphalangeal (IP) joint with a single DoF for flexion/extension, while its MCP and trapeziometacarpal (TM) joints allow for both flexion/extension and adduction/abduction. Consequently, the thumb has five DoFs, making it essential for opposable grasping. Additionally, wrist motion contributes six DoFs, encompassing translation and rotation. The overall range of
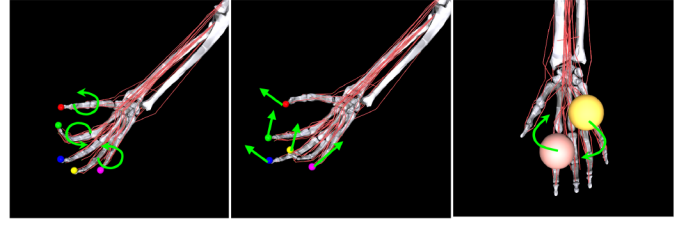

Figure 3. The MyoHand model in MyoSuite. From left to right, green arrows indicate joint articulation, fingertip movement, and in-hand manipulation of baoding balls, showcasing the model's tendon-driven control and dexterous capabilities.

motion varies, with adduction/abduction movements having a more limited range compared to flexion/extension.

To replicate human hand motion in robotic and prosthetic applications, musculoskeletal models such as MyoHand have been developed. Figure 3 shows the MyoSuite MyoHand model, which accurately represents human hand biomechanics, including tendon-driven actuation. Unlike direct joint-controlled robotic hands, MyoHand employs a tendon-based system, where tendons transmit forces to generate joint movements, mimicking biological muscle-tendon dynamics. This allows for greater compliance and adaptability in interaction with objects.

To ensure biomechanical accuracy, MyoHand incorporates constraints derived from human hand kinematics. For example, normal flexion of the DIP joint is often coupled with PIP joint movement, and the MCP joint significantly limits the PIP range of motion. Such constraints are naturally captured in tendon-driven models, making them more effective in replicating human dexterity than traditional joint-based robotic hands. By leveraging inverse dynamics, we use MyoHand to generate tendon control signals corresponding to observed hand poses, providing a biologically plausible pathway for translating sEMG signals into tendon-driven robotic hand movements.

B. EMG Sensor Model

Electromyography (EMG) sensors measure the electrical potentials generated by skeletal muscles during neurological activation. These signals provide essential insights into the temporal dynamics and morphological behavior of motor units responsible for muscle movement [12]. EMG signals play a crucial role in detecting and predicting muscle-driven body movements and are widely used for diagnosing neuromuscular disorders, as well as studying the physiological differences in healthy, aging, or fatigued neuromuscular systems.

To understand the muscular involvement in finger movements, we briefly outline the key muscles responsible for hand articulation. The *Extensor Pollicis Longus* facilitates thumb extension, while the *Abductor Pollicis Proprius* is involved in extending the index finger. The *Extensor Digitorum* controls extension of the four medial fingers, and the *Extensor Digiti Minimi* specifically targets the little finger. Finger adduction and abduction—movements toward or away from the midline—are governed by the *Volar Interossei* and

Dorsal Interossei, which attach to the proximal phalanx and *Extensor Digitorum*. Several additional muscles contribute to broader hand and arm movements, including the *Supinator* for forearm rotation, the *Anconeus* and *Brachioradialis* for elbow control, and the *Extensor Carpi Ulnaris*, *Extensor Carpi Radialis Longus*, and *Extensor Carpi Radialis Brevis* for wrist movement. In this study, we primarily focus on the muscles directly involved in fine finger control.

Some of these muscles are located near the skin surface and can be easily detected using surface EMG (sEMG), while others, such as the *Extensor Indicis*, are deeper within the forearm. Previous research [26] has established a strong correlation between EMG signals and basic hand movements, highlighting the activation patterns of flexors and extensors. Additionally, studies have explored biological models for EMG signal generation in response to muscle activations, though a detailed discussion of these models is beyond the scope of this work.

IV. PROPOSED DATASET

In this section, we begin by summarizing the existing *emg2pose* dataset, which serves as the foundation for our work. We then detail the process of generating *emg2tendon*.

A. emg2pose Dataset: Brief Overview

Here is the revised version, condensed into three paragraphs while maintaining clarity and key details:

In this paper, we build upon the *emg2pose* dataset introduced by Salter et al. [40], the largest publicly available sEMG dataset to date. This dataset was collected using a 16-channel bipolar sEMG-RD wristband developed by CTRL-labs at Reality Labs [11], alongside a 26-camera motion capture array operating at 60 Hz (Prime 13W OptiTrack). Participants wore sEMG-RD bands on both wrists and 19 motion capture markers on each hand, strategically placed at key joint locations, including the fingernail bases, between the DIP, PIP, and MCP joints, and at essential points on the thumb and dorsal hand. This setup enabled simultaneous high-resolution recording of muscle activity and precise hand motion tracking.

sEMG data were sampled at 2 kHz, while motion capture data were recorded at 60 Hz and later processed using an inverse dynamics solver to reconstruct hand joint angles. Due to simultaneous marker occlusions, 12.7% of the frames could not be processed and were interpolated to match the 2 kHz sEMG sampling rate. Participants performed structured 45 to 120-second sessions containing 3-5 gestures in randomized order or unconstrained freeform movements, capturing a wide range of hand motions. Most participants contributed four sessions, with a small subset completing three. The *emg2pose* dataset includes 193 participants, covering 370 hours, 751 sessions, and 29 distinct gesture categories.

B. Pose to Tendon Control Signals

There are no direct pose-to-tendon or EMG-to-tendon datasets available, as capturing tendon movement in vivo is highly challenging. However, biophysical models describing the kinematics of tendon-driven finger joint motion exist in the literature [51]. To overcome this limitation, we use the inverse dynamics of the MyoSuite hand model to compute tendon control signals corresponding to the pose data in the *emg2pose* dataset.

