HL-IK: A Lightweight Implementation of Human-Like Inverse Kinematics in Humanoid Arms

Anonymous ICRA Submission

Abstract— **Traditional IK methods for redundant humanoid manipulators emphasize end-effector (EE) tracking, frequently producing configurations that are valid mechanically but not human-like. We present Human-Like Inverse Kinematics (HL-IK), a lightweight IK framework that preserves EE tracking while shaping whole-arm configurations to appear human-like—without full-body sensing at runtime. The key idea is a learned elbow prior: using large-scale human motion data retargeted to the robot, we train a FiLM-modulated spatio-temporal attention network (FiSTA) to predict the next-step elbow pose from the EE target and a short history of EE–elbow states. This prediction is incorporated as a small residual alongside EE and smoothness terms in a standard Levenberg–Marquardt optimizer, making HL-IK a drop-in addition to numerical IK stacks. Over 183k simulation steps, HL-IK reduces arm-similarity position and direction error by 30.6% and 35.4% on average, and by 42.2% and 47.4% on the most challenging trajectories. Hardware teleoperation on a robot distinct from simulation further confirms the gains in anthropomorphism. HL-IK is simple to integrate, adaptable across platforms via our pipeline, and adds minimal computation, enabling human-like motions for humanoid robots.**
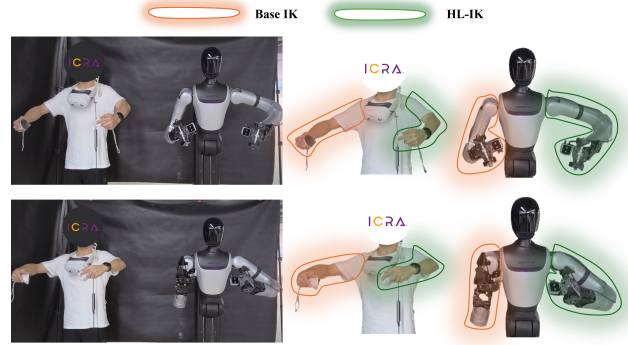

Fig. 1. In physical teleoperation, when only the EE pose is used as input, our HL-IK method achieves a more human-like effect, with the arm configuration more closely resembling that of humans.

I. INTRODUCTION

A robotic arm can be defined as a series of links connected together by joints [1]. Inverse kinematics (IK) is a fundamental problem in such robotics, traditionally formulated to compute joint configurations that achieve a specified end-effector (EE) pose. For industrial manipulators, this formulation is often sufficient, since the primary objective is to place the tool center point at the desired location with high precision. Classical IK solvers—whether based on closed-form derivations [2]–[5], numerical iterations [6], [7], or optimization frameworks [8]–[10] — focus almost exclusively on EE tracking.

For redundant robotic arms, the inverse solution to a given EE pose is often not unique, with infinitely many possible configurations [11], [12]. When only the EE pose is constrained, the intermediate joints remain underdetermined [13], which can lead to solutions that are mechanically valid but visually unnatural and non-human-like. In scenarios such as humanoid robot teleoperation [14]–[16], beyond accurate EE tracking, we also aim for the robot's overall arm configuration to closely resemble that of the human arm, thereby achieving a higher level of anthropomorphism. Existing methods [17], [18] often rely on external cameras to capture human body keypoints and align them with robot joints to improve configuration similarity. Yet, such approaches not only require additional perception inputs but also typically do not treat EE tracking as the primary constraint, and thus cannot be regarded as strict IK solutions. Therefore, our goal

is to develop a system that, given only the desired EE pose as input (as in traditional IK), not only ensures precise EE tracking but also achieves close similarity between the human and robot arm configurations.

To realize this goal, we first model the human arm as a four-point, three-segment kinematic chain comprising the shoulder, elbow, wrist, and fingertips (the human EE) [19]. For a fixed EE pose, the dominant redundancy manifests as the elbow "swivel" about the shoulder–wrist axis. Aligning the elbow pose effectively sets the arm plane and the forearm pointing direction, thereby resolving the main ambiguity and yielding anthropomorphic configurations without sacrificing EE accuracy. In other words, once the elbow is aligned, the overall arm configuration becomes perceptually natural and significantly more similar to that of a human. Furthermore, given a desired EE pose, determining a prior elbow pose that best reflects the natural human form becomes a central aspect of our approach. In summary, the primary contributions of this paper are:

1) Human-like data acquisition framework: We propose an automatic EE–elbow data collection scheme based on large-scale human motion trajectory datasets, which can be readily adapted to different robots.

2) Elbow prediction network: We design a FiLM-modulated Spatio-Temporal Attention Network (FiSTA) that uses only a partial history of EE and elbow frames to predict the desired human-like elbow pose for a given EE target.

