

# Learning External Force Sensing for Commodity Robot Arms Improves Policy Learning

Author Names Omitted for Anonymous Review.

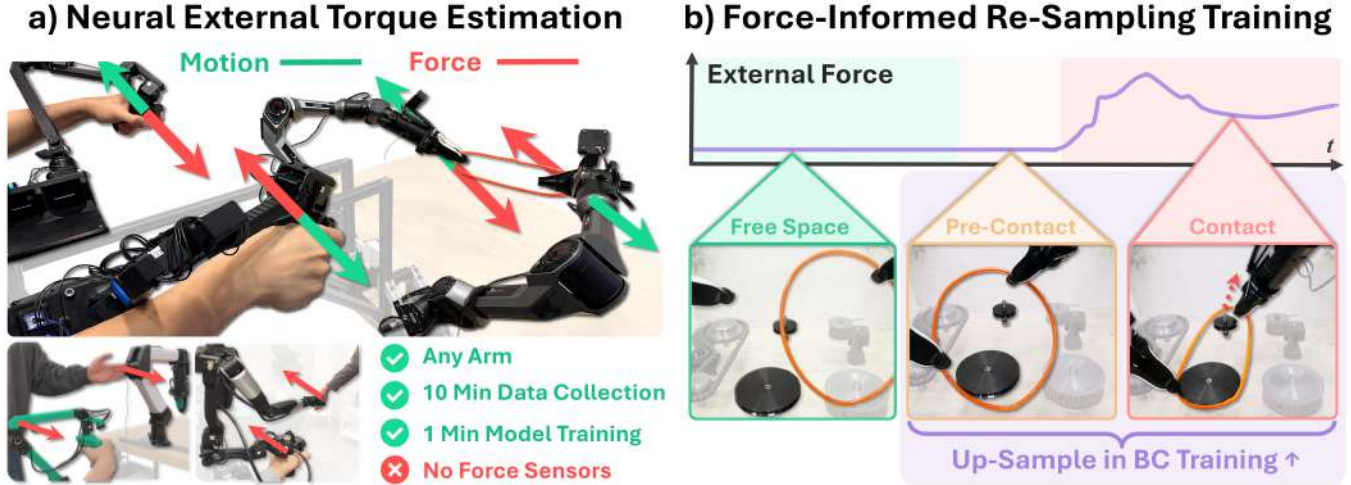


Fig. 1. **Overview of our approach.** (a) Neural External Torque Estimation (NEXT) produces high-quality joint torque estimates using only 10 minutes of data without dedicated force sensors or explicit system identification, enabling force-feedback teleoperation on low-cost arms such as the Piper, YAM, and Nero. (b) Force-Informed Re-Sampling Training (FIRST) uses learned external torque estimates to segment demonstrations into free-space, pre-contact, and contact phases, then up-samples contact-relevant segments during training to improve policy performance.

**Abstract**—Contact-rich manipulation requires force sensitivity, but many robot arms lack dedicated force sensors due to their high cost. We present Neural External Torque Estimation (NEXT), a data-driven method that estimates external joint torques without needing any dedicated force sensors. NEXT trains in 1 minute from only 10 minutes of free-motion data, yet achieves estimates comparable to dedicated joint-torque sensors. NEXT enables force-feedback teleoperation on low-cost arms and improves policy learning through Force-Informed Re-Sampling Training (FIRST), which up-samples pre-contact and contact segments during behavior cloning. Across five long-horizon tasks, FIRST outperforms prior force-aware policies by over 17% in task progress. Together, NEXT and FIRST bring force-aware teleoperation and policy learning to off-the-shelf robots without additional sensing hardware. All parts will be open-sourced. Video results are provided at [learningforce2026.github.io](https://learningforce2026.github.io)

## I. INTRODUCTION

Many everyday manipulation tasks, such as insertion, fastening, and deformable-object handling, require precise force-aware control that humans naturally rely on. In robotics, these subtle contact interactions are difficult to infer reliably from vision alone, making force sensing an essential capability. Dedicated force-torque (FT) sensors are typically available only on expensive platforms such as Franka, Flexiv, or KUKA arms. Many low-cost robot arms lack built-in force sensing, limiting their ability to provide force feedback during data collection and preventing policies from leveraging the force information needed for robust, contact-aware manipulation.

Retrofitting platforms with end-effector force sensors remains challenging and expensive, sometimes costing more than the arms themselves. Conventional piezoresistive FT sensors mounted between the arm and end effector rely on multiple strain gauges attached to precision-machined structures, making them expensive and bulky [3, 5]. Recent capacitive sensors such as CoinFT [10] and magnetic tactile sensors such as ReSkin [4] offer lower-cost alternatives, but their soft sensing materials are not designed to withstand the larger force magnitudes often experienced at the end effector [7].

Instead, we present Neural External Torque estimation (NEXT), a data-driven method for estimating external joint torque, producing high-quality estimates even on low-cost robot arms. NEXT requires only 10 minutes of contact-free trajectories and 1 minute of training to learn a neural network that predicts the torque required for free-space motion. At deployment, NEXT estimates external torque as the residual between measured motor torque, inferred from motor current, and the predicted free-space torque. Importantly, NEXT does not require dedicated force-torque sensors during either data collection or deployment. We show that NEXT produces signals that closely match dedicated joint-torque sensor measurements on high-end robot arms, while outperforming prior estimation methods.

The learned external force signal enables two capabilities: (1) Force-feedback teleoperation on non-sensorized, low-cost

arms (e.g., AgileX Piper, YAM), achieving performance comparable to sensorized systems [26, 32]; and (2) **Force Informed Re-Sampling Training (FIRST)**, a behavior cloning method that uses external force estimates to segment demonstrations into free-space, pre-contact, and contact phases, and up-samples the latter two to improve policy success.

FIRST is motivated by the observation that policy failures concentrate not in free-space motion, but in the brief pre-contact intervals requiring precise alignment as well as in contact phases requiring controlled interaction and fine corrections. We evaluate FIRST on five long-horizon, contact-rich tasks, where it consistently improves success rates over prior baselines, with gains in alignment, recovery, and contact-phase performance.

Together, NEXT and FIRST show that learned free-space inverse dynamics can turn readily available motor currents into a reusable contact signal for commodity robot arms. Our contribution is not just sensorless force estimation or data reweighting in isolation, but showing that the same learned external torque signal enables both force-feedback teleoperation and contact-phase reweighting for improved policy learning, all without dedicated force sensors. Our contributions are:

- **Data-driven external torque estimation without force sensors.** We present NEXT, a neural method for estimating external joint torque without dedicated FT sensors or robot-specific system identification. NEXT trains a recurrent model on only 10 minutes of free-motion data and produces higher-quality estimates than prior methods.
- **Force-feedback teleoperation on low-cost arms.** We use the estimated force signal to enable bilateral teleoperation on non-sensorized low-cost arms, including Piper and YAM, achieving feedback quality comparable to sensorized systems.
- **Force-Informed Re-Sampling Training.** We propose FIRST, a behavior cloning method that uses external torque estimates to identify pre-contact and contact phases in demonstrations, then up-samples these phases during training to improve policy performance.

## II. RELATED WORK

### A. Sensorless External Force Estimation

Estimating external forces without dedicated sensors usually involves modeling robot dynamics and interpreting the residuals between predicted and observed signals as external contact. Traditional approaches based on analytical modeling and system identification [6, 45, 39] can be effective, but they require substantial manual engineering effort and often struggle with actuator nonlinearities like dead zones [20] and torque ripple [51]. Learning-based methods have emerged to address these limitations. Supervised approaches learn from paired proprioceptive and real-world [33] or simulation [24] contact data, but require dedicated data collection and labeling. Unsupervised methods such as autoencoder-based anomaly detection [34] can identify deviations from manipulator free motion, but are primarily suited for contact detection rather

than continuous force estimation. Inverse-dynamics learning has been explored for both individual actuators [51] and complex high-end platforms such as the *da Vinci* system [46], but has not been demonstrated broadly on low-cost robots.

### B. Imitation Learning with Force Information

Recent imitation learning methods incorporate force or torque information in several ways. One line of work uses force-related signals as an additional policy input. For example, TA-VLA [49] uses joint effort information, BiACT [6] uses joint torque, and CompACT [22] conditions the policy on end-effector force. ForceVLA [47] further introduces a mixture-of-experts architecture for context-aware routing across modality-specific experts. Some methods, such as TA-VLA [49] and FoAR [15], use auxiliary objectives to encourage contact-aware representations. A second line of work combines behavior cloning with hybrid force-position or impedance control at inference time [27, 13, 44]. These methods predict both motion and force targets, which are then tracked by a low-level controller to enforce the desired interaction behavior. Finally, FACTR [26] uses force information during training to improve policy generalization to unseen objects. Our method, FIRST, uses external torque not only as policy information, but to infer contact phases and use them to reshape the training distribution. Our ablations further show that pre-contact samples are often more valuable than in-contact samples, providing an empirical guideline for force-aware data reweighting.

### C. Policy Learning with Dataset Reweighting

Policy training dynamics is strongly shaped by how data is used during optimization. In reinforcement learning, prior work has used value and advantage-based weighting to emphasize actions that produce higher returns. AWR [35] weights actions by exponentiated advantages, CRR [41] uses a learned critic to filter or weight actions, and AWAC [31] applies advantage-weighted updates to bridge offline data and online fine-tuning. In imitation learning, data reweighting has also been explored. Re-Mix adapts DoReMi-style [43] Group DRO for robotics [16], while DataMIL uses data model estimators [19] for target-aware selection [11]. DemInf scores state-action mutual information to favor diverse, predictable actions [17], and SCIZOR combines progress prediction and embedding clustering to prune transitions [48]. Other methods leverage learned value functions for filtering: GVL uses vision-language models for value estimation [29], and SARM trains a stage-aware reward model [9]. Finally, rollout-based methods optimize closed-loop performance directly. Demo-SCORE filters data using a success classifier trained on rollouts [8], and CUPID uses influence functions to estimate how individual demonstrations affect expected returns [1].

## III. BACKGROUND

External joint torque  $\tau_{\text{ext}}$  is the joint-space torque induced by physical interactions with the environment, excluding the torque required to move the robot in free space. This signal is

directly available on manipulators with dedicated force-torque sensing, but many arms do not have those sensors.

In the absence of dedicated force sensing, external joint torque can be estimated from motor current measurements and a model of the robot dynamics. Motor current is readily provided on most robot arms, and is approximately related to motor torque through  $\tau_m = KI_m$ , where  $I_m$  is the measured motor current and  $K$  is the torque constant. Given measured motor torque  $\tau_m$  and a known robot dynamics model, the external joint torque can be taken as

$$\tau_{\text{ext}} = \tau_m - \tau_f, \quad \tau_f = \mathbf{M}(\mathbf{q})\ddot{\mathbf{q}} + \mathbf{C}(\mathbf{q}, \dot{\mathbf{q}})\dot{\mathbf{q}} + \mathbf{g}(\mathbf{q}) + \tau_{\text{other}} \quad (1)$$

Here,  $\tau_f$  denotes the torque required to realize the observed motion in free space, i.e., in the absence of external contact. Computing  $\tau_f$  therefore requires an accurate model of the robot mass matrix  $\mathbf{M}$ , Coriolis and centrifugal terms  $\mathbf{C}$ , gravity term  $\mathbf{g}$ , and any other contributing terms  $\tau_{\text{other}}$ .