1) Muscle Actuator Dynamics: In biomechanics, a muscle-tendon actuator consists of a muscle and a tendon connected in series. MuJoCo implements this by modeling the tendon as the entity with spatial properties (such as length and velocity), while the muscle is represented as a force-generating mechanism that applies tension to the tendon. Thus, in MuJoCo, the tendon length corresponds to the sum of the biological muscle and tendon lengths:

$$\text{actuator length} = L_T + L_M \quad (1)$$

where L_T is the biological tendon length, assumed inelastic, and L_M is the biological muscle length, which varies dynamically.

The optimal resting muscle length, denoted as L_0, is the length at which the muscle produces its peak active force at zero velocity. Due to the complexity of tendon routing and wrapping, directly specifying L_0 and L_T is impractical. Instead, these parameters are computed automatically using predefined scaling relationships. The scaled muscle length and velocity are given by:

$$L = \frac{\text{actuator length} - L_T}{L_0}, \quad V = \frac{\text{actuator velocity}}{L_0} \quad (2)$$

where the range constraints for muscle length, typically set as $(0.75, 1.05)$, are used to determine L_0 and L_T.

The force-length-velocity (FLV) function models the active and passive forces generated by the muscle:

$$FLV(L, V, act) = F_L(L) \cdot F_V(V) \cdot act + F_P(L) \quad (3)$$

where: - $F_L(L)$ represents active force as a function of muscle length, - $F_V(V)$ represents active force as a function of velocity, - act is the activation level of the muscle, and - $F_P(L)$ accounts for passive forces that are independent of activation.

The total muscle force applied to the actuator is:

$$\text{actuator force} = -FLV(L, V, act) \cdot F_0 \quad (4)$$

where F_0 is the peak active force at zero velocity, related to the muscle's physiological cross-sectional area (PCSA). If unknown, F_0 is set to -1 by default and scaled using:

$$F_0 = \frac{\text{scale}}{\text{actuator acc}_0} \quad (5)$$

where actuator acc_0 represents the joint acceleration caused by unit force applied to the actuator transmission.

The activation state of a muscle is governed by a first-order nonlinear filter that depends on the control signal:
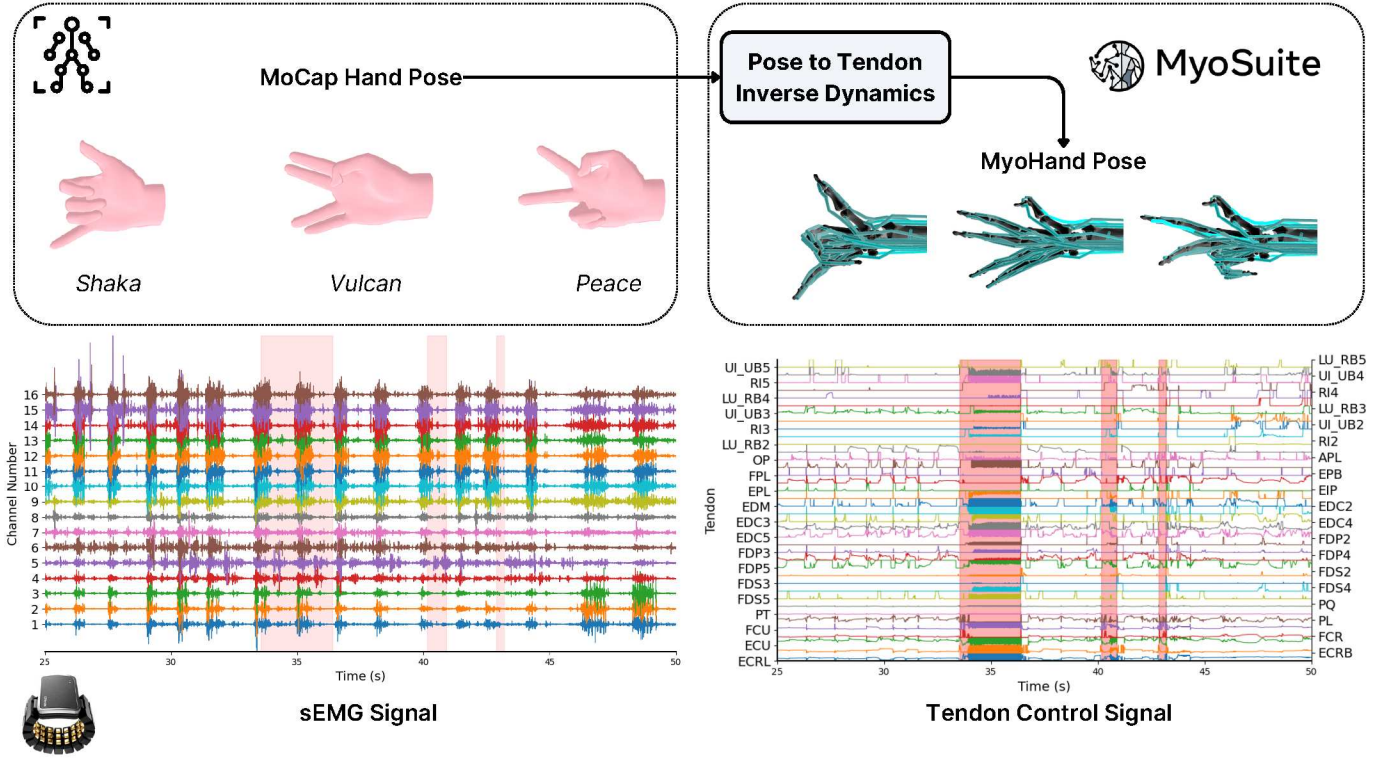
Figure 4. Overview of the *emg2tendon* data collection and processing pipeline. The top-left section illustrates the *emg2pose* dataset collection, where hand gestures ('Shaka,' 'Vulcan,' and 'Peace') are captured using a motion capture (MoCap) system and a wrist-worn sEMG device. The top-right section represents the MyoSuite-based musculoskeletal hand model, which enables pose-to-tendon control mapping through inverse dynamics. The bottom-left plot shows the recorded sEMG signals from 16 channels, while the bottom-right plot depicts the computed 39 channel tendon control signals.