3) Comprehensive validation: The effectiveness of our approach is validated both in simulation and on real hardware, as well as across different robotic configurations.
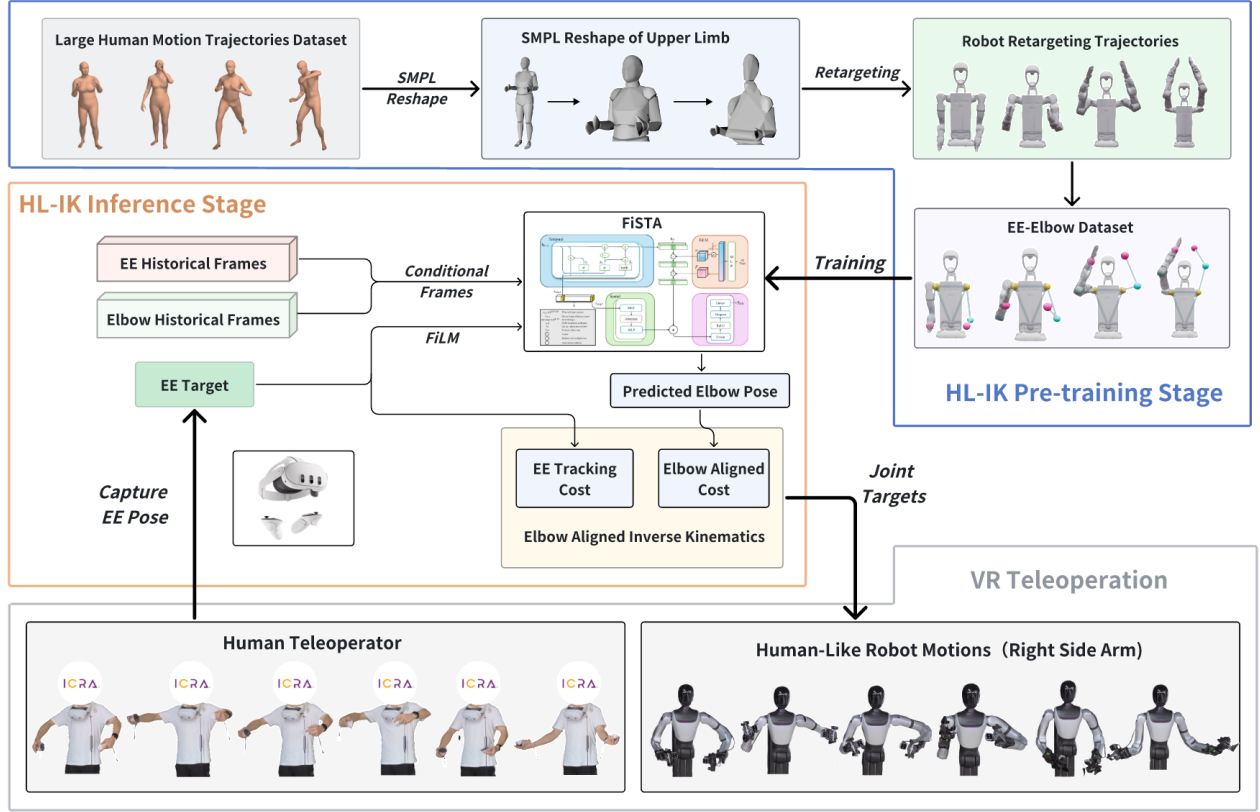
Fig. 2. Our method begins with a large-scale human motion dataset, which we retarget to the robot to extract an EE–elbow mapping dataset used to train our FiSTA network. After training, during human teleoperation we obtain the operator's desired current EE pose via a VR headset. Given this target and a fixed-length history of past frames, FiSTA predicts the elbow position. We then augment the IK objective with an elbow-alignment cost and solve for the desired joint angles using Levenberg–Marquardt iterations. The resulting commands are sent to the robot's low-level controller, yielding human-like motion.

II. RELATED WORK

A. Inverse Kinematics

The goal of IK is to compute joint configurations that realize a specified EE pose. Classical approaches pursue analytical (closed-form) solutions [2]–[4], which are highly efficient but hinge on restrictive assumptions about the manipulator's kinematic structure. More recently, the field has shifted toward iterative optimization formulations [8]–[10], [19] that trade closed-form speed for modeling flexibility and can accommodate secondary objectives arising from kinematic redundancy, including geometric constraints [10], [20] and learned collision penalties [21]. Building on this line of work, our proposed HL-IK explicitly incorporates humanoid-likeness into the IK objective. While ensuring precise EE tracking, HL-IK further guide the arm configuration to resemble natural human arm postures, thereby achieving more anthropomorphic motion behavior.

B. Motion Retargeting

Motion retargeting refers to mapping motions from a source entity to a target entity [22], such as transferring human movements to robots. It plays a crucial role in robotic teleoperation systems [23] and in learning frameworks based on human demonstrations [17], [24], [25]. Existing real-time whole-body teleoperation methods based on motion retargeting rely on sensing devices to capture the full-body state, which in turn requires additional perception algorithms. In contrast, HL-IK pretrains on large-scale retargeting data to learn an elbow-configuration prior, enabling human-like upper-limb motion from EE-only input.

C. Human-Like Motion Planning in IK

To encourage human-like arm configurations in IK, prior work largely takes two paths. Classical methods impose kinematic heuristics—e.g., swivel-angle parameterizations [26], [27] and physiology-inspired criteria [28]—but such hand-tuned rules generalize poorly across subjects and tasks. Recent data-driven method [29] learn wrist-pose→arm-angle mappings and then hard-wire them into closed-form IK; because this prior are entangled with analytic solvers, they transfer poorly to iterative numerical methods and restrict design flexibility. We instead learn an optimizer-agnostic elbow prior from large-scale human trajectories and encode it as a lightweight cost. Decoupled from any specific solver, this prior drops in as a generic objective term, integrating seamlessly with starndad numerical optimizers.
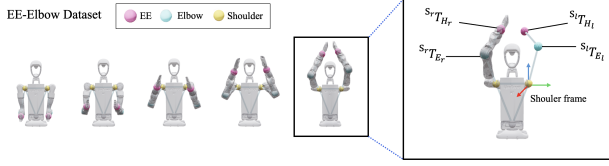
Fig. 3. EE-Elbow Data Collection. For each frame, extract the relative pose of the EE and the elbow in the corresponding shoulder frames.

III. METHODOLOGY

To realize HL-IK, we first retarget large-scale human motion datasets onto the robot to collect and process paired EE-elbow trajectories. We then introduce our FiLM-modulated Spatio-Temporal Attention (FiSTA) network, which learns from these data to predict the elbow pose given a short motion history and the current desired EE target. Finally, we show how the predicted elbow is incorporated into a numerical optimization IK. The full pipeline of our approach is illustrated in Fig. 2.

A. EE-Elbow Data Collection

We use AMASS [30] as the source of human motion trajectories, which provides consistent SMPL sequences with smooth kinematics and reliable pose quality. We retain six upper-limb joints—left/right shoulder, elbow, and EE—to represent the two arms. Following [17], we first estimate an upper-body shape parameter β' for the target robot via gradient-based fitting. Using β', we retarget each AMASS frame to the robot and build a frame-aligned dataset $\mathcal{D}_{\beta'}$ paired with the original AMASS sequence $\mathcal{D}_{\beta}$.

We adopt the convention that ${}^A\mathbf{T}_B \in SE(3)$ maps coordinates from frame $\{B\}$ to frame $\{A\}$. All poses in $\mathcal{D}_{\beta'}$ are expressed in the world frame $\{W\}$. From the retargeted sequence, forward kinematics (using the robot URDF) yields the world transforms of the shoulder, elbow, and hand (EE) for each arm: ${}^W\mathbf{T}_{S_j}, {}^W\mathbf{T}_{E_j}, {}^W\mathbf{T}_{H_j}$ with $j \in \{\text{left}, \text{right}\}$. The quantities we extract are the elbow and EE poses in the corresponding shoulder frames:

$$ {}^{S_j}\mathbf{T}_{E_j} = \left({}^W\mathbf{T}_{S_j}\right)^{-1} {}^W\mathbf{T}_{E_j}, \quad (1a) $$

$$ {}^{S_j}\mathbf{T}_{H_j} = \left({}^W\mathbf{T}_{S_j}\right)^{-1} {}^W\mathbf{T}_{H_j}. \quad (1b) $$

These four $SE(3)$ poses per frame—${}^{S_1}\mathbf{T}_{E_1}, {}^{S_1}\mathbf{T}_{H_1}, {}^{S_r}\mathbf{T}_{E_r}, {}^{S_r}\mathbf{T}_{H_r}$—constitute the supervision we use for learning and downstream IK. Fig. 3 illustrates this data collection process.