However, model-based external torque estimation is highly sensitive to errors in estimating  $\tau_f$ . In practice, accurately identifying the inertial, gravitational, and friction parameters of a robot is difficult, especially when only motor-side measurements are available [21, 14]. Moreover, the residual  $\tau_m - \tau_f$  captures not only contact-induced torque, but also unmodeled actuator and transmission effects, including nonlinear friction, stiction, backlash, hysteresis, temperature-dependent drive behavior, sensing noise, torque ripple, deadzones, and saturation [25, 23, 36]. Because many of these effects are history-dependent and only partially observable from instantaneous joint state, free-space modeling errors can limit purely model-based sensorless force estimation.

#### IV. LEARNING EXTERNAL TORQUE ESTIMATION

We present NEXT, Neural External Torque Estimation, a data-driven method for estimating external joint torque  $\tau_{\text{ext}}$ . Importantly, NEXT does not require dedicated force-torque sensors for either data collection or deployment. Instead of computing free-space torque  $\tau_f$  from model-based inverse dynamics, we learn a free-space inverse dynamics neural network directly from joint states, predicting the motor torque that would be observed in the absence of external contact:

$$\hat{\tau}_f = f_\theta(\mathbf{x}), \quad \hat{\tau}_{\text{ext}} = \tau_m - \hat{\tau}_f. \quad (2)$$

where  $f_\theta$  is a neural network parameterized by  $\theta$ , and  $\mathbf{x}$  is a history of proprioceptive observations. At deployment time, we estimate external joint torque  $\hat{\tau}_{\text{ext}}$  by subtracting the predicted free-space torque  $\hat{\tau}_f$  from the measured motor torque  $\tau_m = KI_m$ , illustrated in Figure 2. As shown in Section VII-A and Appendix B, NEXT yields higher-quality external torque estimates than prior methods.

To train the free-space inverse-dynamics model, we collect a dataset of robot arm trajectories executed without external contact,  $\mathcal{D}_{\text{free}} = \{(\mathbf{x}_i, \tau_{m,i})\}_{i=1}^N$ , where  $\tau_{m,i}$  is the measured motor torque at timestep  $i$ . The input  $\mathbf{x}_i$  is constructed from a temporal history of proprioceptive quantities,

$$\mathbf{x}_i = [\mathbf{q}_{i-H:i}, \dot{\mathbf{q}}_{i-H:i}, \Delta\mathbf{q}_{d,i-H:i}], \quad (3)$$

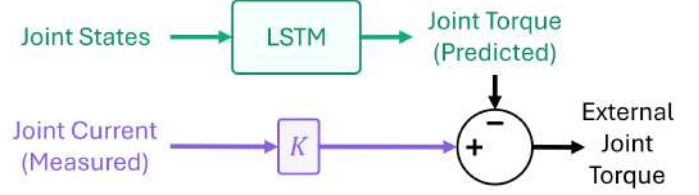


Fig. 2. **External force estimation at deployment.** We obtain measured joint torque by multiplying each joint’s measured current by its torque constant  $K$ . An LSTM trained on free-space data predicts free-space torque, which is subtracted from the measured joint torque to estimate external joint torque.

where  $H$  is the history length,  $\mathbf{q}$  denotes joint position,  $\dot{\mathbf{q}}$  denotes joint velocity, and  $\Delta\mathbf{q}_{d,i}$  is the difference between the commanded joint position target  $\mathbf{q}_{d,i}$  and the current joint position  $\mathbf{q}_i$ .

Because all trajectories in  $\mathcal{D}_{\text{free}}$  are collected without external contact, the measured motor torque equals the free-space torque. Thus, we use  $\tau_{m,i}$  as the supervision target for learning  $\tau_f$ . In practice, the dataset can be collected from approximately 10 minutes of free-space robot motion, generated either by teleoperation, motion planning, or replay of known free-space trajectories. Thus, this procedure is easy to adopt on new robot platforms, as it does not require force-torque sensing, payload calibration, or manual system identification.

We instantiate  $f_\theta$  as an LSTM-based sequence model, which maps the proprioceptive history  $\mathbf{x}_i$  to a predicted free-space torque  $\hat{\tau}_{f,i}$ . The use of temporal history allows the model to capture effects that are difficult to represent with an instantaneous rigid-body dynamics model, such as hysteresis, actuator dynamics, frictional memory, and controller-dependent tracking behavior. The model is trained with an  $L_2$  regression loss,

$$\mathcal{L}(\theta) = \frac{1}{N} \sum_{i=1}^N \|f_\theta(\mathbf{x}_i) - \tau_{m,i}\|_2^2 \quad (4)$$

After training, the NEXT model predicts the free-space torque  $\hat{\tau}_f$ , allowing external torque to be estimated as the residual  $\tau_m - \hat{\tau}_f$ . Training is lightweight, requiring only 1 minute on an NVIDIA RTX 3090. Details on the model architecture and hyperparameters are provided in Appendix A.

#### V. FORCE-INFORMED RE-SAMPLING TRAINING

To leverage external joint torque for improved behavior cloning, we introduce FIRST, Force-Informed Re-Sampling Training. FIRST is motivated by the observation that failures in contact-rich tasks often occur near contact: pre-contact phases require precise alignment, while contact phases demand controlled interaction and fine corrective actions. FIRST therefore re-weights the training data mixture to sample pre-contact and contact segments more frequently, focusing learning on the most failure-prone portions of the trajectory. In Section VII-B, we show that FIRST improves task success by over 17% and reduces validation loss on pre-contact and contact phases compared to standard behavior cloning, explaining its downstream performance gains.

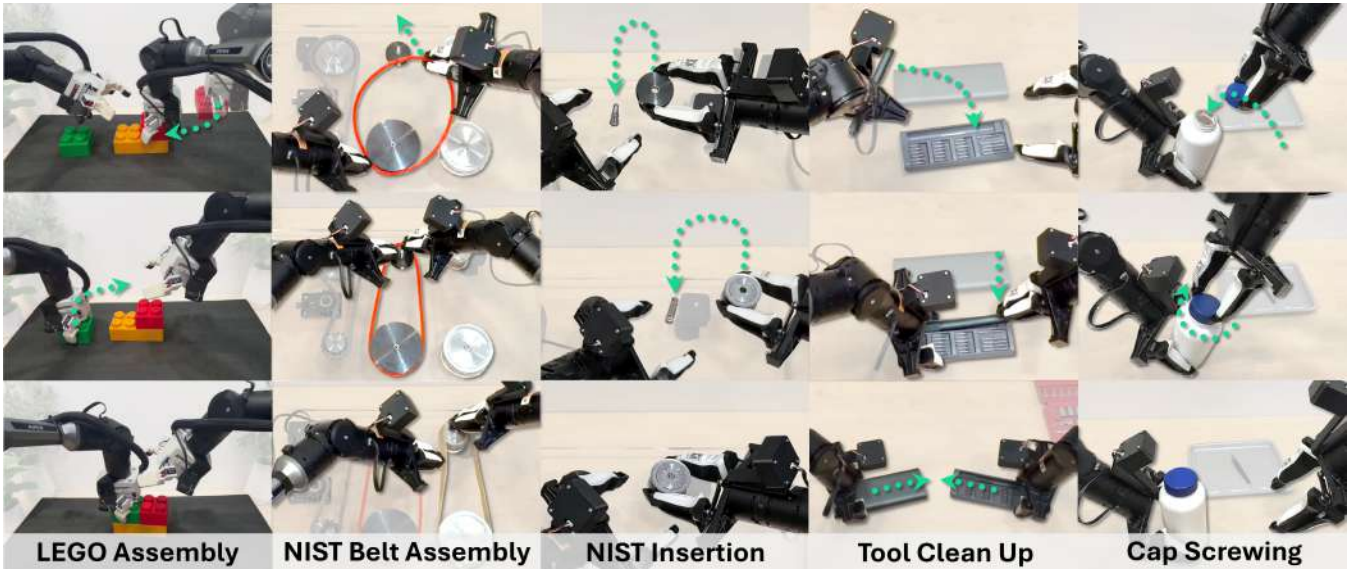


Fig. 3. FIRST is evaluated on five long-horizon, contact-rich tasks. Each task comprises multiple stages, many requiring precise alignment or fine-grained, force-sensitive adjustments.

**Problem Statement.** We consider a policy that, at timestep  $t$ , conditions on an observation  $\mathbf{o}_t = (\mathbf{I}_t^{1:M}, \mathbf{q}_t, \hat{\tau}_{\text{ext},t})$ , consisting of  $M$  camera images, joint position  $\mathbf{q}_t$ , and the estimated external joint torque  $\hat{\tau}_{\text{ext},t}$  from Section IV. The policy predicts a chunk of future joint position targets,  $\hat{\mathbf{a}}_{t:t+k} \sim \pi_\theta(\cdot | \mathbf{o}_t)$ , where  $\mathbf{a}_{t:t+k}$  denotes a corresponding action chunk.

We train the policy by behavior cloning on expert demonstrations  $\mathcal{D} = \{(\mathbf{o}_t, \mathbf{a}_{t:t+k})\}$ . We instantiate  $\pi_\theta$  as a flow-matching policy parameterized by a conditional velocity field  $v_\theta$ . Let  $\mathbf{x}_1 = \mathbf{a}_{t:t+k}$  be an expert action chunk,  $\mathbf{x}_0 \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$  be noise, and  $\mathbf{x}_\alpha = (1 - \alpha)\mathbf{x}_0 + \alpha\mathbf{x}_1$  for  $\alpha \sim \mathcal{U}(0, 1)$ . The policy is trained with a velocity prediction objective [28], where  $v_\theta$  predicts the flow from noise  $\mathbf{x}_0$  to the expert action chunk  $\mathbf{x}_1$ :

$$\mathcal{L}_{\text{FM}}(\theta) = \mathbb{E} \left[ \|v_\theta(\mathbf{x}_\alpha, \alpha, \mathbf{o}_t) - (\mathbf{x}_1 - \mathbf{x}_0)\|_2^2 \right]. \quad (5)$$

**Segmenting Trajectories into Contact Phases.** To begin, FIRST labels each datapoint in the demonstration dataset as free motion, pre-contact, or contact. Contact corresponds to timesteps where the robot is physically interacting with the environment, pre-contact corresponds to the one-second window before contact onset, and all remaining timesteps are labeled as free motion.

We identify contact from the estimated external joint torque  $\hat{\tau}_{\text{ext},t}$  using the scalar contact score  $f_t = \|\hat{\tau}_{\text{ext},t}\|_1$ . To avoid noisy transitions, we use hysteresis thresholding: the binary contact indicator  $c_t$  turns on when  $f_t \geq T_{\text{high}}$ , turns off when  $f_t \leq T_{\text{low}}$ , and otherwise retains its previous value.