$$\frac{d}{dt}act = \frac{ctrl - act}{\tau(ctrl, act)} \tag{6}$$

where $ctrl$ represents the input control signal, clamped within $[0, 1]$. The activation time constant $\tau(ctrl, act)$ is defined as:

$$\tau(ctrl, act) = \begin{cases} \tau_{act} \cdot (0.5 + 1.5 \cdot act), & ctrl - act > 0 \\ \tau_{deact} \cdot (0.5 + 1.5 \cdot act), & ctrl - act \leq 0 \end{cases} \tag{7}$$

where τ_{act} and τ_{deact} represent the activation and deactivation time constants, respectively. To ensure smooth transitions, MuJoCo introduces an optional smoothing parameter, τ_{smooth}, which interpolates between activation and deactivation dynamics using a sigmoid function:

$$s(x) = \begin{cases} 0, & x \leq 0 \\ 6x^5 - 15x^4 + 10x^3, & 0 < x < 1 \\ 1, & x \geq 1 \end{cases} \tag{8}$$

2) Computation of control signals: The control signal $ctrl$ is determined by solving the following muscle actuator equation:

$$AM \cdot \left(\text{gain} \odot \left(act + \text{timestep} \cdot \frac{ctrl - act}{\tau} \right) + \text{bias} \right) - q_{\text{frc}} = 0 \tag{9}$$

where: - AM is the actuator transmission matrix, - gain and bias represent actuator gain and bias, - q_{frc} is the joint force, and - τ is computed as:

$$\tau = \tau_D + (\tau_A - \tau_D) \cdot \text{sigmoid}\left(\frac{ctrl - act}{\tau_{\text{smooth}}} + 0.5 \right) \tag{10}$$

To efficiently solve for $ctrl$, we reformulate the equation as a quadratic program (QP):

$$\min_x \frac{1}{2}x^T P x + q^T x, \quad \text{s.t.} \quad lb \leq x \leq ub \tag{11}$$

where the sigmoid function is approximated to enable efficient optimization. Using this approximation, the equation is rewritten as:

$$AM \cdot x + k = 0 \tag{12}$$

where:

$$x = \left(\text{timestep} \cdot \text{gain} \odot \frac{ctrl - act}{(ctrl - act) \cdot \tau_1 + \tau_2} \right) \tag{13}$$

$$k = AM \cdot (\text{gain} \odot act) + AM \cdot \text{bias} - q_{\text{frc}} \qquad (14)$$

From the QP formulation, we define:

$$P = 2 \cdot AM^T \cdot AM, \quad q = 2 \cdot AM^T \cdot k \qquad (15)$$

The bounds for optimization are:

$$lb = \text{timestep} \cdot \text{gain} \odot \frac{1 - act}{(1 - act) \cdot \tau_1 + \tau_2} \qquad (16)$$

$$ub = \text{timestep} \cdot \text{gain} \odot \frac{-act}{-act \cdot \tau_1 + \tau_2} \qquad (17)$$

After solving for x, the control signal $ctrl$ is computed as:

$$ctrl = act + \frac{x \cdot \tau_2}{\text{timestep} \cdot \text{gain} - x \cdot \tau_1} \qquad (18)$$

This framework enables us to compute tendon control signals for the *emg2pose* dataset, allowing its extension into *emg2tendon* for tendon-driven robotic hand control.

C. Tendon Control Signals Generation

Figure 4 provides an overview of the *emg2tendon* data processing workflow. The *emg2pose* dataset provides hand pose recordings with 20 joints, whereas the MyoHand musculoskeletal model [5] consists of 23 joints. To ensure compatibility between these datasets, we mapped all finger joints from *emg2pose* to MyoHand, covering five fingers, each with four joints. However, three wrist-related joints—pro-supination (pro_sup), radial-ulnar deviation (deviation), and flexion-extension (flexion)—are excluded, as they are not explicitly modeled in *emg2pose*.

To compute tendon control signals, we applied inverse dynamics to all pose recordings in *emg2pose* using the Myo-Hand model, yielding 39 tendon control signals. However, in 21 sessions, the IK solver failed due to multiple occlusions in the pose data, resulting in infeasible joint configurations for MyoHand. To expedite these computations, we leveraged the MyoSuite MyoHand model, which we observed to run approximately five times faster on Apple Silicon compared to x86_64 or ARM-based systems. Using a distributed setup of 15 Mac Mini devices (4 × M1, 6 × M2, 5 × M3), we processed inverse dynamics across all 25,254 sessions, completing the full dataset transformation in 46 hours.

To align the data format with MyoHand's simulation requirements, we downsampled the pose recordings from 2 kHz to 500 Hz before running the IK solver. The computed tendon control signals were then upsampled back to 2 kHz to match the original *emg2pose* dataset, ensuring that sEMG signals, joint poses, and tendon control forces remain synchronized at a uniform sampling rate. This pipeline facilitates seamless model training and analysis by providing a structured approach to tendon-driven hand motion estimation.

V. EXPERIMENTS

In this section, we demonstrate the applicability of the proposed *emg2tendon* dataset by training regression models to predict tendon control signals from sEMG recordings. To evaluate the effectiveness of this approach, we compare it with EMG-to-pose regression and tendon-to-pose regression models. This enables us to assess the impact of an intermediate tendon representation on final pose estimation by analyzing the difference between direct EMG-to-pose mapping and a two-step approach (EMG → tendon → pose).

Additionally, we evaluate an alternative physics-informed setup, where predicted tendon control signals from the regression model are applied to the MyoSuite MyoHand model to generate hand pose estimations through simulation. This allows us to compare learned regression models with biomechanically accurate physics-based pose estimation.