B. FiLM-modulated Spatio-Temporal Attention Network for Elbow Prediction

Inputs and outputs. For a single arm, the history window contains T frames. Each frame encodes the 7D pose (3D position + 4D unit quaternion) of the EE and the elbow in the shoulder-local frame, giving $7 \times 2 \times T = 14T$ dimensions in total. The future conditioning input is the next-step EE target pose (7D). The network predicts a 7D elbow pose for the next step.
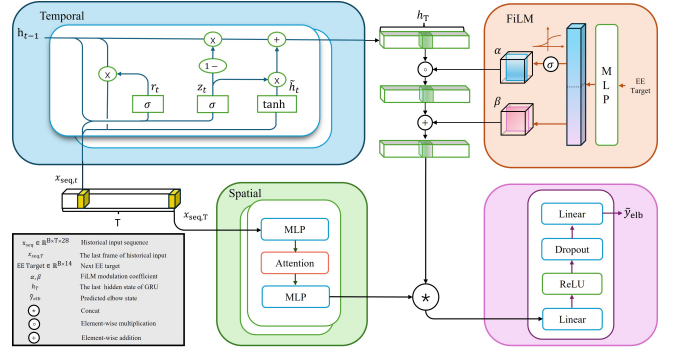

Fig. 4. Model Architecture. A GRU encodes the full history to produce temporal features, which are FiLM-conditioned using the next EE target. In parallel, an attention module computes spatial features from the last frame. The two streams are then concatenated to predict the elbow configuration.

FiSTA architecture. Fig. 4 illustrates the network architecture of FiSTA and the model comprises four cooperating modules:

1) Temporal Encoder: A GRU consumes the 5-frame history and produces a temporal summary that captures position, velocity changes, and inertial trends.

2) Spatial Attention Module: On the most recent frame, the EE pose and elbow pose are treated as two tokens; a lightweight self-attention layer models their instantaneous dependence to yield a spatial summary.

3) Goal Modulator: The next-step EE target is mapped to FiLM-style feature-wise scale and shift, which affine-modulate the temporal summary and thus condition historical dynamics on the imminent goal.

4) Fusion and Output Head: The modulated temporal features are concatenated with the spatial summary and passed through an MLP regressor to produce the 7D elbow prediction.

Design rationale. This design explicitly decouples temporal dynamics from instantaneous spatial coordination while enabling goal-conditioned modulation. Compared with naive target concatenation, FiLM modulation provides more expressive conditioning, and the lightweight spatial attention captures crucial EE–elbow coordination with negligible computational overhead.

C. Elbow Aligned Inverse Kinematics

Kinematic optimization benefits from quasi-Newton approaches, which accelerate convergence by exploiting curvature information of the cost function [31]. In our IK computation, we employ the Levenberg–Marquardt (LM) optimizer [32]–[36], which minimizes the sum of squared residuals defined below.

1) EE Pose Cost: This EE pose cost c_{ee} allows the IK to track the given EE target as closely as possible.

$$ c_{\text{ee}}(\boldsymbol{q}) = \mathbf{W}_{\text{ee}}^{1/2} \log\left(\left(\mathbf{T}_{\text{base}}^{\text{tar}_{\text{ee}}}\right)^{-1} \mathbf{T}_{\text{base}}^{\text{ee}}(\boldsymbol{q})\right) \in \mathbb{R}^6, \quad (2) $$

2) Elbow Pose Cost: This elbow pose cost c_{elbow} enables the IK to track the given EE elbow to a certain extent.

$$ c_{\text{elbow}}(\boldsymbol{q}) = \mathbf{W}_{\text{elbow}}^{1/2} \log\left(\left(\mathbf{T}_{\text{base}}^{\text{tar}_{\text{el}}}\right)^{-1} \mathbf{T}_{\text{base}}^{\text{el}}(\boldsymbol{q})\right) \in \mathbb{R}^6 \quad (3) $$

3) Smoothness Cost: The smoothness cost c_{smooth} encourages small changes in joint positions, and is useful for generating smooth trajectories.

$$c_{\text{smooth}}(\boldsymbol{q}) = \mathbf{W}_{\text{smooth}}^{1/2}\left(\boldsymbol{q}_t - \boldsymbol{q}_{t-1}\right) \in \mathbb{R}^d \qquad (4)$$

where

- $\mathbf{T}_{\text{base}}^{\text{tar}_{\text{ee}}} \in SE(3)$ is the target EE pose expressed in the base frame,
- $\mathbf{T}_{\text{base}}^{\text{ee}}(\boldsymbol{q}) \in SE(3)$ is the forward-kinematics EE pose at configuration $\boldsymbol{q}$,
- $\mathbf{T}_{\text{base}}^{\text{tar}_{\text{el}}} \in SE(3)$ is the target elbow pose in the base frame,
- $\mathbf{T}_{\text{base}}^{\text{el}}(\boldsymbol{q}) \in SE(3)$ is the elbow pose from forward kinematics,
- $\log(\cdot) : SE(3) \to \mathbb{R}^6$ is the matrix logarithm mapping rigid-body transformations to 6D twist coordinates (translation + rotation error),
- $\mathbf{W}_{(\cdot)}^{1/2}$ denotes the square-root of the weight matrix used for residual scaling,
- $\boldsymbol{q}_t \in \mathbb{R}^d$ is the robot joint configuration at time step t and the total number of joints is d.

Then, the three types of residuals are stacked into a single residual vector $\boldsymbol{c}(\boldsymbol{q})$, and its Jacobian $\mathbf{J}(\boldsymbol{q})$ with respect to the joint configuration $\boldsymbol{q}$ is computed. The Levenberg–Marquardt iteration then updates the joint configuration by solving a damped least-squares problem, which balances fast convergence with robustness against ill-conditioning:

$$\boldsymbol{c}(\boldsymbol{q}) = \begin{bmatrix} \boldsymbol{c}_{\text{ee}}(\boldsymbol{q}) \\ \boldsymbol{c}_{\text{elbow}}(\boldsymbol{q}) \\ \boldsymbol{c}_{\text{smooth}}(\boldsymbol{q}) \end{bmatrix}, \qquad (5a)$$

$$\mathbf{J}(\boldsymbol{q}) = \frac{\partial \boldsymbol{c}(\boldsymbol{q})}{\partial \boldsymbol{q}}, \qquad (5b)$$

$$\boldsymbol{q}_n = \boldsymbol{q}_{n-1} - \left(\mathbf{J}^\top \mathbf{J} + \lambda \mathbf{I}\right)^{-1} \mathbf{J}^\top \boldsymbol{c}(\boldsymbol{q}_{n-1}). \qquad (5c)$$

Through successive iterations, the joint configuration $\boldsymbol{q}_n$ converges to a solution $\boldsymbol{q}^*$ that minimizes the stacked costs, yielding an IK solution with accurate EE tracking, improved elbow alignment, and smooth joint motions.