Contact onsets are timesteps where the label  $c_t$  switches from non-contact to contact,

$$\mathcal{T}_{\text{onset}} = \{t \mid c_{t-1} = 0, c_t = 1\}. \quad (6)$$

Given a control frequency  $F$ , we assign the phase label  $s_t$  by marking contact timesteps as C, the  $F$  timesteps before

each contact onset as PC, and all remaining timesteps as F:

$$s_t = \begin{cases} \text{C}, & c_t = 1, \\ \text{PC}, & \exists t_o \in \mathcal{T}_{\text{onset}}: t_o - F \leq t < t_o, \\ \text{F}, & \text{otherwise.} \end{cases} \quad (7)$$

**Up-Sampling Pre-Contact and Contact Phases in Training.** After assigning each datapoint a phase label  $s_t \in \{\text{Free Space}, \text{Pre-Contact}, \text{Contact}\}$ , FIRST modifies only the sampling distribution used for batch construction. We assign phase-dependent sampling weights  $w_{\text{F}}$ ,  $w_{\text{PC}}$ , and  $w_{\text{C}}$  to free-space, pre-contact, and contact datapoints, respectively, with  $w_{\text{PC}}, w_{\text{C}} > w_{\text{F}}$ . During training, datapoints are sampled with probability  $p_t = w(s_t) / \sum_{j=1}^N w(s_j)$ , where  $w(s_t)$  denotes the weight associated with the phase label of datapoint  $t$ . This weighted random sampling increases the fraction of pre-contact and contact examples in each minibatch, emphasizing trajectory segments where precise alignment and force-sensitive corrections are most critical.

## VI. EXPERIMENTS

### A. External Force Estimation and Force-Feedback Teleoperation

We evaluate external torque estimation on a Franka arm in both contact and free-space settings. We use the Franka because it has built-in joint torque sensors and a factory-calibrated external torque estimate, providing a high-quality reference which we use as ground truth during contact. For contact, we collect 5 minutes of held-out data, unseen during training, while a human applies external forces to the robot. We report the  $L_1$  error averaged across joints between the ground-truth sensor signal and estimates from NEXT (ours), FILIC [13], and a disturbance observer (DO) [30]. For free-space motion, where the true external torque should be zero,

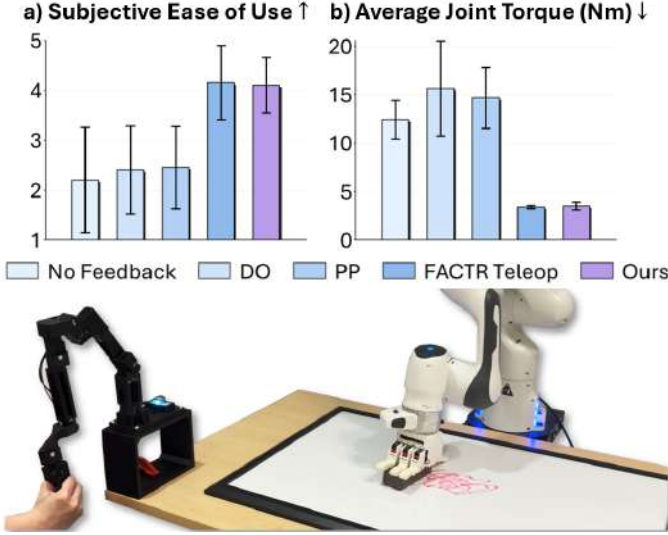


Fig. 4. **User study.** NEXT-based external force estimates improve teleoperation over the baselines and remain comparable to FACTR Teleop, which relies on dedicated external force sensors. Lower applied joint torque corresponds to less unnecessary exertion.

we collect 5 minutes of held-out data without contact and use zero torque as ground truth, treating the Franka sensor signal as another baseline. Baseline details are provided in Appendix D. We repeat the free-space evaluation on the AgileX Piper, a low-cost arm without force sensors, again using zero torque as ground truth. Since the Piper lacks external torque sensing, we cannot repeat the contact evaluation; instead, we assess contact performance through a force-feedback teleoperation user study.

**Force-Feedback Teleoperation User Study.** We evaluate whether the learned external torque estimate provides useful teleoperation feedback. In a user study with 20 participants, users perform a wiping task on the Franka arm, as shown in Figure 4, with five feedback conditions: no feedback, disturbance-observer based feedback (DO), leader-follower position-position feedback [18] (PP), FACTR Teleop using Franka’s built-in torque sensors, and FACTR Teleop with NEXT’s external torque estimates. Note that both FACTR Teleop variants implement the following force-feedback control law [26]:

$$\tau_{feedback} = \mathbf{K}_{f,p} \tau_{ext} \quad (8)$$

where  $\tau_{feedback}$  is joint torque applied to the leader arm, and  $\mathbf{K}_{f,p}$  scales the external joint torque  $\tau_{ext}$  sensed by the follower arm. In this study,  $\tau_{ext}$  either comes from Franka’s factory external torque estimate or comes from NEXT. We collect 1–5 ease-of-use ratings and measure the applied joint torque. We repeat the study on the Piper arm. Details about the baselines are described in Appendix E.

### B. Imitation Learning with Force Information

**Experiment Settings.** We evaluate FIRST against baselines on five long-horizon, contact-rich manipulation tasks, shown

TABLE I  
AVERAGE  $L_1$  JOINT TORQUE ERROR ON THE FRANKA.

| Motion     | External Torque Method    | $L_1$ Error (Nm)                    |
|------------|---------------------------|-------------------------------------|
| Contact    | External Sensor (GT)      | $0.000 \pm 0.000$                   |
|            | FILIC [13]                | $4.395 \pm 1.531$                   |
|            | Disturbance Observer [30] | $1.471 \pm 0.761$                   |
|            | <b>NEXT (Ours)</b>        | <b><math>0.547 \pm 0.348</math></b> |
| Free Space | Free Space (GT)           | $0.000 \pm 0.000$                   |
|            | FILIC [13]                | $2.460 \pm 1.358$                   |
|            | Disturbance Observer [30] | $2.429 \pm 1.298$                   |
|            | External Sensor           | $0.449 \pm 0.208$                   |
|            | <b>NEXT (Ours)</b>        | <b><math>0.414 \pm 0.278</math></b> |

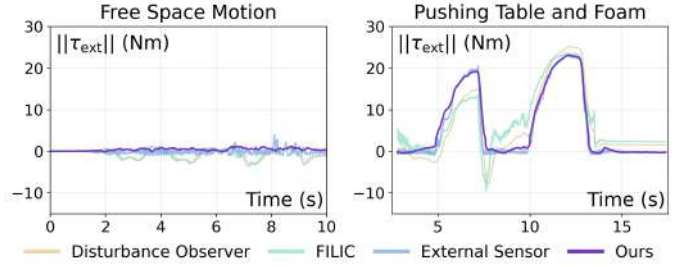


Fig. 5. External torque estimation results. **Left:** In free space, the true external torque should remain near zero. NEXT remains closer to zero and is less noisy than the external sensor, while FILIC and the disturbance observer drift. **Right:** During contact, NEXT closely tracks the external sensor, whereas the baselines deviate substantially.

in Figure 3. Detailed task descriptions are provided in Appendix G. For each task, we collect 250 demonstrations on a bimanual Piper setup with four RGB cameras: two wrist cameras and two overhead cameras. Each policy is evaluated over 20 rollouts per task. We report task progress, with scoring criteria defined in Appendix H.

**Policy Details and Baselines.** We compare FIRST with four baselines while keeping the policy architecture and behavior cloning objective fixed. All methods use a flow-matching policy trained for 20 epochs, with an architecture adapted from Multi-modal Diffusion Transformer [37] and an action chunk size of 30. We also repeat the experiments with Action Chunking Transformer (ACT) [50]. Implementation details for both are provided in Appendix F. We compare the following variants:

- **Base Policy:** Policy takes in images and robot joint positions as inputs, with no force information.
- **Base Policy + Torque:** Base policy with the learned external torque estimate as additional input.
- **FACTR [26]:** Uses external torque and blurs input images during training with a noise schedule.
- **TA-VLA [49]:** Uses an auxiliary objective to reconstruct the input external torque.
- **FIRST (Ours):** Uses learned external torque as input and up-samples pre-contact and contact data during training. We ablate up-sampling ratios in Appendix M.

## VII. RESULTS

In our external force estimation experiments, we ask: (i) How do NEXT’s external joint torque estimates compare to ground-truth sensor measurements and baselines? (ii) How effective is NEXT for force-feedback teleoperation compared to alternative feedback methods?

In our imitation learning experiments, we ask: (i) How does FIRST compare to other baselines leveraging force? (ii) Which contact phases should FIRST up-sample for best performance?

### A. NEXT Evaluation

**NEXT predicts external joint torque comparable to dedicated sensors.** We present quantitative results for NEXT on the Franka system in Table I. Under external contact, NEXT closely matches the Franka external joint torque sensor, achieving an average  $L_1$  error of only 0.547 Nm, which is 87.6% lower than FILIC and 62.8% lower than the disturbance observer. This trend is also shown qualitatively on the right side of Figure 5: NEXT tracks the sensor signal closely, while the baselines deviate substantially, especially during contact transitions.

In free space, where the true external torque should be zero, we treat the Franka’s external torque sensor signals as an additional baseline; NEXT achieves the lowest error, even outperforming this dedicated sensor. As shown on the left side of Figure 5, its estimate is less noisy and remains closer to zero than the sensor. This can be attributed to NEXT learning hardware-specific effects directly from data collected on the hardware itself, yielding high-quality force estimates.

We observe similar results on the AgileX Piper arm, where NEXT achieves an average joint  $L_1$  error of only 0.018 Nm, substantially lower than the baselines; full results are provided in Table V in Appendix F. We also ablate the NEXT model architecture in Appendix B, showing that the LSTM outperforms both GRU and MLP variants.

**NEXT enables high-quality force-feedback teleoperation on low-cost arms.** In the Franka user study, participants rated NEXT-based force feedback as easier to use than the baselines, shown in Figure 4 (a). NEXT also reduced the force required to complete the task, performing comparably to FACTR Teleop, which uses dedicated joint torque sensors, seen in Figure 4 (b). Lower applied force indicates greater force awareness and less unnecessary exertion. On the Piper arm, which lacks dedicated force sensors, NEXT also outperforms the baselines, as detailed in Appendix F. Overall, these results show that NEXT provides high-quality force feedback without needing force sensors, comparable to sensor-based methods.

### B. FIRST Evaluation

**FIRST leads to improved policy performance.** In Figure 7, we present the main quantitative results for FIRST against baselines for the flow-matching policy, with ACT results shown in Appendix K. FIRST achieves the highest task progress across all five real-world tasks. The improvement is especially clear in long-horizon, multi-stage tasks,

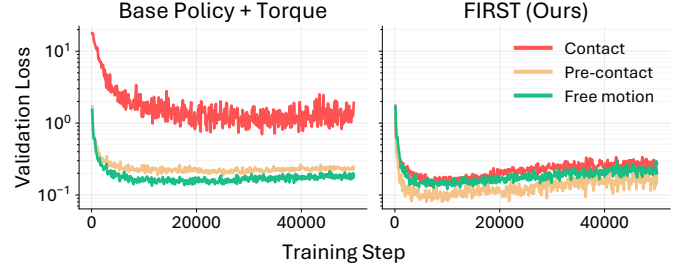


Fig. 6. Default sampling yields higher validation loss on pre-contact and contact phases. By up-sampling these phases during training, FIRST reduces their validation losses.

TABLE II  
UP-SAMPLING ABLATION. C DENOTES CONTACT, PC DENOTES PRE-CONTACT, AND PC+C DENOTES PRE-CONTACT PLUS CONTACT.

| Task           | Up-sampling Phase |              |              |
|----------------|-------------------|--------------|--------------|
|                | C                 | PC           | PC+C         |
| LEGO Assembly  | 0.697             | <b>0.913</b> | 0.875        |
| NIST Belt      | 0.494             | <b>0.767</b> | 0.649        |
| NIST Insertion | 0.800             | <b>0.933</b> | <b>0.933</b> |
| Tool Cleanup   | 0.650             | <b>0.805</b> | 0.775        |
| Cap Screwing   | 0.707             | 0.673        | <b>0.825</b> |
| <b>Average</b> | 0.670             | <b>0.818</b> | 0.811        |

such as NIST Belt, where successful execution requires reliable transitions between free motion, pre-contact alignment, and contact-rich manipulation. Qualitatively, we observe that FIRST produces more robust recovery behaviors before key contact events, such as fine-grained realignment motions before insertion, wrapping, or screwing.