For these experiments, we use three baseline models from *emg2pose*: Time-Depth Separable Convolution (TDS) Network, NeuroPose, and SensingDynamics. We also introduce a Conditional Latent Diffusion Model (CLDM) as our proposed approach for EMG-to-tendon regression.

A. Conditional Latent Diffusion Model (CLDM)

Several methods have been developed to apply diffusion models to time-series data, particularly in healthcare applications. We follow a similar approach, structuring the training into three main stages. First, we train two autoencoders with Kullback-Leibler (KL) regularization to compress data into a latent vector representation. One autoencoder is designed for the output signal (tendon control signals), while the other is for the input signal (sEMG recordings), which serves as the conditioning input for the diffusion model. Figure 5 shows the CLDM architecture including the autoencoders.

Once trained, we use the frozen encoder of the output autoencoder to extract latent vectors, which are then processed through a diffusion model. The diffusion process involves gradually adding noise to the latent representation over multiple time steps and then using a learned reverse process to reconstruct the original data. The reverse diffusion process is conditioned on the latent encoding of the input signal, ensuring that the model learns meaningful correlations between sEMG and tendon control signals. The final output is decoded using the decoder of the output autoencoder, reconstructing the predicted tendon control signals. This approach ensures that both input and output signals share a common latent space, effectively capturing the relationship between muscle activations and tendon-driven hand movements.

During inference, we generate a random noise sample from a Gaussian distribution, learned during training, and condition it using the latent vector from the input autoencoder. This conditioned latent representation is then passed through the reverse diffusion process, and the final predicted tendon control signals are obtained by decoding the output.

The autoencoder architecture consists of an Encoder and Decoder, both featuring multiple layers of ResNet Blocks, Attention Blocks, and Downsampling/Upsampling Blocks. For
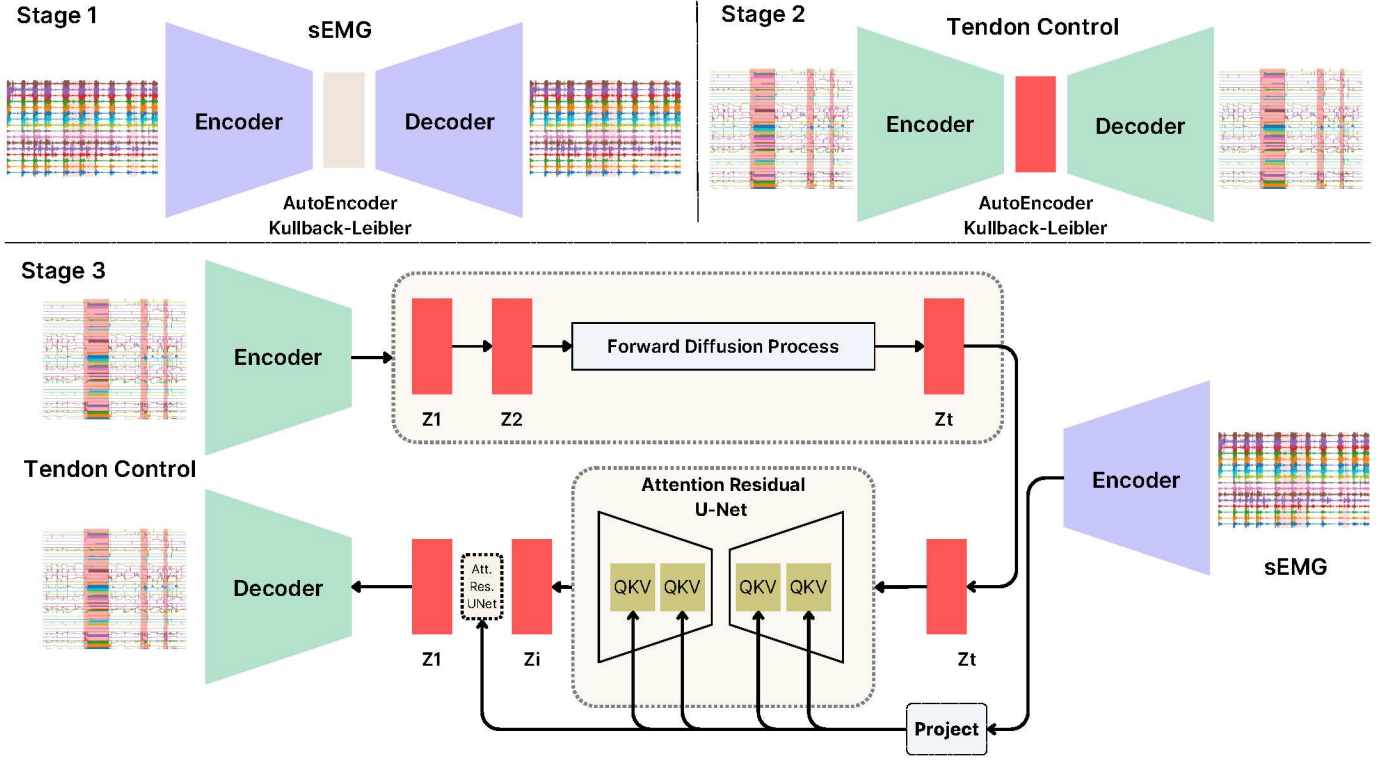
Figure 5. Training stages of Conditional Latent Diffusion Model (CLDM) to predict tendon control signals from sEMG recordings.

the diffusion process, we employ a UNet-based architecture, which effectively refines the latent representations over multiple diffusion steps.

B. Networks Training

To evaluate different approaches for predicting hand pose and tendon control signals from sEMG signals, we train multiple neural network architectures under three key tasks.