IV. EXPERIMENTS

A. Model Evaluation

To comprehensively assess the elbow–prediction network, we conduct systematic experiments along three axes: (i) Network Comparisons; (ii) Tests on history length and ablation experiments on the FiSTA network; (iii) Runtime tests. To ensure broad coverage—from everyday activities to dynamic, in the AMASS dataset, we specifically employ the following three items:

a. ACCAD Dataset, 252 trajectories [37]
b. CMU Dataset, 2079 trajectories [38]
c. SFU Dataset, 44 trajectories [39]

To balance efficiency and reliability, we adopt a two-stage training strategy: (1) All network comparison experiments, ablation studies, and hyperparameter searches are conducted on the ACCAD Dataset. (2) We then perform final training
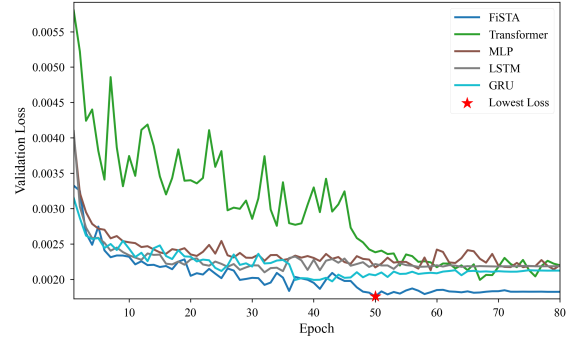

Fig. 5. Model Evaluation Comparison. The y-axis shows validation MSE over epochs. The red star marks the lowest loss of FiSTA. All models use a 5-frame history, with other hyperparameters set to their best configuration.
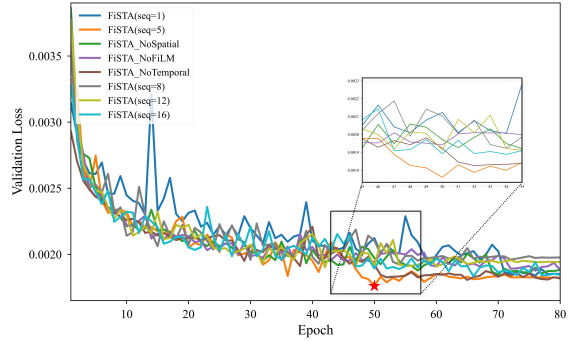

Fig. 6. Ablation and sequence length comparison of FiSTA. Validation loss with different history lengths (1, 5, 8, 12, 16 frames) and ablation variants (without spatial, FiLM, or temporal).

on the full dataset using the selected configuration. During comparison, we split the dataset by trajectory, with 90% for training and 10% for validation.

1) Network Comparisons: We evaluated five representative architectures—MLP, LSTM, GRU, Transformer, and FiSTA—using the ACCAD subset, where each was trained with tuned hyperparameters under identical normalization and loss settings. The history length was fixed at 5 for all models. Validation errors are reported in Fig. 5. Among the candidates, FiSTA achieved the lowest validation error (`MSE=0.001760`), substantially outperforming all baselines. By contrast, the explicit architectural separation of spatial and temporal processing in FiSTA provides a stronger inductive bias than standard recurrent models or Transformers, which must learn these relationships from a less structured hidden state.

2) Tests on history length and ablation experiments on the FiSTA network: To rigorously evaluate our proposed model, we conducted a series of tests to analyze the impact of history sequence length and core architectural components on the final validation loss. First, we assessed the model's performance with varying lengths of historical context, testing sequence lengths of 1, 5, 8, 12, and 16 frames. As shown in Fig. 6, the results clearly indicate an optimal contextual window. The model with a sequence length of 5 achieved the lowest validation loss, demonstrating maximum predictive
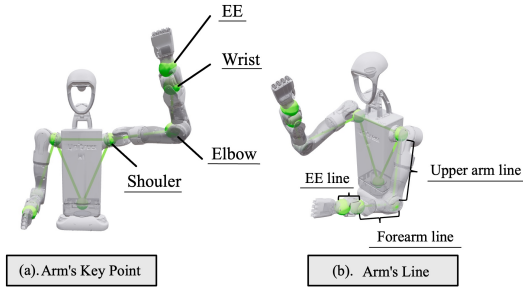
Fig. 7. Keypoints and segments for metric computation. Display the four key points (Shoulder, Elbow, Wrist, EE) and three line segments (Upperarm, Forearm, EE line) used for quantitative indicators of anthropomorphism.

precision. We observed that $L = 5$ strikes a favorable balance: too short a context fails to provide sufficient dynamic information for inferring motion trends, while an excessively long context introduces irrelevant historical data, reducing the signal-to-noise ratio. To quantify the contribution of each component, we performed an ablation study on `FiSTA (seq=5)` (all deltas relative to the full model). Removing the Spatial Attention and Temporal modules increased the loss by 4.9% and 2.8%, respectively, indicating that both are important for building a robust spatiotemporal representation. The largest effect came from ablating the Goal Modulation module, which led to a 6.6% drop in accuracy. This module, implemented as a FiLM layer, conditions the GRU-encoded temporal features on the EE target. Its removal therefore highlights the critical role of FiLM-based goal conditioning in injecting target information into the temporal feature stream for this task.

3) Runtime Comparison: To assess the deployment potential of our model in practical applications, we conducted a comprehensive runtime tests on an NVIDIA RTX 4070 GPU. The total runtime for each model was calculated by averaging the runtime for one arm across the ACCAD trajectories. In this comparison, the total time is divided into two main stages as shown in Table I: (i) Preprocess, which includes the neural network's input/output handling and the model inference itself. (ii) IK, which represents the time taken by the subsequent IK solver. Our optimal model, `FiSTA (seq=5)`, achieved the best performance, though its overall running time ranked fourth among all the models. The final computation took 7.0840 (ms), which is well within the acceptable range.

4) Conclusion: We set the default history length for FiSTA to $L = 5$. The network was then trained on the entire dataset for a total of 1 hour and 50 minutes on an NVIDIA RTX 4090 GPU.