Figure 6 further explains this improvement through phase-wise validation loss for the NIST Belt task. For the Base Policy + Torque baseline, loss is higher on pre-contact and contact samples than on free-motion samples, indicating that these phases are harder to model. By up-sampling pre-contact and contact data, FIRST reduces their validation losses and brings them closer to the free-motion loss, focusing training on the phases most critical for contact-rich manipulation.

**FIRST up-sampling ablation.** We ablate which phases to up-sample for best policy performance. As shown in Table II, up-sampling pre-contact samples achieves the highest average task progress,

suggesting that the moments before contact are especially important as they determine reaching the correct pose and alignment. Up-sampling both pre-contact and contact phases is also competitive, and performs best on Cap Screwing, where sustained contact is essential. In contrast, up-sampling only contact samples does not improve performance. Overall, these results show that emphasizing contact-relevant phases, especially pre-contact, improves long-horizon manipulation. Additional results on the highest performing up-sampling ratio for each task are provided in Appendix M.

**Why not use raw joint current as input?** Since low-cost arms often expose motor current, we compare policies

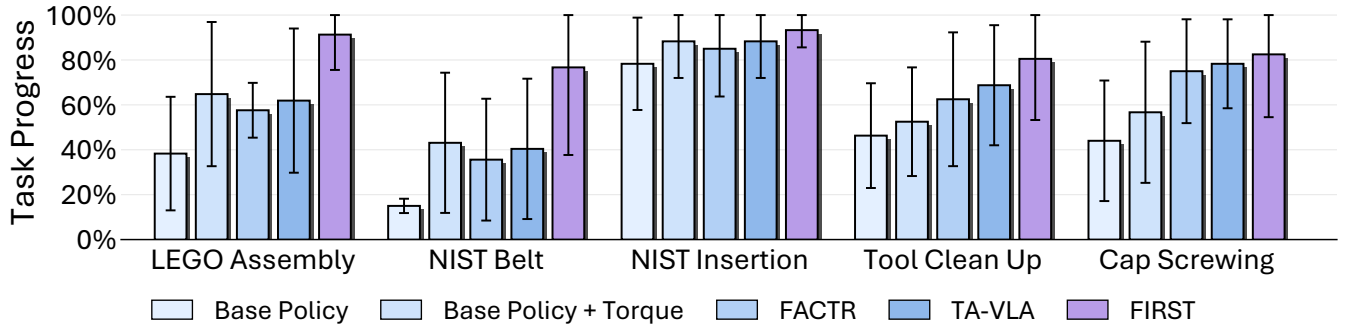


Fig. 7. Task progress across five contact-rich manipulation tasks using the flow-matching policy.

conditioned on raw current with those conditioned on NEXT-estimated external torque. As shown in Table X, raw current improves NIST Insertion but not Cap Screwing, while NEXT-estimated external torque performs best on both tasks. This suggests that subtracting learned free-space dynamics yields a cleaner contact representation than raw current alone, further highlighting the usefulness of NEXT. Full results and attention analysis are in Appendix O.

## VIII. CONCLUSION AND LIMITATIONS

We present NEXT, a data-driven method for estimating external joint torque without dedicated force sensors. With only 10 minutes of free-space data and 1 minute of training, NEXT produces estimates comparable to dedicated sensor measurements and enables force-feedback teleoperation on low-cost arms. We further introduce FIRST, which uses NEXT torque estimates to up-sample pre-contact and contact segments during behavior cloning, improving performance on contact-rich tasks by emphasizing the phases most critical for alignment and interaction.

**Limitations.** NEXT has two main limitations. First, its absolute torque scale depends on the motor torque constant  $K$ , which is typically provided by the manufacturer or derived from a current-torque curve. If  $K$  is inaccurate, absolute scale requires calibration against an external force sensor. In practice, this is less limiting for force-feedback teleoperation, where forces are often scaled for comfort [26], and also for policy learning, where inputs are normalized. Second, NEXT is robot-specific and may require retraining across arms due to hardware variation.

## REFERENCES

- [1] Christopher Agia, Rohan Sinha, Jingyun Yang, Rika Antonova, Marco Pavone, Haruki Nishimura, Masha Itkina, and Jeannette Bohg. Cupid: Curating data your robot loves with influence functions. In Joseph Lim, Shuran Song, and Hae-Won Park, editors, *Proceedings of The 9th Conference on Robot Learning*, volume 305 of *Proceedings of Machine Learning Research*, pages 2907–2932. PMLR, 27–30 Sep 2025. URL <https://proceedings.mlr.press/v305/agia25a.html>.
- [2] AgileX Robotics. agx\_arm\_urdf: Agilix robot arm urdf models. [https://github.com/agilexrobotics/agx\\_arm\\_urdf](https://github.com/agilexrobotics/agx_arm_urdf), 2026. Accessed: 2026-05-11.
- [3] ATI Industrial Automation. Gamma Force/Torque sensor. [https://www.ati-ia.com/products/ft/ft\\_models.aspx?id=gamma](https://www.ati-ia.com/products/ft/ft_models.aspx?id=gamma), 2026. Accessed: April 28, 2026.
- [4] Raunaq Bhirangi, Tess Hellebrekers, Carmel Majidi, and Abhinav Gupta. Reskin: versatile, replaceable, lasting tactile skins. In *CoRL*, 2021.
- [5] Bota Systems. 6-Axis Force Torque Sensors for Robotics. <https://www.botasys.com/force-torque-sensors>, 2024. Accessed: April 28, 2026.
- [6] Thanpimon Buamanee, Masato Kobayashi, Yuki Urinishi, and Haruo Takemura. Bi-act: Bilateral control-based imitation learning via action chunking with transformer. In *2024 IEEE International Conference on Advanced Intelligent Mechatronics (AIM)*, pages 410–415, 2024. doi: 10.1109/AIM55361.2024.10637173.
- [7] Max Yiye Cao, Stephen Laws, and Ferdinando Rodriguez y Baena. Six-axis force/torque sensors for robotics applications: A review. *IEEE Sensors Journal*, 21(24): 27238–27251, 2021.
- [8] Annie S. Chen, Alec M. Lessing, Yuejiang Liu, and Chelsea Finn. Curating demonstrations using online experience. In *Proceedings of Robotics: Science and Systems*, 2025. doi: 10.15607/RSS.2025.XXI.071.
- [9] Qianzhong Chen, Justin Yu, Mac Schwager, Pieter Abbeel, Fred Shentu, and Philipp Wu. SARM: Stage-aware reward modeling for long horizon robot manipulation. In *The Fourteenth International Conference on Learning Representations*, 2026. URL <https://openreview.net/forum?id=aemqAxScI9>.
- [10] Hojung Choi, Jun En Low, Tae Myung Huh, Seongheon Hong, Gabriela A Uribe, Kenneth AW Hoffmann, Julia Di, Tony G Chen, Andrew A Stanley, and Mark R Cutkosky. Coinft: A coin-sized, capacitive 6-axis force torque sensor for robotic applications. *arXiv preprint arXiv:2503.19225*, 2025.
- [11] Shivin Dass, Alaa Khaddaj, Logan Engstrom, Aleksander Madry, Andrew Ilyas, and Roberto Martín-Martín. DataMIL: Selecting data for robot imitation learning with

- datamodels. In *The Fourteenth International Conference on Learning Representations*, 2026. URL <https://openreview.net/forum?id=AcTsKglDdh>.
- [12] Claudio Gaz, Marco Cognetti, Alexander Oliva, Paolo Robuffo Giordano, and Alessandro De Luca. Dynamic identification of the franka emika panda robot with retrieval of feasible parameters using penalty-based optimization. *IEEE Robotics and Automation Letters*, 4(4): 4147–4154, 2019. doi: 10.1109/LRA.2019.2931248.
- [13] Haizhou Ge, Yufei Jia, Zheng Li, Yue Li, Zhixing Chen, Ruqi Huang, and Guyue Zhou. Filic: Dual-loop force-guided imitation learning with impedance torque control for contact-rich manipulation tasks, 2025. URL <https://arxiv.org/abs/2509.17053>.
- [14] Sami Haddadin, Alessandro De Luca, and Alin Albu-Schäffer. Robot collisions: A survey on detection, isolation, and identification. *IEEE Transactions on Robotics*, 33(6):1292–1312, 2017.
- [15] Zihao He, Hongjie Fang, Jingjing Chen, Hao-Shu Fang, and Cewu Lu. Foar: Force-aware reactive policy for contact-rich robotic manipulation. *IEEE Robotics and Automation Letters*, 10(6):5625–5632, 2025. doi: 10.1109/LRA.2025.3560871.
- [16] Joey Hejna, Chethan Anand Bhateja, Yichen Jiang, Karl Pertsch, and Dorsa Sadigh. Remix: Optimizing data mixtures for large scale imitation learning. In Pulkit Agrawal, Oliver Kroemer, and Wolfram Burgard, editors, *Proceedings of The 8th Conference on Robot Learning*, volume 270 of *Proceedings of Machine Learning Research*, pages 145–164. PMLR, 06–09 Nov 2025. URL <https://proceedings.mlr.press/v270/hejna25a.html>.
- [17] Joey Hejna, Suvir Mirchandani, Ashwin Balakrishna, Annie Xie, Ayzaan Wahid, Jonathan Tompson, Pannag Sanketi, Dhruv Shah, Coline Devin, and Dorsa Sadigh. Robot data curation with mutual information estimators. In *Robotics: Science and Systems (RSS)*, 2025.
- [18] Peter F. Hokayem and Mark W. Spong. Bilateral teleoperation: An historical survey. *Automatica*, 42(12):2035–2057, 2006. ISSN 0005-1098. doi: <https://doi.org/10.1016/j.automatica.2006.06.027>. URL <https://www.sciencedirect.com/science/article/pii/S0005109806002871>.
- [19] Andrew Ilyas, Sung Min Park, Logan Engstrom, Guillaume Leclerc, and Aleksander Madry. Datamodels: Understanding predictions with data and data with predictions. In *ICML*, 2022.
- [20] Koki Inami, Koki Yamane, and Sho Sakaino. Loss function considering dead zone for neural networks, 2024. URL <https://arxiv.org/abs/2402.00393>.
- [21] A. Jubien, M. Gautier, and A. Janot. Dynamic identification of the kuka lwr robot using motor torques and joint torque sensors data. In *19th IFAC World Congress*, 2014.
- [22] Tatsuya Kamijo, Cristian C. Beltran-Hernandez, and Masashi Hamaya. Learning variable compliance control from a few demonstrations for bimanual robot with haptic feedback teleoperation system. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2024.
- [23] Nenad M. Kircanski and Andrew A. Goldenberg. An experimental study of nonlinear stiffness, hysteresis, and friction effects in robot joints with harmonic drives and torque sensors. *The International Journal of Robotics Research*, 16(2):214–239, 1997.
- [24] Jacky Liang and Oliver Kroemer. Contact localization for robot arms in motion without torque sensing. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pages 6322–6328. IEEE, 2021.
- [25] Magnus Linderöth, Andreas Stolt, Anders Robertsson, and Rolf Johansson. Robotic force estimation using motor torques and modeling of low velocity friction disturbances. In *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3550–3556, 2013.
- [26] Jason Jingzhou Liu, Yulong Li, Kenneth Shaw, Tony Tao, Russ Salakhutdinov, and Deepak Pathak. FACTR: Force-Attending Curriculum Training for Contact-Rich Policy Learning. In *Proceedings of Robotics: Science and Systems*, Los Angeles, CA, USA, June 2025. doi: 10.15607/RSS.2025.XXI.079.
- [27] Wenhai Liu, Junbo Wang, Yiming Wang, Weiming Wang, and Cewu Lu. Forcemimic: Force-centric imitation learning with force-motion capture system for contact-rich manipulation. In *2025 IEEE International Conference on Robotics and Automation (ICRA)*, 2025.
- [28] Xingchao Liu, Chengyue Gong, and qiang liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. In *The Eleventh International Conference on Learning Representations*, 2023. URL <https://openreview.net/forum?id=XVjTT1nw5z>.
- [29] Yecheng Jason Ma, Joey Hejna, Chuyuan Fu, Dhruv Shah, Jacky Liang, Zhuo Xu, Sean Kirmani, Peng Xu, Danny Driess, Ted Xiao, et al. Vision language models are in-context value learners. In *The Thirteenth International Conference on Learning Representations*, 2024.
- [30] Shamil Mamedov and Stanislav Mikhel. Practical aspects of model-based collision detection. *Frontiers in Robotics and AI*, 7:571574, 2020.
- [31] Ashvin Nair, Abhishek Gupta, Murtaza Dalal, and Sergey Levine. Awac: Accelerating online reinforcement learning with offline datasets. *arXiv preprint arXiv:2006.09359*, 2020.
- [32] Steven Oh, Tomoya Takahashi, Cristian C. Beltran-Hernandez, Yuki Kuroda, and Masashi Hamaya. A soft wrist with anisotropic and selectable stiffness for robust robot learning in contact-rich manipulation, 2026. URL <https://arxiv.org/abs/2602.14434>.
- [33] Kyu Min Park, Jihwan Kim, Jinhyuk Park, and Frank C. Park. Learning-based real-time detection of robot collisions without joint torque sensors. *IEEE Robotics and Automation Letters*, 6(1):103–110, 2021. doi: 10.1109/LRA.2020.3033269.