1) **Pose regression from EMG:** We train Time-Depth Separable (TDS) Network, NeuroPose, and Sensing-Dynamics to directly predict joint poses from EMG signals. Additionally, we introduce a Conditional Latent Diffusion Model (CLDM) where two autoencoders are first trained—one for joint poses and another for sEMG recordings. The CLDM is then trained to predict joint poses conditioned on the latent representation of EMG.

2) **Tendon control signals regression from EMG:** We train the same baseline models (TDS, NeuroPose, and SensingDynamics) to predict tendon control signals from EMG recordings. Similar to pose regression, we train an autoencoder specifically for tendon control signals and then apply CLDM to model tendon control signal prediction conditioned on EMG signals.

3) **Pose regression from tendon control signals:** Here, we train TDS, NeuroPose, and SensingDynamics to predict joint poses from tendon control signal data. Additionally, we train a CLDM model, where tendon control signals serve as the conditioning input for predicting hand pose.

We train all three autoencoders using a 16-channel latent space with a four-layer encoder-decoder structure $[32, 64, 64, 64]$ and attention at all levels. Using these pre-trained encoders, we train three CLDMs. The CLDM architecture employs a UNet-based diffusion process with a sample size of 4000, extra conditioning channels set to 16, and downsampling across four blocks $[32, 64, 128, 256]$. The noise schedule follows a linear progression with 1000 timesteps ($\beta_{start} = 0.0001, \beta_{end} = 0.02$), and training is performed using the Adam optimizer (10^{-4} learning rate) with a Reduce-on-Plateau scheduler. For baseline models (TDS, NeuroPose, SensingDynamics), we adopt the same architecture configurations and training settings as in *emg2pose*. All models are trained on a single NVIDIA GH200 system with a 96GB H100 GPU, 64-core ARM64 CPU, and 463.9GB RAM running Ubuntu 22.04.5. To evaluate performance, we use angular error as the primary metric for pose regression. For the prediction of the tendon control signal, we report the root mean square error (RMSE) and the mean absolute error (MAE).

VI. RESULTS

In this section, we present the evaluation results of our trained models for predicting both tendon control signals and joint poses from sEMG signals. We analyze the performance of different regression approaches and assess the impact of using tendon control signals as an intermediate representation for pose estimation.

We begin by reporting results for direct EMG-to-pose

Test Set	Network	sEMG-to-Pose	Tendon-to-Pose	Two-Step	Two-Step Simulation
User	SensingDynamics	15.5 ± 1.4	14.6 ± 1.2	18.4 ± 1.8	16.2 ± 1.5
	NeuroPose	13.2 ± 1.1	12.4 ± 1.3	16.7 ± 1.3	14.4 ± 1.1
	TDS	12.2 ± 1.3	11.1 ± 1.1	14.5 ± 1.4	13.6 ± 1.3
	CLDM	**11.3 ± 1.0**	**10.8 ± 0.9**	**12.6 ± 1.2**	**11.9 ± 1.3**
Stage	SensingDynamics	18.8 ± 1.6	17.6 ± 1.8	20.9 ± 1.5	19.1 ± 1.7
	NeuroPose	17.2 ± 1.7	16.2 ± 1.6	19.8 ± 1.7	17.8 ± 1.5
	TDS	15.2 ± 1.6	14.3 ± 1.4	17.9 ± 1.6	16.1 ± 1.6
	CLDM	**14.3 ± 1.5**	**12.5 ± 1.3**	**16.2 ± 1.5**	**14.9 ± 1.3**
User, Stage	SensingDynamics	18.7 ± 1.6	17.9 ± 1.5	21.3 ± 1.9	19.2 ± 1.7
	NeuroPose	17.5 ± 1.5	16.5 ± 1.7	21.1 ± 1.7	18.3 ± 1.6
	TDS	15.8 ± 1.4	14.5 ± 1.6	18.7 ± 1.4	16.4 ± 1.5
	CLDM	**14.7 ± 1.4**	**13.1 ± 1.4**	**16.9 ± 1.3**	**15.2 ± 1.2**

Table I. Comparison of mean angular error (lower is better) for different pose estimation methods. CLDM consistently outperforms the three baselines, achieving the most accurate pose predictions. Bold values indicate statistically significant improvements.

regression, followed by EMG-to-tendon control signal regression and tendon control signal-to-pose regression. Next, we evaluate the two-step regression approach (EMG → tendon control signals → pose) to determine whether tendon control signals provide a meaningful intermediary representation for improved pose estimation. Finally, we compare this method with a physics-informed approach, where the predicted tendon control signals are applied to the MyoSuite MyoHand model to generate pose estimations through simulation.

Table I presents the pose estimation results for different models trained to predict joint angles using five different approaches: 1) pose regression from sEMG signals (direct EMG-to-Pose mapping), 2) pose regression from tendon control signals (tendon-to-pose mapping), 3) two-step pose estimation: predicting tendon control signals from EMG, followed by joint pose estimation, and 4) two-step pose estimation with simulation: using MyoSuite's physics-based model to infer joint poses from predicted tendon control signals.

The table reports mean angular error ± standard deviation across users, computed in the same manner as *emg2pose*. Bold values indicate statistical significance under a Wilcoxon signed-rank test, comparing *emg2pose* against NeuroPose and SensingDynamics.

Table II shows Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) for networks trained to predict tendon control signals from sEMG recordings. These models are critical in both two-step regression variants, as they enable tendon control signal estimation from EMG before final pose inference.

A. Pose Regression from sEMG Signals

Directly predicting hand pose from sEMG signals shows CLDM achieving the lowest angular error (11.3° ± 1.0°) across users, significantly outperforming all baseline models. Compared to TDS (12.2° ± 1.3°) and NeuroPose (13.2° ±

Test Set	Network	RMSE	MAE
User	SensingDynamics	0.243	0.169
	NeuroPose	0.226	0.155
	TDS	0.215	0.150
	CLDM	**0.201**	**0.139**
Stage	SensingDynamics	0.281	0.182
	NeuroPose	0.274	0.163
	TDS	0.263	0.157
	CLDM	**0.249**	**0.151**
User, Stage	SensingDynamics	0.285	0.187
	NeuroPose	0.281	0.171
	TDS	0.269	0.165
	CLDM	**0.253**	**0.159**

Table II. Comparison of RMSE and MAE for tendon force prediction models. CLDM achieves the lowest errors, demonstrating superior accuracy compared to the three baselines. Bold values indicate statistically significant improvements.