B. Comparative Analysis of Elbow Alignment

When incorporating elbow alignment into IK, we set the weight coefficients as follows.

$$\mathbf{W}_{\text{ee}} = \text{diag}(50\,\mathbf{I}_3,\, 40\,\mathbf{I}_3), \tag{6a}$$

$$\mathbf{W}_{\text{elbow}} = \text{diag}(20\,\mathbf{I}_3,\, 5\,\mathbf{I}_3), \tag{6b}$$

$$\mathbf{W}_{\text{smooth}} = 0.35\,\mathbf{I}_d. \tag{6c}$$

In $\mathbf{W}_{\text{ee}}$ and $\mathbf{W}_{\text{elbow}}$, the first 3×3 block weights translational residuals and the second weights rotational residuals; $\mathbf{I}_d$ denotes the $d\times d$ identity where d is the joints number.

For quantitative evaluation, we use a new human-motion dataset (not for network training) - BMLhandball (649 trajectories, 183,806 steps) [40]. The trajectory generated by the original retargeting method serves as the reference, and we compare two IK solvers: (i) a baseline using only the EE pose as input (ii) our elbow-aligned method HL-IK. We conduct a comprehensive evaluation using four metrics: two that quantify arm-configuration similarity and two that EE tracking precision.

1) Arm's Key Point Position Error: Measures the absolute positional deviation of key arm points $K_{\text{arm}} = \{\text{elbow, wrist, EE}\}$ from the reference trajectory. This is mainly for calculating the error of key points on the arm configuration after IK solving. Fig. 7 (a) demonstrated these arm's key points.

$$E_{\text{arm}}^p(t) = \sum_{k \in K_{\text{arm}}} \left\| \boldsymbol{p}_k^{\text{re}}(t) - \boldsymbol{p}_k^{\text{ik}}(t) \right\|^2, \tag{7}$$

where $\boldsymbol{p}_k^{\text{re}}(t)$ and $\boldsymbol{p}_k^{\text{ik}}(t)$ represent the 3D positions of the key points from the reference trajectory and the IK solution at time t, respectively.

2) Arm' Line Angle Error: Quantifies the angular deviation of $L_{\text{arm}} = \{$upper arm line (shoulder-elbow), forearm line (elbow-wrist), EE line (wrist-EE) $\}$. This is mainly for calculating the error of the direction of line segments on the arm configuration after IK solving. Fig. 7 (b) demonstrated these arm's line.

$$E_{\text{arm}}^\theta(t) = \sum_{l \in L_{\text{arm}}} \arccos\left(\frac{\boldsymbol{v}_l^{\text{re}}(t) \cdot \boldsymbol{v}_l^{\text{ik}}(t)}{\left\| \boldsymbol{v}_l^{\text{re}}(t) \right\| \left\| \boldsymbol{v}_l^{\text{ik}}(t) \right\| + \alpha} \right), \tag{8}$$

where $\boldsymbol{v}_l^{\text{re}}(t)$ and $\boldsymbol{v}_l^{\text{ik}}(t)$ denote the arm-segment vectors of the reference trajectory and the IK solution at time t, respectively. $\alpha > 0$ is a small numerical constant added for stability to avoid division by zero.

3) EE Position Tracking Error: Evaluates the 3D positional tracking accuracy of the EE relative to the reference trajectory.

$$E_{\text{ee}}^p(t) = \left\| \boldsymbol{p}_{\text{ee}}^{\text{re}}(t) - \boldsymbol{p}_{\text{ee}}^{\text{ik}}(t) \right\|^2, \tag{9}$$

where $\boldsymbol{p}_{\text{ee}}^{\text{re}}(t)$ and $\boldsymbol{p}_{\text{ee}}^{\text{ik}}(t)$ represent EE position of the reference trajectory and the IK solution at time t, respectively.

TABLE I
THE AVERAGE RUNTIME AND THE MINIMUM LOSS COMPARISON OF DIFFERENT MODELS

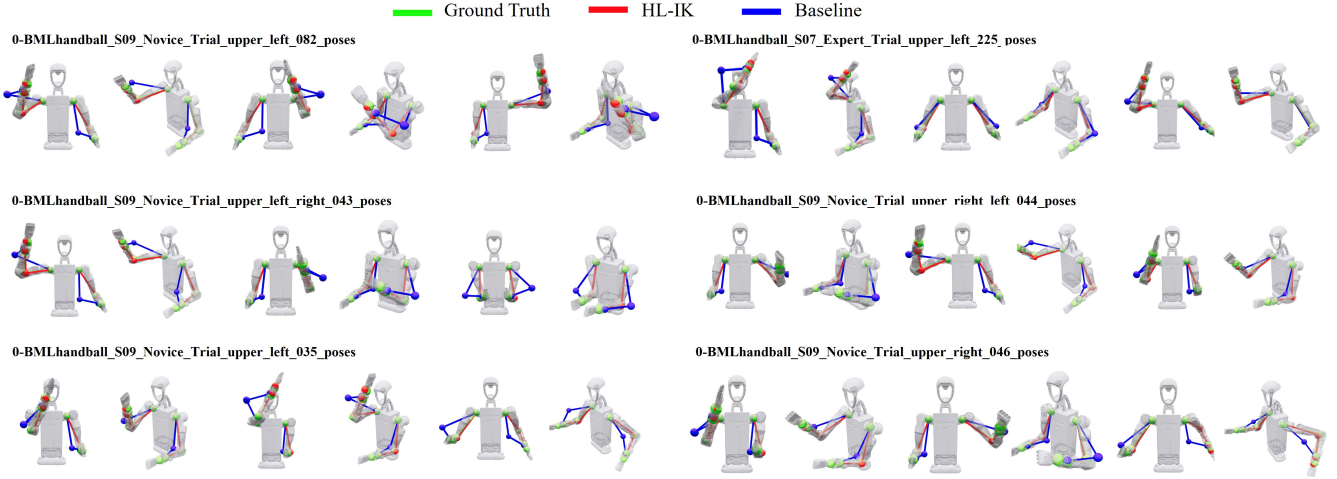| Model Name	Preprocess (ms)	IK (ms)	Total (ms)	Loss ($\times 10^{-3}$)
MLP	**1.3104 $\pm$ 0.153**	5.4896 $\pm$ 0.951	6.8000 $\pm$ 0.963	2.130
GRU	1.7565 $\pm$ 0.210	6.4981 $\pm$ 1.102	8.2547 $\pm$ 1.122	1.973
LSTM	1.5480 $\pm$ 0.185	4.3283 $\pm$ 0.824	**5.8763 $\pm$ 0.842**	2.197
Transformer	7.3642 $\pm$ 0.551	4.6888 $\pm$ 0.915	12.0530 $\pm$ 1.068	1.994
FiSTA (seq=1)	2.8091 $\pm$ 0.303	4.8272 $\pm$ 1.040	7.6363 $\pm$ 1.173	1.825
FiSTA (seq=5)	2.8665 $\pm$ 0.325	4.2175 $\pm$ 0.562	7.0840 $\pm$ 0.772	**1.760**
FiSTA_NoSpatial	2.0050 $\pm$ 0.126	**3.9653 $\pm$ 0.679**	5.9703 $\pm$ 0.698	1.846
FiSTA_NoTemporal	1.9993 $\pm$ 0.107	5.7616 $\pm$ 0.626	7.7609 $\pm$ 0.637	1.809
FiSTA_NoFiLM	2.1463 $\pm$ 0.072	9.0212 $\pm$ 3.291	11.1674 $\pm$ 3.303	1.876
FiSTA (seq=8)	3.0448 $\pm$ 0.339	4.2995 $\pm$ 0.745	7.3443 $\pm$ 0.954	1.923
FiSTA (seq=12)	3.1464 $\pm$ 0.380	5.4129 $\pm$ 1.188	8.5593 $\pm$ 1.405	1.902
FiSTA (seq=16)	3.1618 $\pm$ 0.368	4.2071 $\pm$ 0.658	7.3689 $\pm$ 0.878	1.926
Baseline IK (No NN)	/	5.0818 $\pm$ 0.755	5.0818 $\pm$ 0.755	/