- [34] Kyu Min Park, Younghyo Park, Sangwoong Yoon, and Frank C. Park. Collision detection for robot manipulators using unsupervised anomaly detection algorithms. *IEEE/ASME Transactions on Mechatronics*, 27(5):2841–2851, 2022. doi: 10.1109/TMECH.2021.3119057.
- [35] Xue Bin Peng, Aviral Kumar, Grace Zhang, and Sergey Levine. Advantage-weighted regression: Simple and scalable off-policy reinforcement learning. *arXiv preprint arXiv:1910.00177*, 2019.
- [36] Moritz Reuss, Niels van Duijkeren, Robert Krug, Philipp Becker, Vaisakh Shaj, and Gerhard Neumann. End-to-end learning of hybrid inverse dynamics models for precise and compliant impedance control. In *Robotics: Science and Systems*, 2022.
- [37] Moritz Reuss, Ömer Erdiñç Yağmurlu, Fabian Wenzel, and Rudolf Lioutikov. Multimodal diffusion transformer: Learning versatile behavior from multimodal goals. In *Proceedings of Robotics: Science and Systems*, Delft, Netherlands, July 2024. doi: 10.15607/RSS.2024.XX.121.
- [38] Kenneth Shaw and Deepak Pathak. Leap hand v2: Dexterous, low-cost anthropomorphic hybrid rigid soft hand for robot learning. In *2nd Workshop on Dexterous Manipulation: Design, Perception and Control (RSS)*.
- [39] Haochen Shi, Songbo Hu, Yifan Hou, Weizhuo Wang, Karen Liu, and Shuran Song. Minimalist compliance control, 2026. URL <https://arxiv.org/abs/2603.00913>.
- [40] Oriane Siméoni, Huy V Vo, Maximilian Seitzer, Federico Baldassarre, Maxime Oquab, Cijo Jose, Vasil Khalidov, Marc Szafraniec, Seungeun Yi, Michaël Ramamonjisoa, et al. Dinov3. *arXiv preprint arXiv:2508.10104*, 2025.
- [41] Ziyu Wang, Alexander Novikov, Konrad Zolna, Jost Tobias Springenberg, Scott Reed, Bobak Shahriari, Noah Siegel, Josh Merel, Caglar Gulcehre, Nicolas Heess, and Nando de Freitas. Critic regularized regression. *Advances in Neural Information Processing Systems*, 33: 7768–7778, 2020.
- [42] Philipp Wu, Yide Shentu, Zhongke Yi, Xingyu Lin, and Pieter Abbeel. GELLO: A general, low-cost, and intuitive teleoperation framework for robot manipulators. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 12156–12163. IEEE, 2024. doi: 10.1109/IROS58592.2024.10801581.
- [43] Sang Michael Xie, Hieu Pham, Xuanyi Dong, Nan Du, Hanxiao Liu, Yifeng Lu, Percy S Liang, Quoc V Le, Tengyu Ma, and Adams Wei Yu. Doremi: Optimizing data mixtures speeds up language model pretraining. *Advances in Neural Information Processing Systems*, 36: 69798–69818, 2023.
- [44] Han Xue, Jieji Ren, Wendi Chen, Gu Zhang, Yuan Fang, Guoying Gu, Huazhe Xu, and Cewu Lu. Reactive diffusion policy: Slow-fast visual-tactile policy learning for contact-rich manipulation. In *Proceedings of Robotics: Science and Systems (RSS)*, 2025.
- [45] Koki Yamane, Yunhan Li, Masashi Konosu, Koki Inami, Junji Oaki, Toshiaki Tsuji, and Sho Sakaino. Design and experimental validation of sensorless 4-channel bilateral teleoperation for low-cost manipulators, 2026. URL <https://arxiv.org/abs/2507.06174>.
- [46] Nural Yilmaz, Jie Ying Wu, Peter Kazanzides, and Ugur Tumerdem. Neural network based inverse dynamics identification and external force estimation on the da vinci research kit. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1387–1393. IEEE, 2020.
- [47] Jiawen Yu, Hairuo Liu, Qiaojun Yu, Jieji Ren, Ce Hao, Haitong Ding, Guangyu Huang, Guofan Huang, Yan Song, Panpan Cai, Wenqiang Zhang, and Cewu Lu. ForceVLA: Enhancing VLA models with a force-aware moe for contact-rich manipulation. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2026. URL <https://openreview.net/forum?id=2845H8Ua5D>.
- [48] Yu Zhang, Yuqi Xie, Huihan Liu, Rutav Shah, Michael Wan, Linxi Fan, and Yuke Zhu. Scizor: Self-supervised data curation for large-scale imitation learning. In *IEEE International Conference on Robotics and Automation (ICRA)*, 2026.
- [49] Zongzheng Zhang, Haobo Xu, Zhuo Yang, Chenghao Yue, Zehao Lin, Huan-ang Gao, Ziwei Wang, and Hao Zhao. Elucidating the design space of torque-aware vision-language-action models. In Joseph Lim, Shuran Song, and Hae-Won Park, editors, *Proceedings of The 9th Conference on Robot Learning*, volume 305 of *Proceedings of Machine Learning Research*, pages 4019–4037. PMLR, 27–30 Sep 2025. URL <https://proceedings.mlr.press/v305/zhang25k.html>.
- [50] Tony Z. Zhao, Vikash Kumar, Sergey Levine, and Chelsea Finn. Learning Fine-Grained Bimanual Manipulation with Low-Cost Hardware. In *Proceedings of Robotics: Science and Systems*, Daegu, Republic of Korea, July 2023. doi: 10.15607/RSS.2023.XIX.016.
- [51] Alvin Zhu, Yusuke Tanaka, Fadi Rafeedi, and Dennis Hong. Cycloidal quasi-direct drive actuator designs with learning-based torque estimation for legged robotics. In *2025 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1–8. IEEE, 2025.

## APPENDIX

### A. Free-Space Inverse Dynamics Model Architecture

We use an LSTM-based sequence model for free-space inverse dynamics prediction. Rather than maintaining a persistent hidden state across the entire trajectory, we adopt a stateless sliding-window formulation. The network therefore implicitly models temporal dynamics within the look-back window, while the hidden state is reset for each sequence. This formulation provides recent motion context without accumulating hidden-state drift over long trajectories. The local history allows the model to capture velocity-dependent, actuator-dependent, and unmodeled friction effects without explicitly estimating acceleration. For deployment, the estimator can be evaluated using a rolling buffer of the most recent  $N$  observations. External joint torque is then estimated as the residual between the measured joint torque and the predicted free-space torque. The network architecture and training hyperparameters are listed in Table III.

On a single thread of an Intel Core i9-10920X CPU, the free-space inverse dynamics model required approximately 1.76 ms per forward pass, corresponding to an effective update rate of 568 Hz. This enables high-frequency force feedback teleoperation. In practice, we run the model at 100 Hz.

TABLE III  
NETWORK ARCHITECTURE AND TRAINING HYPERPARAMETERS FOR THE FREE-SPACE INVERSE DYNAMICS MODEL USED IN NEXT.

| Hyperparameter     | Value                                                              |
|--------------------|--------------------------------------------------------------------|
| Recurrent backbone | 2-layer LSTM with hidden size 128                                  |
| Prediction head    | 2-layer MLP with hidden size 256                                   |
| Dropout            | 0.1                                                                |
| Loss function      | Mean squared error (MSE)                                           |
| Optimizer          | AdamW                                                              |
| Base learning rate | $1 \times 10^{-3}$                                                 |
| Weight decay       | $1 \times 10^{-6}$                                                 |
| Gradient clipping  | 1.0                                                                |
| Early stopping     | Validation loss; patience 20; warmup 10; $\Delta_{\min} = 10^{-5}$ |
| Hardware           | NVIDIA RTX 3090, 24 GB                                             |
| Training time      | Approximately 1 minute with early stopping                         |
| Data frequency     | 100 Hz                                                             |

### B. Free-Space Inverse Dynamics Network Ablations

To select the free-space inverse dynamics model used in NEXT, we ablate model architecture, input modality, and history length using external torque estimation error in both free-space and contact settings. Prior work has used MLPs for torque estimation on specialized *da Vinci* systems [46] and GRUs for single-actuator external torque estimation [51]. We therefore compare MLP, GRU, and LSTM architectures.

Table IV reports the architecture ablation results, with values corresponding to the average joint  $L_1$  validation error between predicted free-space torque and ground-truth free-space torque on the held-out set. MLPs perform worse than recurrent models, showing the importance of temporal history for free-space torque estimation. GRUs perform well in free

space, but perform worse during contact. LSTMs achieve the best free-space and contact performance, where longer-term actuator, frictional, and controller-dependent effects may play a larger role.