1.1°), the diffusion-based model benefits from better temporal smoothing and uncertainty modeling, reducing errors by 7.4% over TDS and 14.3% over NeuroPose.

B. Tendon Control Signals Regression from sEMG Signals

CLDM also achieves the best RMSE (0.201) and MAE (0.139) for tendon control signal regression, outperforming NeuroPose (0.226 RMSE, 0.155 MAE) and TDS (0.215 RMSE, 0.150 MAE). This suggests that CLDM captures fine-grained muscle activation patterns more effectively, leading to better tendon control signal predictions that enhance two-step pose estimation models.

C. Pose Regression from Tendon Control Signals

When predicting joint angles from tendon control signals, CLDM again outperforms baselines with an angular error of 10.8° ± 0.9°, demonstrating that it learns a more accurate tendon-to-pose mapping than TDS (11.1° ± 1.1°) and NeuroPose (12.4° ± 1.3°).

D. Two-Step Pose Estimation: EMG → Tendon Control Signals → Pose

Using tendon control signals as an intermediary, CLDM achieves 12.6° ± 1.2°, which is lower than direct EMG-to-Pose regression using TDS (14.5° ± 1.4°). This suggests that tendon control signals act as a meaningful latent representation, improving pose estimation.

E. Two-Step Pose Estimation with Physics-Based Simulation: EMG → Tendon Control Signals → Simulated Pose

Incorporating physics-based constraints in MyoSuite, CLDM achieves the best results (11.9° ± 1.3°), which is a 5.6% improvement over TDS (13.6° ± 1.3°) and a 7.6% improvement over NeuroPose (14.4° ± 1.1°). This demonstrates the importance of integrating learned tendon control signals with biomechanical simulation for achieving realistic and accurate pose estimation.

The results demonstrate that CLDM consistently outperforms all baseline models across all five evaluation settings. Direct EMG-to-Pose regression benefits from the diffusion process's ability to capture temporal dependencies, leading to lower angular error compared to conventional deep learning architectures. Using tendon control signals as an intermediary representation further improves pose estimation accuracy, highlighting the importance of tendon-driven modeling for robotic hand control. The two-step regression approach (EMG → tendon control signals → pose) shows that incorporating tendon control as an intermediate layer enhances precision over direct pose regression. Furthermore, integrating biomechanical constraints through MyoSuite's physics-based simulation significantly improves final pose estimation, reducing errors caused by overfitting and unrealistic joint kinematics. CLDM-based models achieve the lowest errors across all tasks, underscoring their effectiveness in learning the complex relationships between muscle activations and tendon control signals. These findings indicate that the use of latent diffusion and tendon-based modeling can drive substantial progress in tendon-driven robotic hand control, prosthetics, and teleoperation applications.

VII. LIMITATIONS

Through our experiments, we identified two main limitations that highlight potential areas for improvement in our approach. First, our method heavily relies on accurate pose estimation and inverse dynamics for mapping joint poses to tendon control signals. The *emg2pose* dataset already exhibits a 12.7% failure rate in its inverse dynamics solver due to marker occlusions and infeasible joint positions. Additionally, failures in the pose-to-tendon mapping further compound the

issue. These limitations stem primarily from data collection constraints rather than the model itself.

Second, the process of obtaining tendon control signals from joint poses via inverse dynamics applied to a simulated musculoskeletal hand model inherently places the dataset within a synthetic space. While the MyoHand model serves as a useful approximation, a definitive evaluation of our system requires real-world testing on physical tendon-driven robotic hands. The current MyoHand model has been simplified to balance computational efficiency and interpretability, but it omits several secondary muscles that contribute to the full range of motion of the human hand. This simplification may introduce discrepancies between the predicted tendon forces and those observed in actual biomechanical systems.

VIII. CONCLUSION

In this work, we extend the *emg2pose* dataset to create the first large-scale dataset for mapping sEMG signals to tendon control signals, *emg2tendon*. We establish three baseline models on this dataset and evaluate their performance in predicting tendon control signals. Additionally, we introduce a diffusion-based method that demonstrates strong performance not only for *emg2pose* but also for the proposed *emg2tendon* mapping problem, highlighting its potential for accurate and robust tendon-driven hand control.

Moving forward, we aim to validate our approach on real tendon-driven robotic hands to bridge the gap between synthetic and physical systems. Additionally, we seek to explore reinforcement learning (RL) techniques to enhance the performance of diffusion-based models for tendon control signal prediction. Lastly, we plan to extend the dataset to include interactions between hands and objects, further improving its applicability to real-world dexterous manipulation tasks.

REFERENCES

[1] Adnane Boukhayma and Rodrigo de Bem and Philip H. S. Torr. 3D Hand Shape and Pose From Images in the Wild. *Computer Vision and Pattern Recognition*, 2019.

[2] Adrian Spurr and Jie Song and Seonwook Park and Otmar Hilliges. Cross-Modal Deep Variational Hand Pose Estimation. *Computer Vision and Pattern Recognition*, 2018.

[3] Adrian Spurr and Umar Iqbal and Pavlo Molchanov and Otmar Hilliges and Jan Kautz. Weakly Supervised 3D Hand Pose Estimation via Biomechanical Constraints. *ArXiv*, abs/2003.09282, 2020.

[4] Al Borno, Mazen and Vyas, Saurabh and Shenoy, Krishna V and Delp, Scott L. High-fidelity musculoskeletal modeling reveals that motor planning variability contributes to the speed-accuracy tradeoff. *eLife*, 2020.