Fig. 8. The robot motions. Comparison of robot motions reconstructed by different methods: Ground Truth (green), HL-IK (red), and Baseline (blue). In each visualization, four spheres on each arm represent the shoulder, elbow, wrist, and EE, respectively. The gray robot mesh is rendered from the ground-truth poses, while the colored skeletons illustrate the predicted trajectories of different methods.

4) EE Orientation Tracking Error: Evaluates the 3D orientational tracking accuracy of the EE relative to the reference trajectory.

$$E_{ee}^o(t) = \phi\left(\mathbf{R}_{ee}^{re}(t) \cdot \left(\mathbf{R}_{ee}^{ik}(t)\right)^{-1}\right)^2, \quad (10a)$$

$$\phi = \text{acos}\left(\frac{\text{tr}(\mathbf{R}) - 1}{2}\right) \quad (10b)$$

where $\mathbf{R}_{ee}^{re}(t)$ and $\mathbf{R}_{ee}^{ik}(t)$ represent the rotation matrices of the EE from the reference trajectory and the IK solution at time t, respectively. ϕ quantifies the rotational difference as a specific scalar value representing the angle and $\text{tr}(\cdot)$ calculates the trace of the matrix.

Based on the four metrics mentioned above, we conducted a comparative analysis on the BMLhandball's [40] human retargeting dataset. First, we calculated the average error across all steps in the trajectory. The first two columns in Table. II reports the mean errors across all 183,806 steps. Each indicator in the table shows the average error for all steps. With elbow alignment, the arm's key point position error and the line angle error are reduced of 30.6% and 35.4%, while the EE tracking accuracy drops slightly—an

acceptable trade-off for teleoperation.

Since our aim is to probe configuration-tracking advantage in our HL-IK, we rank-ordered trajectories by the "Arm's Key Point Position Error" achieved by the baseline method and selected the top 20% as "challenging trajectories" for further evaluation. We then compared the baseline with HL-IK on this subset. As reported in the last two columns in Table. II, and relative to the first two columns, the baseline shows large increases in both arm's key point position error and line angel error, whereas HL-IK maintains nearly the same error levels as on the full set. On this subset, HL-IK reduces arm keypoint and direction errors by 42.2% and 47.4%, respectively. Fig. 8 visualizes this gap: the baseline (blue) deviates substantially, while HL-IK (red) remains closely aligned with the reference configuration.

C. Teleoperation Experiments on a Physical Robot

We tested a robot model in the physical world that is completely different from the one used in simulation—a wheeled-arm robot's upper limb. We use the same weight coefficients as those used in the simulation. We applied the HL-IK method to the robot's left arm and the original IK

TABLE II

COMPARISON OF MEAN METRICS ON THE FULL DATASET AND CHALLENGING DATESET

Quantitative Indicators	HL-IK	Baseline	HL-IK (Challenging)	Baseline (Challenging)
Key Point Position Error (m)	$5.944 \times 10^{-2} \pm 0.00412$	$8.571 \times 10^{-2} \pm 0.00942$	$5.977 \times 10^{-2} \pm 0.00436$	$10.338 \times 10^{-2} \pm 0.00812$
Line Angle Error (rad)	$3.565 \times 10^{-1} \pm 0.02438$	$5.515 \times 10^{-1} \pm 0.08580$	$3.575 \times 10^{-1} \pm 0.02190$	$6.791 \times 10^{-1} \pm 0.06978$
EE Position Error (m)	$4.575 \times 10^{-3} \pm 0.00067$	$1.847 \times 10^{-3} \pm 0.00028$	$5.405 \times 10^{-3} \pm 0.00218$	$3.642 \times 10^{-3} \pm 0.00483$
EE Orientation Error (rad)	$8.110 \times 10^{-4} \pm 0.00021$	$4.621 \times 10^{-4} \pm 0.00008$	$10.728 \times 10^{-4} \pm 0.00331$	$9.495 \times 10^{-4} \pm 0.00141$