We further ablate input modalities and history length. Specifically, we compare histories of joint positions  $q$ , joint positions and velocities  $(q, \dot{q})$ , and joint positions, velocities, and tracking error  $(q, \dot{q}, \Delta q_d)$ , with history lengths of 10, 25, and 50 steps. Using  $(q, \dot{q}, \Delta q_d)$  consistently outperforms the other inputs, suggesting that tracking error provides useful information about controller effort and actuator response. Short histories of 10 steps generally yield higher error, while increasing the history to 50 steps improves performance, especially for LSTM models. Based on these results, we use an LSTM with history length  $H = 50$  and input  $(q, \dot{q}, \Delta q_d)$  as the final NEXT free-space inverse dynamics model.

### C. Free-Space Data Collection Guidelines for NEXT

Figure 8 shows the distribution of the 10-minute free-motion dataset collected on the Piper arms for training the free-space inverse dynamics LSTM for NEXT. During data collection, we follow a simple coverage-oriented procedure: first, each joint is moved independently across its range of motion; next, the end effector is moved through Cartesian-like motions involving multiple joints; finally, both slow and fast motions are repeated to capture velocity-dependent dynamics. Although these trajectories can be generated automatically through motion planning, in practice they can also be recorded once and replayed on new robot setups to quickly collect training data.

### D. External Joint Torque Estimation Baseline Implementation

We implement two model-based external joint torque estimation baselines for NEXT: FILIC [13] and a disturbance observer [30]. We evaluate both methods on the Franka Panda and AgileX Piper arms. For the Franka, we use the dynamics model from [12], whose parameters were obtained through optimization-based system identification. For the AgileX Piper, we use the dynamics model derived from the manufacturer-provided URDF [2].

**FILIC [13].** For FILIC, we use the standard residual-based estimate introduced in Sec. III. Given the measured motor torque  $\tau_m$ , external torque is estimated by subtracting the nominal free-space torque computed using MuJoCo inverse dynamics:

$$\hat{\tau}_{\text{ext}}^{\text{FILIC}} = \tau_m - \tau_f, \quad (9)$$

$$\tau_f = \mathbf{M}(\mathbf{q})\ddot{\mathbf{q}} + \mathbf{C}(\mathbf{q}, \dot{\mathbf{q}})\dot{\mathbf{q}} + \mathbf{g}(\mathbf{q}). \quad (10)$$

This method directly attributes the residual between measured motor torque and modeled free-space dynamics to external interaction.

**Disturbance Observer [30].** As a second baseline, we implement a momentum-based disturbance observer. Instead of directly computing acceleration-dependent inverse dynamics,

TABLE IV

ABLATION OF MODEL FAMILY, INPUT MODALITY, AND TEMPORAL HISTORY LENGTH FOR EXTERNAL TORQUE ESTIMATION. VALUES ARE THE AVERAGE JOINT  $L_1$  ERROR BETWEEN PREDICTED AND GROUND-TRUTH FREE-SPACE TORQUE ON THE HELD-OUT SET, REPORTED IN NM.

| Input                    | H  | MLP               |                   | GRU               |                   | LSTM (Ours)                         |                                     |
|--------------------------|----|-------------------|-------------------|-------------------|-------------------|-------------------------------------|-------------------------------------|
|                          |    | Free Motion<br>↓  | Contact<br>↓      | Free Motion<br>↓  | Contact<br>↓      | Free Motion<br>↓                    | Contact<br>↓                        |
| $q$                      | 10 | $2.35 \pm 2.34$   | $1.78 \pm 1.64$   | $2.25 \pm 1.79$   | $1.12 \pm 1.09$   | $2.53 \pm 2.52$                     | $2.61 \pm 2.29$                     |
|                          | 25 | $2.02 \pm 2.01$   | $1.34 \pm 1.32$   | $1.68 \pm 1.30$   | $0.665 \pm 0.478$ | $2.21 \pm 2.20$                     | $1.45 \pm 1.39$                     |
|                          | 50 | $1.76 \pm 1.76$   | $1.13 \pm 1.06$   | $1.62 \pm 1.24$   | $0.843 \pm 0.835$ | $2.15 \pm 2.15$                     | $1.09 \pm 0.908$                    |
| $q, \dot{q}$             | 10 | $1.04 \pm 1.03$   | $0.681 \pm 0.647$ | $0.863 \pm 0.862$ | $0.580 \pm 0.535$ | $1.15 \pm 1.15$                     | $0.715 \pm 0.553$                   |
|                          | 25 | $0.980 \pm 0.976$ | $0.908 \pm 0.802$ | $0.814 \pm 0.814$ | $0.596 \pm 0.506$ | $1.11 \pm 1.11$                     | $0.551 \pm 0.511$                   |
|                          | 50 | $1.02 \pm 1.02$   | $0.905 \pm 0.892$ | $0.795 \pm 0.794$ | $0.653 \pm 0.549$ | $1.12 \pm 1.12$                     | $0.578 \pm 0.526$                   |
| $q, \dot{q}, \Delta q_d$ | 10 | $0.706 \pm 0.702$ | $1.89 \pm 1.71$   | $0.423 \pm 0.422$ | $1.94 \pm 1.75$   | $0.504 \pm 0.503$                   | $0.876 \pm 0.244$                   |
|                          | 25 | $0.755 \pm 0.747$ | $1.56 \pm 1.45$   | $0.473 \pm 0.472$ | $1.47 \pm 1.29$   | $0.490 \pm 0.489$                   | $0.789 \pm 0.453$                   |
|                          | 50 | $0.813 \pm 0.811$ | $1.47 \pm 1.33$   | $0.429 \pm 0.429$ | $1.62 \pm 1.45$   | <b><math>0.414 \pm 0.278</math></b> | <b><math>0.547 \pm 0.348</math></b> |

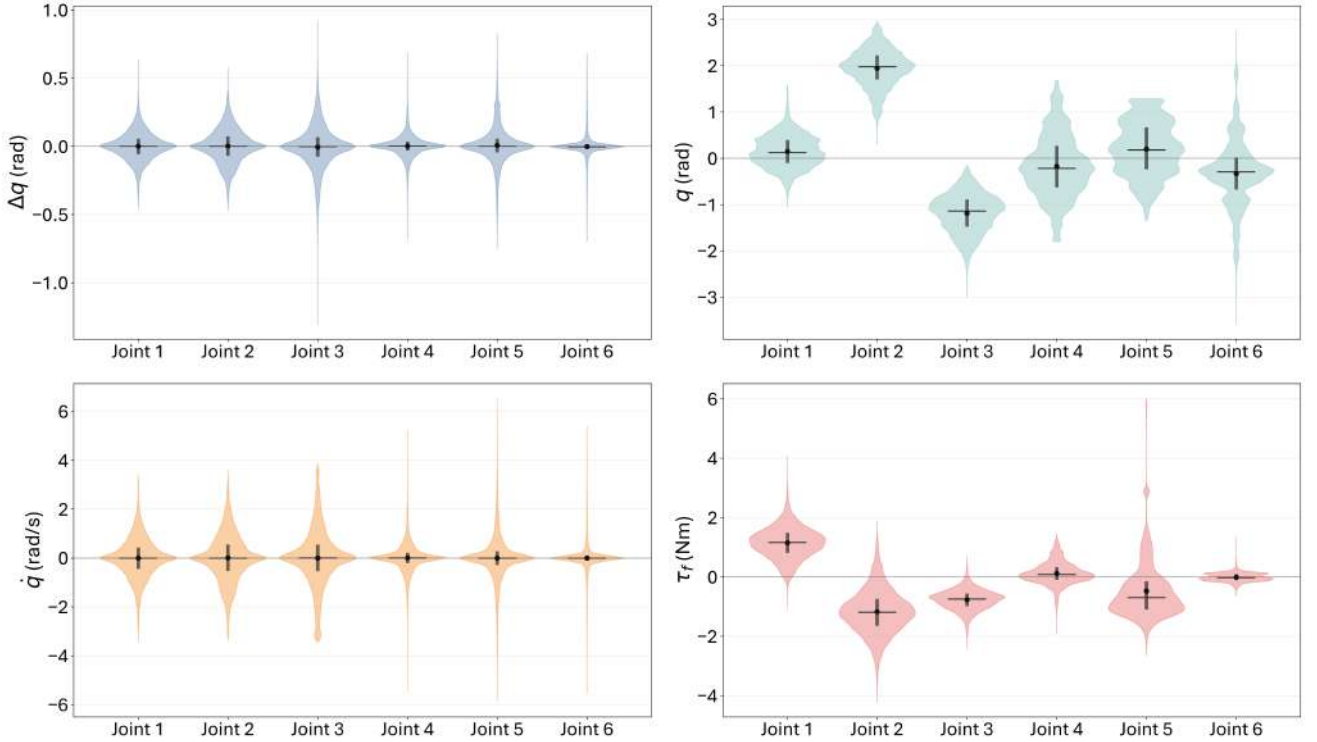


Fig. 8. Distribution of the 10-minute free-motion dataset used to train NEXT on the Piper arm. The plots show coverage of joint-position error  $\Delta q$ , joint position  $q$ , joint velocity  $\dot{q}$ , and free-space torque  $\tau_f$  across all six joints.

the observer estimates external torque from the residual in generalized momentum,

$$\mathbf{p} = \mathbf{M}(\mathbf{q})\dot{\mathbf{q}}. \quad (11)$$

It maintains an internal momentum estimate  $\hat{\mathbf{p}}$  with dynamics

$$\begin{aligned} \dot{\hat{\mathbf{p}}} &= \tau_m - \mathbf{C}(\mathbf{q}, \dot{\mathbf{q}})\dot{\mathbf{q}} - \mathbf{g}(\mathbf{q}) \\ &\quad + \dot{\mathbf{M}}(\mathbf{q})\dot{\mathbf{q}} + \hat{\tau}_{\text{ext}}^{\text{MDO}}, \end{aligned} \quad (12)$$

$$\hat{\tau}_{\text{ext}}^{\text{MDO}} = \mathbf{K}_o(\mathbf{p} - \hat{\mathbf{p}}). \quad (13)$$

By avoiding direct computation of  $\ddot{\mathbf{q}}$ , the momentum observer is less sensitive to encoder noise and produces smoother estimates than FILIC. As shown in Figure 5, it is less noisy than FILIC but remains less accurate than NEXT.

#### E. Force-Feedback Teleoperation Baseline Implementation

For the force-feedback teleoperation user study, we implement five feedback conditions on the Franka Panda:

- *No Feedback*: A GELLO leader-follower system [42] without force feedback.
- *Disturbance Observer (DO)*: External joint torque is estimated on the follower arm using the disturbance observer described in Appendix D, then relayed to the leader arm as force feedback.
- *Position-Position Feedback (PP)*: Feedback torque is computed by scaling the joint position difference between the leader and follower arms, following [18].
- *FACTR Teleop*: The FACTR teleoperation system [26],

TABLE V  
AVERAGE  $L_1$  JOINT TORQUE ERROR ON THE AGILEX PIPER IN  
FREE-SPACE SETTINGS. ERRORS ARE MEASURED AGAINST ZERO  
EXTERNAL TORQUE. NEXT ACHIEVES THE LOWEST ERROR.

| External Torque Method | Error (Nm) ↓                        |
|------------------------|-------------------------------------|
| Free Motion (GT)       | $0.000 \pm 0.000$                   |
| FILIC                  | $0.435 \pm 0.225$                   |
| Disturbance Observer   | $0.248 \pm 0.132$                   |
| <b>NEXT (Ours)</b>     | <b><math>0.018 \pm 0.012</math></b> |

Fig. 9. User-study results comparing external torque feedback methods on the AgileX Piper.

which feeds back external joint torque from dedicated joint torque sensors, along with gravity compensation, friction compensation, and null-space regulation.