[5] Caggiano, Vittorio and Wang, Huawei and Durandau, Guillaume and Sartori, Massimo and Kumar, Vikash. MyoSuite–A contact-rich simulation suite for musculoskeletal motor control. *arXiv:2205.13600*, 2022.

[6] Cameron Berg and Vittorio Caggiano and Vikash Kumar. SAR: Generalization of Physiological Agility and

Dexterity via Synergistic Action Representation. *Auton. Robots*, 48, 2023.

[7] Chen, Tao and Xu, Jie and Agrawal, Pulkit. A system for general in-hand object re-orientation. In *Conference on Robot Learning*, 2022.

[8] Cheng Chi and Siyuan Feng and Yilun Du and Zhenjia Xu and Eric A. Cousineau and others. Diffusion Policy: Visuomotor Policy Learning via Action Diffusion. *ArXiv*, abs/2303.04137, 2023.

[9] Christiane Zimmermann and Duygu Ceylan and Jimei Yang and Bryan C. Russell and Max Argus and others. FreiHAND: A Dataset for Markerless Capture of Hand Pose and Shape From Single RGB Images. *International Conference on Computer Vision*, 2019.

[10] Christiane Zimmermann and Thomas Brox. Learning to Estimate 3D Hand Pose from Single RGB Images. *International Conference on Computer Vision*, 2017.

[11] CTRL-labs at Reality Labs and Sussillo, David and Kaifosh, Patrick and Reardon, Thomas. A generic noninvasive neuromotor interface for human-computer interaction. *bioRxiv*, 2024.

[12] Dan Stashuk. EMG signal decomposition: how can it be accomplished and used? *Journal of Electromyography and Kinesiology*, 11(3), 2001.

[13] Delp, Scott L and Anderson, Frank C and Arnold, Allison S and Loan, Peter and Habib, Ayman and others. OpenSim: open-source software to create and analyze dynamic simulations of movement. *IEEE transactions on biomedical engineering*, 54(11), 2007.

[14] Franziska Mueller and Dushyant Mehta and Oleksandr Sotnychenko and Srinath Sridhar and Dan Casas and others. Real-Time Hand Tracking under Occlusion from an Egocentric RGB-D Sensor. *International Conference on Computer Vision*, 2017.

[15] Gyeongsik Moon and Shunsuke Saito and Weipeng Xu and Rohan Joshi and Julia Buffalini and others. A Dataset of Relighted 3D Interacting Hands. *ArXiv*, abs/2310.17768, 2023.

[16] Hasenclever, Leonard and Pardo, Fabio and Hadsell, Raia and Heess, Nicolas and Merel, Josh. CoMic: Complementary task learning & mimicry for reusable skills. In *International Conference on Machine Learning*, 2020.

[17] Ikkala, Aleksi and Fischer, Florian and Klar, Markus and Bachinski, Miroslav and Fleig, Arthur and others. Breathing life into biomechanical user models. In *ACM Symposium on User Interface Software and Technology*, 2022.

[18] Jiayi Wang and Franziska Mueller and Florian Bernard and Suzanne Sorli and Oleksandr Sotnychenko and others. RGB2Hands. *ACM Transactions on Graphics*, 39, 2020.

[19] Joos, Emanuel and Péan, Fabien and Goksel, Orcun. Reinforcement learning of musculoskeletal control from functional simulations. In *Medical Image Computing and Computer Assisted Intervention*, 2020.

[20] Kenneth Shaw and Ananye Agarwal and Deepak Pathak. LEAP Hand: Low-Cost, Efficient, and Anthropomorphic Hand for Robot Learning. *ArXiv*, abs/2309.06440, 2023.

[21] Kourosh Darvish and Luigi Penco and João Ramos and Rafael Cisneros and Jerry E. Pratt and others. Teleoperation of Humanoid Robots: A Survey. *IEEE Transactions on Robotics*, 39, 2023.

[22] Krasoulis, Agamemnon and Kyranou, Iris and Erden, Mustafa Suphi and Nazarpour, Kianoush and Vijayakumar, Sethu. Improved prosthetic hand control with concurrent use of myoelectric and inertial measurements. *Journal of NeuroEngineering and Rehabilitation*, 2017.

[23] Kumar, Vikash. *Manipulators and Manipulation in high dimensional spaces*. PhD thesis, 2016.

[24] Laura Carreras. Increasing Haptic Fidelity and Ergonomics in Teleoperated Surgery. 2012.

[25] Lee, Jong Hwa and Asakawa, Deanna S and Dennerlein, Jack T and Jindrich, Devin L. Finger muscle attachments for an OpenSim upper-extremity model. *PloS one*, 10(4), 2015.

[26] Liu, Yilin and Zhang, Shijia and Gowda, Mahanth. NeuroPose: 3D Hand Pose Tracking using EMG Wearables. In *Proceedings of the Web Conference 2021*, 2021.

[27] Ma, Raymond and Dollar, Aaron. Yale OpenHand Project: Optimizing Open-Source Hand Designs for Ease of Fabrication and Adoption. *IEEE Robotics & Automation Magazine*, 24(1), 2017.

[28] Marcin Andrychowicz and Bowen Baker and Maciek Chociej and Rafal Józefowicz and Bob McGrew and others. Learning dexterous in-hand manipulation. *The International Journal of Robotics Research*, 39, 2018.

[29] McCann, Connor and Patel, Vatsal and Dollar, Aaron. The Stewart Hand: A Highly Dexterous, Six-Degrees-of-Freedom Manipulator Based on the Stewart-Gough Platform. *IEEE Robotics & Automation Magazine*, 28 (2), 2021.

[30] McFarland, Daniel C and Binder-Markey, Benjamin I and Nichols, Jennifer A and Wohlman, Sarah J and De Bruin, Marije and others. A musculoskeletal model of the hand and wrist capable of simulating functional tasks. *IEEE transactions on biomedical engineering*, 70(5), 2022.