(a) Low Elbow Height Situations

(b) Mid Elbow Height Situations

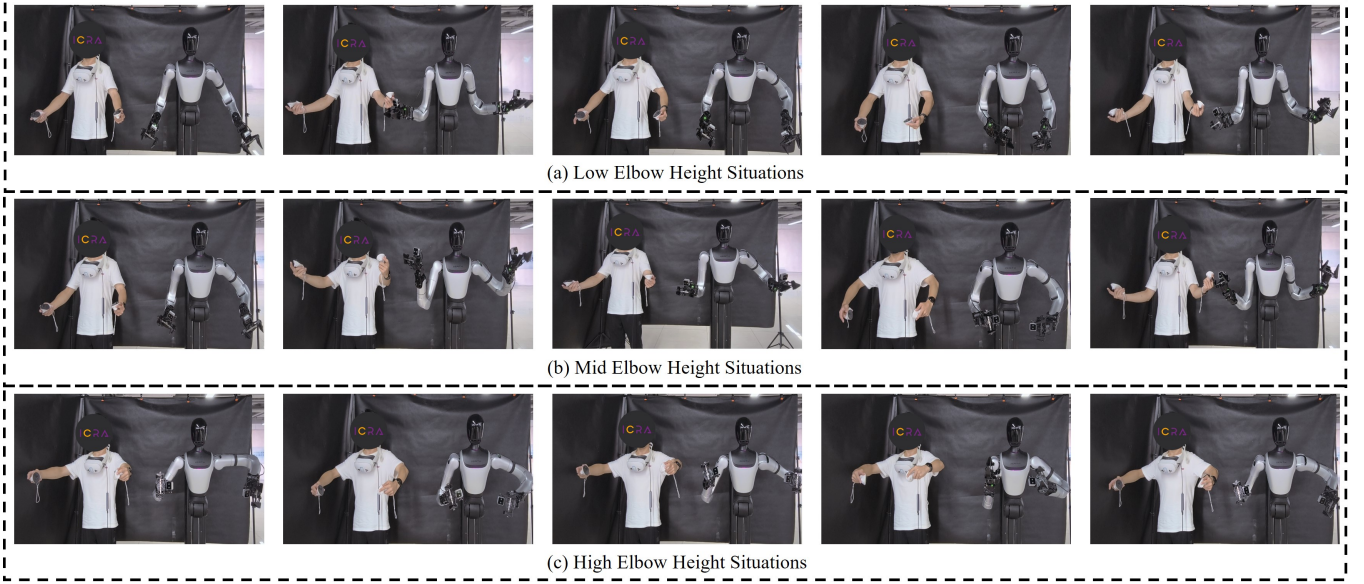(c) High Elbow Height Situations

Fig. 9. Teleoperation on a physical robot. We deployed the HL-IK framework on the robot's left arm (shown on the right side of each figure) and a standard IK solver without elbow-joint optimization on the right arm (shown on the left side of each figure). With HL-IK, the robot not only tracked the end effector but, without any direct sensing, also inferred and aligned with the human elbow position. As a result, its overall arm configuration closely matched the human's, whereas the baseline IK achieved only end effector tracking and failed to reproduce configuration level similarity. Specifically, the higher the elbow joint height, the more apparent the difference.

method to the right arm, observing the differences in arm configurations during the teleoperation task. We use a Meta VR headset as the sensing input for the EE target, and no other sensing is employed. As shown in Fig. 9, HL-IK enabled the robot to track the end effector and, without direct sensing, infer and align with the human elbow position. As a result, the robot's overall arm configuration closely matched the human's, whereas the baseline IK achieved only end effector tracking and failed to reproduce configuration level similarity. In particular, this difference becomes increasingly apparent with higher elbow joint heights.

V. CONCLUSIONS

In this paper, we introduced HL-IK, a lightweight IK framework that preserves EE tracking while producing human-like arm configurations without full-body sensing. The key idea is a learned elbow prior: a FiSTA network predicts the next-step elbow pose from a short motion history and the desired hand target, and this prediction is injected as a small residual cost alongside EE tracking and smoothness in a standard Levenberg–Marquardt IK stack. This design integrates seamlessly with generic kinematics libraries, adds negligible overhead (best trade-off at a 5-frame history), and significantly improves anthropomorphism. On a 183k-step

test, HL-IK reduces arm key-point and directional errors by 30.6% and 35.4% on average across all steps—and by 42.2% and 47.4% on the most challenging trajectories—while maintaining acceptable EE accuracy. Teleoperation experiments on a physical robot likewise clearly demonstrate the effectiveness of our method in improving arm-configuration similarity in IK. Looking ahead, we will explore richer upper-limb priors, extend to whole-body control, enable online adaptation and personalization, and unify collision and task constraints within a single optimization framework to further improve generalization and robustness in real deployments.

REFERENCES

[1] R. P. Paul, *Robot Manipulators : Mathematics, Programming, and Control : The Computer Control of Robot Manipulators*. Cambridge, Mass. : MIT Press, 1981.

[2] X. Tian, Q. Xu, and Q. Zhan, "An analytical inverse kinematics solution with joint limits avoidance of 7-dof anthropomorphic manipulators without offset," *Journal of the Franklin Institute*, vol. 358, no. 2, pp. 1252–1272, Jan. 2021.

[3] L. Zheng, P. Lin, M. Liang, C. Wang, Y. Li, J. Sun, Y. Han, and H. Liu, "Analytical inverse kinematics for a prismatic-revolute hybrid joints radiography robot mounted on the ambulance," *IEEE/ASME Transactions on Mechatronics*, vol. 30, no. 2, pp. 1366–1376, Apr. 2025.

[4] M. N. Vu, F. Beck, M. Schwegel, C. Hartl-Nesic, A. Nguyen, and A. Kugi, "Machine learning-based framework for optimally solving

the analytical inverse kinematics for redundant manipulators," *Mechatronics*, vol. 91, p. 102970, May 2023.

[5] R. Diankov, "Automated construction of robotic manipulation programs," Ph.D. dissertation, Carnegie Mellon University, USA, 2010.

[6] R. M. Murray, Z. Li, and S. S. Sastry, *A Mathematical Introduction to Robotic Manipulation*, 1st ed. CRC Press, Dec. 2017.

[7] A. Yonezawa, H. Yonezawa, and I. Kajiwara, "Simple inverse kinematics computation considering joint motion efficiency," *IEEE Transactions on Cybernetics*, vol. 54, no. 9, pp. 4903–4914, Sep. 2024.

[8] D. O. Santos, L. Molina, J. G. N. Carvalho, E. A. N. Carvalho, and E. O. Freire, "Modifications of fully resampled pso in the inverse kinematics of robot manipulators," *IEEE Robotics and Automation Letters*, vol. 9, no. 2, pp. 1923–1928, Feb. 2024.

[9] F. Marić, M. Giamou, A. W. Hall, S. Khoubyarian, I. Petrović, and J. Kelly, "Riemannian optimization for distance-geometric inverse kinematics," *IEEE Transactions on Robotics*, vol. 38, no. 3, pp. 1703–1722, Jun. 2022.

[10] B. Sundaralingam, S. K. S. Hari, A. Fishman, C. Garrett, K. Van Wyk, V. Blukis, A. Millane, H. Oleynikova, A. Handa, F. Ramos, N. Ratliff, and D. Fox, "Curobo: Parallelized collision-free robot motion generation," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, May 2023, pp. 8112–8119.

[11] B. Ames, J. Morgan, and G. Konidaris, "Ikflow: Generating diverse inverse kinematics solutions," *IEEE Robotics and Automation Letters*, vol. 7, no. 3, pp. 7177–7184, Jul. 2022.