- *Ours*: The same teleoperation framework as FACTR Teleop, but using NEXT-predicted external joint torque instead of dedicated joint torque sensor measurements.

We repeat the same study on the AgileX Piper arm, excluding FACTR Teleop because the Piper does not provide dedicated joint torque sensing.

#### F. AgileX Piper NEXT Evaluations

Since the AgileX Piper does not provide dedicated external joint torque sensing, we evaluate torque estimation only in the free-space setting, where the ground-truth external torque is known to be zero. As shown in Table V, NEXT achieves the lowest error, with an average  $L_1$  error of only 0.018 Nm, outperforming both FILIC and the disturbance observer. These results show that NEXT can produce accurate external torque estimates even on a low-cost arm without dedicated force sensors.

To evaluate NEXT in contact settings, we perform a force-feedback teleoperation user study. As shown in Figure 9, participants rated teleoperation with NEXT-based force feedback as easier to use than the baseline feedback methods.

Our behavior cloning policy takes RGB images, proprioceptive inputs, and force estimates as input. Each modality is tokenized before being passed to either an ACT or Flow Matching DiT policy architecture. Images are encoded with a DINOv3-Base encoder [40], while proprioceptive and force inputs are projected into tokens using MLP layers.

Proprioceptive inputs include joint positions and gripper widths, and force inputs correspond to estimated external joint torques. We use only the most recent observation at the current timestep. The action space consists of absolute joint angles and gripper width, with an action prediction horizon of 30. Training details for both ACT and Flow Matching policies are provided in Table VI.

#### G. Task Description

We evaluate FIRST along with baselines on five long-horizon, contact-rich manipulation tasks:

- *LEGO assembly (LEAP Hand V2 [38])*: bimanual hand assembly of two small LEGO blocks onto a bigger LEGO block.
- *NIST belt assembly (gripper)*: bimanual gripper manipulation of two belts around two pulley sets.
- *NIST insertion (gripper)*: insertion of two pulley sets onto pegs on a NIST board.
- *Tool Clean Up (gripper)*: screwdriver insertion followed by case assembly.
- *Cap screwing (gripper)*: cap pickup, alignment, and screwing onto a container until tightened.

#### H. Task Progress Rate Calculation

We report task progress rate in addition to binary success rate to evaluate partial completion in multi-stage manipulation tasks. Each task is divided into several stages, and each stage is marked as either completed (s) or failed (f) for every rollout. The task progress rate is calculated as the percentage of completed stages across all rollouts.

The task stages are listed in Table VII. For example, the LEGO task consists of grasping, aligning, and inserting both the red and green cubes. The NIST Belt task includes picking up and aligning both belt components before pushing them into place. The NIST Insertion task measures progress through picking, aligning, and inserting the black part, followed by picking up the aluminum part. The Tool Clean Up task evaluates whether the robot can pick up, align, and insert the driver, then assemble the box.

For the cap screwing task, we use a modified progress score because a policy can fail by continuing to rotate the cap after it is already tightened. Therefore, we subtract a penalty based on the number of over-turns. Each over-turn reduces the progress score by 0.02. This penalizes policies that complete the screwing motion but do not stop properly.

#### I. Imitation Learning with Force Information Baseline Implementations

**Base Policy**: We use the base ACT or flow-matching policy described in Sec. VI-B. The policy takes multi-view RGB images and robot proprioception, including joint positions and gripper states, as input and predicts a chunk of future joint and gripper commands.

**Base Policy + Torque**: This baseline augments the base policy with external joint torque predictions from NEXT. The predicted torque signal is embedded with an MLP and added to the policy context as an additional observation token.

**FACTR [26]**: FACTR keeps the policy architecture unchanged and applies a visual curriculum to all image observations. Following the original FACTR implementation, we use a Gaussian blur operator with a linear curriculum schedule. Specifically, at training step  $n$ , the blur strength is set as

$$\sigma_n = \sigma_0 \left(1 - \frac{n}{N}\right),$$

where  $\sigma_0$  is the initial blur strength and  $N$  is the total number of curriculum steps. Thus, the input images are strongly

TABLE VI  
POLICY ARCHITECTURE AND TRAINING HYPERPARAMETERS.

| Hyperparameter                             | Value                             |
|--------------------------------------------|-----------------------------------|
| <b>Behavior Policy Training</b>            |                                   |
| Optimizer                                  | AdamW                             |
| Base learning rate                         | $3 \times 10^{-4}$                |
| Weight decay                               | 0.05                              |
| Optimizer momentum                         | $\beta_1, \beta_2 = 0.9, 0.95$    |
| Batch size                                 | 128                               |
| Learning-rate schedule                     | Cosine decay                      |
| Total epochs                               | 15–20                             |
| Warmup steps                               | 500                               |
| Augmentation                               | RandomResizeCrop                  |
| GPU                                        | NVIDIA A6000, 48 GB               |
| Wall-clock time                            | 12–24 hours                       |
| <b>Action Chunking Transformer Details</b> |                                   |
| Encoder layers                             | 4                                 |
| Decoder layers                             | 6                                 |
| MHSA heads                                 | 8                                 |
| Hidden dimension                           | 768                               |
| Feed-forward dimension                     | 3072                              |
| Dropout                                    | 0.1                               |
| Positional encoding                        | Sinusoidal                        |
| Action chunk                               | 30                                |
| <b>Flow-Matching Details</b>               |                                   |
| Decoder layers                             | 6                                 |
| MHSA heads                                 | 8                                 |
| Hidden dimension                           | 768                               |
| Feed-forward dimension                     | 3072                              |
| Dropout                                    | 0.1                               |
| Positional encoding                        | Sinusoidal                        |
| Action chunk                               | 30                                |
| Evaluation steps                           | 10                                |
| Time sampling                              | $\tau \sim \text{Beta}(1.0, 1.5)$ |
| ODE solver                                 | Euler                             |
| Noise distribution                         | $\mathcal{N}(0, I)$               |

TABLE VII  
STAGE DECOMPOSITION USED FOR TASK-PROGRESS CALCULATION.

| Task           | Stages                                                                   |
|----------------|--------------------------------------------------------------------------|
| LEGO           | Red grasp, red align, red insert, green grasp, green align, green insert |
| NIST Belt      | Pick orange, align big, align small, push, pick gray, align gray, push   |
| NIST Insertion | Pick black, align black, insert black, pick aluminum                     |
| Tool Clean Up  | Pick screwdriver, align screwdriver, insert screwdriver, assemble box    |
| Cap Screwing   | Pick cap, align cap, screw on, end properly                              |

blurred at the beginning of training, and the blur is gradually removed as training progresses.

**TA-VLA [49]:** TA-VLA uses the predicted torque/contact signal as an auxiliary learning target in addition to the action. During inference, only the robot-action dimensions are executed, while the auxiliary torque/contact prediction is used only to shape the learned representation during training.

### J. Segmenting Trajectories into Contact Phases Visualization

Figure 10 shows an example of phase labels produced by our hysteresis-threshold-based labeling procedure. The estimated external torque from NEXT remains near zero during free motion and rises clearly during physical interaction, allowing the trajectory to be separated into free-motion, pre-contact, and contact phases. The pre-contact label captures the short interval immediately before contact onset, where precise alignment is required, while the contact label captures periods of sustained interaction. This sensor-level signal enables automatic segmentation of demonstrations even on robot arms without dedicated force sensors, providing the phase annotations used by FIRST for force-informed re-sampling.

### K. Task Progress Evaluation on ACT

We also apply FIRST to ACT, in addition to the flow-matching policy presented in Section VII-B. As shown in Figure 11, we observe a similar trend: up-sampling pre-contact and contact phases improves task progress over the baselines.

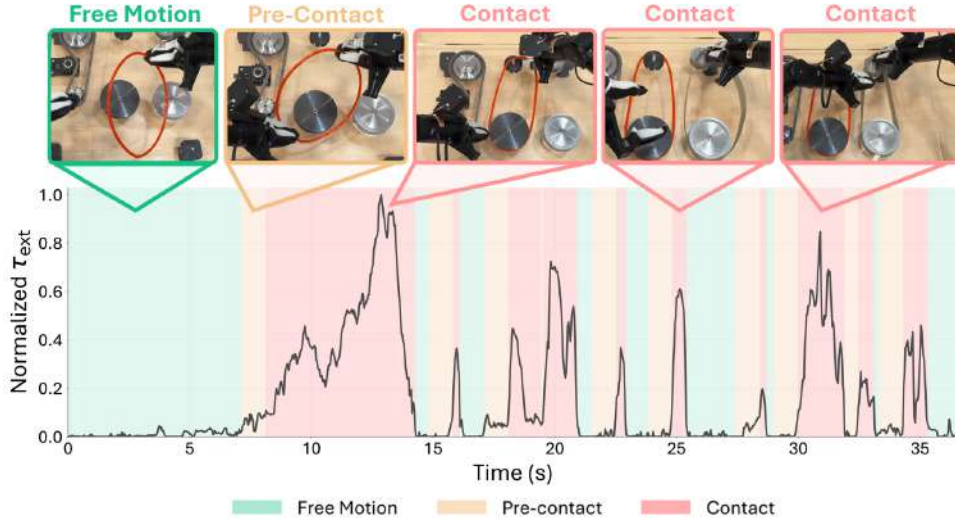


Fig. 10. **Contact-phase segmentation from learned external torque.** The normalized  $\tau_{\text{ext}}$  from NEXT is used to segment demonstrations into free-motion, pre-contact, and contact phases. The image overlays show that the predicted labels align with the robot’s interaction state during the task.

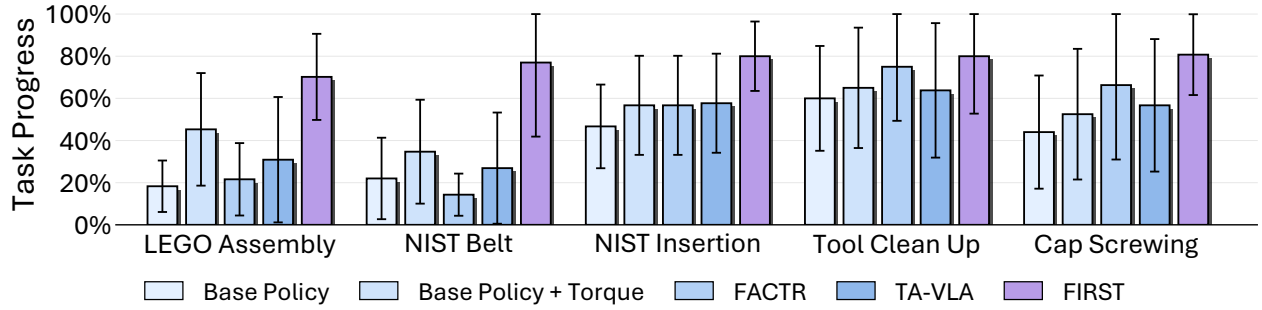


Fig. 11. Task progress across five contact-rich manipulation tasks using the ACT policy.