[31] Merel, Josh and Hasenclever, Leonard and Galashov, Alexandre and Ahuja, Arun and Pham, Vu and others. Neural probabilistic motor primitives for humanoid control. *arXiv:1811.11711*, 2018.

[32] Merel, Josh and Tassa, Yuval and TB, Dhruva and Srinivasan, Sriram and Lemmon, Jay and others. Learning human behaviors from motion capture by adversarial imitation. *arXiv:1707.02201*, 2017.

[33] Roberto Merletti and Dario Farina, editors. *Surface Electromyography: Physiology, Engineering, and Applications*. 2016.

[34] Michael Ahn and Henry Zhu and Kristian Hartikainen and Hugo Ponte and Abhishek Gupta and others. ROBEL: Robotics Benchmarks for Learning with Low-Cost Robots. In *Conference on Robot Learning*, 2019.

[35] Moritz Reuss and Maximilian Xiling Li and Xiaogang Jia and Rudolf Lioutikov. Goal-Conditioned Imitation Learning using Score-based Diffusion Policies. *ArXiv*, abs/2304.02532, 2023.

[36] Mueller, Franziska and Bernard, Florian and Sotnychenko, Oleksandr and Mehta, Dushyant and Sridhar, Srinath and others. GANerated Hands for Real-Time 3D Hand Tracking from Monocular RGB. In *Computer Vision and Pattern Recognition*, 2018.

[37] Nagabandi, Anusha and Konolige, Kurt and Levine, Sergey and Kumar, Vikash. Deep dynamics models for learning dexterous manipulation. In *Conference on Robot Learning*, 2020.

[38] Rajeswaran, Aravind and Kumar, Vikash and Gupta, Abhishek and Vezzani, Giulia and Schulman, John and others. Learning complex dexterous manipulation with deep reinforcement learning and demonstrations. *arXiv:1709.10087*, 2017.

[39] Rückert, Elmar and d'Avella, Andrea. Learned parametrized dynamic movement primitives with shared synergies for controlling robotic and musculoskeletal systems. *Frontiers in computational neuroscience*, 7, 2013.

[40] Sasha Salter and Richard Warren and Collin Schlager and Adrian Spurr and others. emg2pose: A Large and Diverse Benchmark for Surface Electromyographic Hand Pose Estimation. *ArXiv*, abs/2412.02725, 2024.

[41] Saul, Katherine R and Hu, Xiao and Goehler, Craig M and others. Benchmarking of dynamic simulation predictions in two software platforms using an upper limb musculoskeletal model. *Computer methods in biomechanics and biomedical engineering*, 18(13), 2015.

[42] Schumacher, Pierre and Häufle, Daniel and Büchler, Dieter and Schmitt, Syn and Martius, Georg. Dep-rl: Embodied exploration for reinforcement learning in overactuated and musculoskeletal systems. *arXiv:2206.00484*, 2022.

[43] Seth, Ajay and Hicks, Jennifer L and Uchida, Thomas K and Habib, Ayman and Dembia, Christopher L and others. OpenSim: Simulating musculoskeletal dynamics and neuromuscular control to study human and animal movement. *PLoS Computational Biology*, 14(7), 2018.

[44] Shadow Robot Company. Shadow Dexterous Hand. https://www.shadowrobot.com/dexterous-hand-series/, 2023.

[45] Shanxin Yuan and Qianru Ye and Björn Stenger and Siddhant Jain and Tae-Kyun Kim. BigHand2.2M Benchmark: Hand Pose Dataset and State of the Art Analysis. *Computer Vision and Pattern Recognition*, 2017.

[46] Sîmpetru, Raul and Arkudas, Andreas and Braun, Dominik and others. Sensing the Full Dynamics of the Human Hand with a Neural Interface and Deep Learning, 08 2022.

[47] Spurr, Adrian and Dahiya, Aneesh and Wang, Xi and Zhang, Xucong and Hilliges, Otmar. Self-Supervised 3D Hand Pose Estimation from monocular RGB via Contrastive Learning. In *International Conference on Computer Vision*, 2021.

[48] Steffen Puhlmann and Jason Harris and Oliver Brock. RBO Hand 3: A Platform for Soft Dexterous Manipulation. *IEEE Transactions on Robotics*, 38, 2022.

[49] Vittorio Caggiano and Sudeep Dasari and Vikash Kumar. MyoDex: A Generalizable Prior for Dexterous Manipulation. *ArXiv*, abs/2309.03130, 2023.

[50] Wan, Chengde and Probst, Thomas and Van Gool, Luc and Yao, Angela. Self-Supervised 3D Hand Pose Estimation Through Training by Fitting. In *Computer Vision and Pattern Recognition*, 2019.

[51] Wang, Huawei and Caggiano, Vittorio and Durandau, Guillaume and Sartori, Massimo and Kumar, Vikash. MyoSim: Fast and physiologically realistic MuJoCo models for musculoskeletal and exoskeletal studies. In *International Conference on Robotics and Automation*, 2022.

[52] Wonik Robotics. Allegro Hand. http://wiki.wonikrobotics.com/AllegroHandWiki/index.php/Allegro_Hand_v4.0, 2023.

[53] Zhendong Wang and Jonathan J. Hunt and Mingyuan Zhou. Diffusion Policies as an Expressive Policy Class for Offline Reinforcement Learning. *ArXiv*, abs/2208.06193, 2022.

[54] Zicong Fan and Omid Taheri and Dimitrios Tzionas and Muhammed Kocabas and others. ARCTIC: A Dataset for Dexterous Bimanual Hand-Object Manipulation. *Computer Vision and Pattern Recognition*, 2022.

[55] Zilin Si and Kevin Zhang and Oliver Kroemer and Fatma Zeynep Temel. DeltaHands: A Synergistic Dexterous Hand Framework Based on Delta Robots. *IEEE Robotics and Automation Letters*, 9, 2023.