[12] A. C. A. Dione, "Stable posture tracking framework for modular hyper-redundant serial arms."

[13] M. Boukheddimi, R. Kumar, S. Kumar, J. Carpentier, and F. Kirchner, "Investigations into exploiting the full capabilities of a series-parallel hybrid humanoid using whole body trajectory optimization," in *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Oct. 2023, pp. 10 433–10 439.

[14] K. Darvish, L. Penco, J. Ramos, R. Cisneros, J. Pratt, E. Yoshida, S. Ivaldi, and D. Pucci, "Teleoperation of humanoid robots: A survey," *IEEE Transactions on Robotics*, vol. 39, no. 3, pp. 1706–1727, Jun. 2023.

[15] M. Seo, S. Han, K. Sim, S. H. Bang, C. Gonzalez, L. Sentis, and Y. Zhu, "Deep imitation learning for humanoid loco-manipulation through human teleoperation," in *2023 IEEE-RAS 22nd International Conference on Humanoid Robots (Humanoids)*, Dec. 2023, pp. 1–8.

[16] A. Handa, K. Van Wyk, W. Yang, J. Liang, Y.-W. Chao, Q. Wan, S. Birchfield, N. Ratliff, and D. Fox, "Dexpilot: Vision-based teleoperation of dexterous robotic hand-arm system," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, May 2020, pp. 9164–9170.

[17] T. He, Z. Luo, W. Xiao, C. Zhang, K. Kitani, C. Liu, and G. Shi, "Learning human-to-humanoid real-time whole-body teleoperation," in *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Oct. 2024, pp. 8944–8951.

[18] T. Zhou, Q. Zhu, Y. Ye, and J. Du, "Humanlike inverse kinematics for improved spatial awareness in construction robot teleoperation: Design and experiment," *Journal of Construction Engineering and Management*, vol. 149, no. 7, p. 04023044, Jul. 2023.

[19] Y. Wang, P. Praveena, D. Rakita, and M. Gleicher, "Rangedik: An optimization-based robot motion generation method for ranged-goal tasks," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, May 2023, pp. 9700–9706.

[20] D. Rakita, B. Mutlu, and M. Gleicher, "Relaxedik: Real-time synthesis of accurate and feasible robot arm motion," in *Robotics: Science and Systems XIV*. Robotics: Science and Systems Foundation, Jun. 2018.

[21] D. Rakita, H. Shi, B. Mutlu, and M. Gleicher, "Collisionik: A per-instant pose optimization method for generating robot motions with environment collision avoidance," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, May 2021, pp. 9995–10 001.

[22] M. Gleicher, "Retargetting motion to new characters," in *Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques*, ser. SIGGRAPH '98. New York, NY, USA: Association for Computing Machinery, Jul. 1998, pp. 33–42.

[23] T. Schmidt, R. Newcombe, and D. Fox, "Dart: Dense articulated real-time tracking," in *Robotics: Science and Systems X*. Robotics: Science and Systems Foundation, Jul. 2014.

[24] Z. Fu, Q. Zhao, Q. Wu, G. Wetzstein, and C. Finn, "Humanplus: Humanoid shadowing and imitation from humans," Jun. 2024.

[25] H. Zhang, W. Li, J. Liu, Z. Chen, Y. Cui, Y. Wang, and R. Xiong, "Kinematic motion retargeting via neural latent optimization for learning sign language," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 4582–4589, Apr. 2022.

[26] H. Kim, Zhi Li, D. Milutinovic, and J. Rosen, "Resolving the redundancy of a seven dof wearable robotic system based on kinematic and dynamic constraint," in *2012 IEEE International Conference on Robotics and Automation*. St Paul, MN, USA: IEEE, May 2012, pp. 305–310.

[27] W. Liu, D. Chen, and J. Steil, "Analytical inverse kinematics solver for anthropomorphic 7-dof redundant manipulators with human-like configuration constraints," *Journal of Intelligent & Robotic Systems*, vol. 86, no. 1, pp. 63–79, Apr. 2017.

[28] X. Zheng, Y. Han, and J. Liang, "Anthropomorphic motion planning for multi-degree-of-freedom arms," *Frontiers in Bioengineering and Biotechnology*, vol. 12, p. 1388609, May 2024.

[29] S. Li, K. Han, P. He, Z. Li, Y. Liu, and Y. Xiong, "Human-like redundancy resolution: An integrated inverse kinematics scheme for anthropomorphic manipulators with radial elbow offset," *Advanced Engineering Informatics*, vol. 54, p. 101812, Oct. 2022.

[30] N. Mahmood, N. Ghorbani, N. F. Troje, G. Pons-Moll, and M. J. Black, "AMASS: Archive of motion capture as surface shapes," in *International Conference on Computer Vision*, Oct. 2019, pp. 5442–5451.

[31] C. M. Kim, B. Yi, H. Choi, Y. Ma, K. Goldberg, and A. Kanazawa, "Pyroki: A modular toolkit for robot kinematic optimization," May 2025.

[32] K. Levenberg, "A method for the solution of certain non-linear problems in least squares," *Quarterly of Applied Mathematics*, vol. 2, no. 2, pp. 164–168, 1944.

[33] D. W. Marquardt, "An algorithm for least-squares estimation of non-linear parameters," *Journal of the Society for Industrial and Applied Mathematics*, vol. 11, no. 2, pp. 431–441, 1963.

[34] C. W. Wampler, "Manipulator inverse kinematic solutions based on vector formulations and damped least-squares methods," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 16, no. 1, pp. 93–101, Jan. 1986.

[35] Y. Nakamura and H. Hanafusa, "Inverse kinematic solutions with singularity robustness for robot manipulator control," *Journal of Dynamic Systems, Measurement, and Control*, vol. 108, no. 3, pp. 163–171, Sep. 1986.

[36] S. R. Buss and J.-S. Kim, "Selectively damped least squares for inverse kinematics," *Journal of Graphics Tools*, vol. 10, no. 3, pp. 37–49, Jan. 2005.

[37] Advanced Computing Center for the Arts and Design, "ACCAD MoCap Dataset." [Online]. Available: https://accad.osu.edu/research/motion-lab/mocap-system-and-data

[38] Carnegie Mellon University, "CMU MoCap Dataset." [Online]. Available: http://mocap.cs.cmu.edu

[39] S. F. University and N. U. of Singapore, "SFU Motion Capture Database." [Online]. Available: http://mocap.cs.sfu.ca/

[40] F. Helm, N. Troje, M. Reiser, and J. Munzert, "Bewegungsanalyse getäuschter und nicht-getäuschter 7m-würfe im handball," 01 2015.