TABLE VIII

BEST RE-SAMPLING RATIOS USED BY FIRST FOR EACH TASK. RATIOS ARE REPORTED AS  $w_F : w_{PC} : w_C$ , CORRESPONDING TO FREE-MOTION, PRE-CONTACT, AND CONTACT SAMPLING WEIGHTS.

| Task               | Best Ratio |
|--------------------|------------|
| LEGO Assembly      | 1 : 5 : 1  |
| NIST Belt Assembly | 1 : 5 : 1  |
| NIST Insertion     | 1 : 5 : 1  |
| Tool Clean Up      | 1 : 5 : 1  |
| Cap Screwing       | 1 : 3 : 3  |

#### L. Selected Re-Sampling Ratios for Each Task

We report the task-specific re-sampling ratios used by FIRST in Table VIII. All ratios are written as  $w_F : w_{PC} : w_C$ , where  $w_F$ ,  $w_{PC}$ , and  $w_C$  denote the sampling weights for free-motion, pre-contact, and contact samples, respectively. For most tasks, FIRST uses 1 : 5 : 1, which up-samples pre-contact samples by 5 $\times$  relative to free-motion and contact samples. For Cap Screwing, FIRST uses 1 : 3 : 3, up-sampling both pre-contact and contact samples, since sustained contact is central to successful execution. These ratios are selected based on task progress for each task.

#### M. Up-Sampling Magnitude Ablation

Figure 12 ablates the effect of up-sampling magnitude on task progress. We define the magnitude  $m$  as the sampling weight assigned to the up-sampled phase relative to free motion. For pre-contact up-sampling, the phase weights are

$$w_F : w_{PC} : w_C = 1 : m : 1,$$

whereas for pre-contact + contact up-sampling, the phase weights are

$$w_F : w_{PC} : w_C = 1 : m : m.$$

We run this ablation on NIST Belt Assembly, NIST Insertion, and Tool Clean Up using ACT. Due to limited time and compute, we additionally run the flow-matching ablation only on Tool Clean Up. Across tasks, moderate up-sampling improves task progress, with performance generally peaking around  $m = 5$ . Larger values lead to saturation or degradation, suggesting that overly aggressive up-sampling can over-bias the training distribution toward contact-relevant segments and reduce coverage of the full task trajectory.

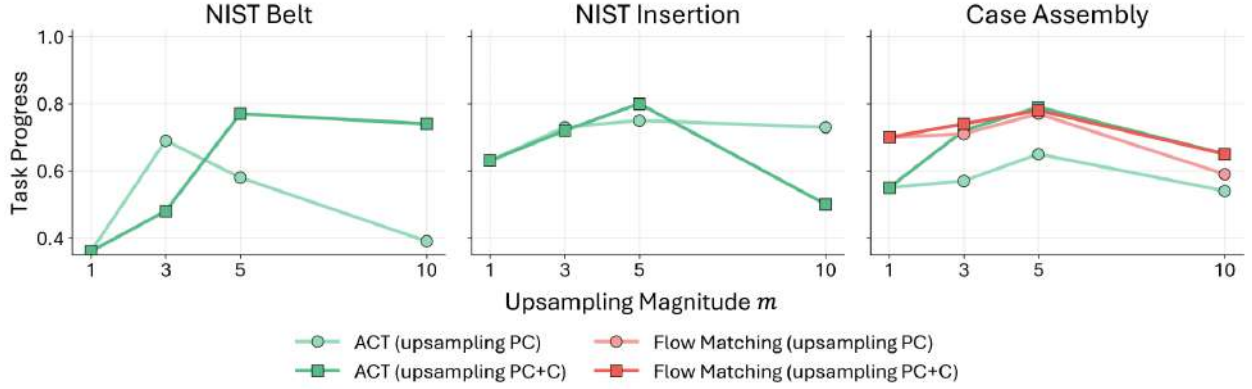


Fig. 12. Moderate up-sampling generally improves performance, with the best results around  $m = 5$ , while larger values lead to saturation or degradation.

#### N. Up-Sampling Policies Without Force Input

While Section VII-B evaluates FIRST on policies that condition on external joint torque, FIRST can also be applied without using force as a policy input. In this setting, NEXT is used only offline to segment demonstrations into free-motion, pre-contact, and contact phases. These phase labels are then used by FIRST to re-sample the training data, increasing the frequency of pre-contact and contact examples in each minibatch. Importantly, the policy itself receives only the standard visual and proprioceptive observations, without the predicted external joint torque.

Table IX highlights two trends. First, when comparing methods within the columns without external joint torque, phase-based up-sampling improves over standard behavior cloning. For both ACT and Flow Matching, using PC or PC+C up-sampling generally increases task success compared to no up-sampling, showing that NEXT-derived phase labels are useful even when force estimates are used only at training time for data re-sampling. This indicates that FIRST can improve standard behavior cloning by reshaping the training distribution toward task-critical segments where alignment and contact-sensitive corrections are most important.

Second, when comparing horizontally between policies trained with and without external joint torque, the force-conditioned policies consistently achieve higher success rates. This shows that while phase-based up-sampling alone provides a clear benefit, directly feeding the predicted external joint torque into the policy provides additional state information that further improves contact-rich manipulation. Overall, the best performance is achieved when both components are used together: NEXT provides external joint torque as a policy input, while FIRST uses NEXT-derived phase labels to up-sample the most relevant pre-contact and contact phases during training.

#### O. Inputting Joint Current vs. External Joint Torque on Policy Performance

We further examine whether raw joint current is sufficient for contact-aware policy learning, or whether the learned exter-

nal joint torque estimate provides a more useful representation. Prior work such as TA-VLA [49] uses readily available joint current measurements from low-cost robot arms as a force-related input for imitation learning. However, joint current is only an indirect measure of physical interaction: it contains actuator effort required for both free-space motion and external contact, as well as hardware- and controller-dependent effects. As a result, the policy must implicitly disentangle contact-induced torque from free-space dynamics.

In contrast, NEXT explicitly estimates external joint torque by subtracting a learned free-space torque prediction from the measured motor torque. This distinction is important because estimating a clean external torque signal from current alone generally requires temporal context. As shown in our NEXT free-space inverse dynamics ablations in Appendix B, longer proprioceptive histories improve torque estimation, suggesting that instantaneous current and state observations are insufficient to reliably separate free-space actuation from contact. A standard behavior cloning policy, however, typically receives only the current observation rather than a long proprioceptive history such as the 50-step history used by NEXT. While one could provide such a history directly to the policy, this increases the burden on the policy to simultaneously infer dynamics, localize contact, and solve the visuomotor task. In practice, long proprioceptive histories may also encourage overfitting to trajectory-specific motion patterns rather than using vision appropriately to react to scene variation. NEXT avoids this issue by using history in a specialized force-estimation module, while providing the policy with a compact, contact-relevant external torque signal.

We compare policies conditioned on raw joint current and learned external joint torque on two contact-sensitive tasks: NIST Insertion and Cap Screwing. Both tasks require force cues for successful execution: NIST Insertion depends on detecting alignment and contact during insertion, while Cap Screwing requires identifying when the cap has been fully tightened. As shown in Table X, conditioning on joint current improves performance on NIST Insertion compared to the base ACT policy, but does not improve Cap Screwing. In

TABLE IX  
EFFECT OF PHASE-BASED UP-SAMPLING ON POLICIES TRAINED WITH AND WITHOUT EXTERNAL JOINT TORQUE. RESULTS ARE REPORTED AS TASK SUCCESS RATES.

| Policy        | Up-sampling Strategy | Without $\tau_{\text{ext}}$ |               | With $\tau_{\text{ext}}$ |               |
|---------------|----------------------|-----------------------------|---------------|--------------------------|---------------|
|               |                      | Belt Assembly               | Tool Clean Up | Belt Assembly            | Tool Clean Up |
| ACT           | None                 | 0.220                       | 0.600         | 0.347                    | 0.650         |
|               | PC                   | 0.360                       | 0.600         | 0.687                    | 0.650         |
|               | PC+C                 | 0.283                       | 0.763         | 0.770                    | 0.788         |
| Flow Matching | None                 | 0.150                       | 0.463         | 0.431                    | 0.525         |
|               | PC                   | 0.213                       | 0.575         | 0.767                    | 0.805         |
|               | PC+C                 | 0.241                       | 0.525         | 0.649                    | 0.775         |

TABLE X  
POLICY PERFORMANCE WHEN CONDITIONING ON RAW JOINT CURRENT VERSUS LEARNED EXTERNAL JOINT TORQUE. HIGHER IS BETTER, AND BOLD INDICATES THE BEST RESULT FOR EACH TASK.

| Method                                       | NIST Insertion $\uparrow$ | Cap Screwing $\uparrow$ |
|----------------------------------------------|---------------------------|-------------------------|
| ACT                                          | 0.220                     | 0.440                   |
| ACT w/ current                               | 0.431                     | 0.440                   |
| <b>ACT w/ <math>\tau_{\text{ext}}</math></b> | <b>0.567</b>              | <b>0.525</b>            |

contrast, conditioning on learned external joint torque achieves the best performance on both tasks. Qualitatively, we observe that policies conditioned on raw joint current often continue rotating the cap after it is already tightened, leading to over-screwing failures. Policies conditioned on learned external torque are better able to detect the tightened contact state and terminate the screwing motion appropriately.

The attention analysis in Figure 13 further supports this observation. To visualize how the policy uses force-related information during rollout, we compute the average cross-attention from the action tokens to the observation tokens in the first decoder layer. The decoder attends over both visual tokens and force-related tokens, but in Figure 13 we plot only the attention mass assigned to the force-related tokens: raw joint-current tokens for the current-conditioned policy, and learned external-torque tokens for our policy. This visualization therefore measures how strongly the policy attends to the force input when predicting actions.

During cap-screwing manipulation, the policy conditioned on learned external torque shows a clear increase in attention around key contact events, especially when the cap becomes tightened and the screwing motion should stop. In contrast, the policy conditioned on raw joint current exhibits weaker and less event-aligned attention to its current tokens. While attention is only a diagnostic and does not by itself prove causality, this pattern suggests that learned external joint torque provides a cleaner and more task-relevant contact representation than raw joint current for force-sensitive manipulation.

#### P. Per-Batch Up-Sampling vs. Weighted Regression

We further compare FIRST against a weighted-regression baseline, where the loss is reweighted according to whether

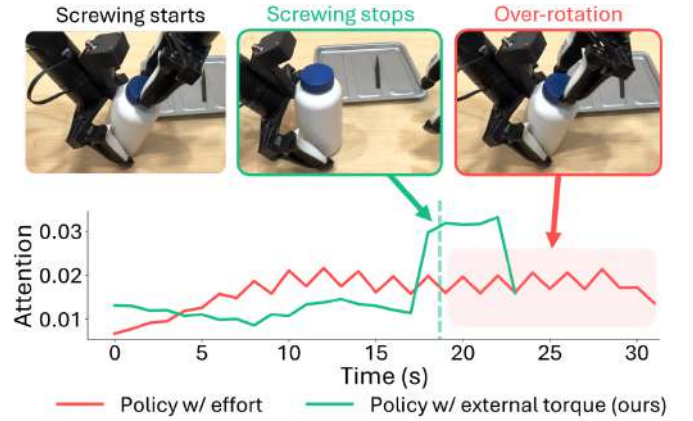


Fig. 13. Attention values during cap-screwing manipulation. Learned external torque produces clearer attention around key contact events than raw joint current.

each sample belongs to the free-motion, pre-contact, or contact phase. Both approaches aim to emphasize contact-relevant data, but they do so in different ways: weighted regression modifies the optimization objective, whereas FIRST changes the training distribution by increasing the frequency of informative samples within each batch.

TABLE XI  
COMPARISON OF PHASE-EMPHASIS MECHANISMS FOR LEGO ASSEMBLY USING ACT.

| Phase-Emphasis Mechanism | Phase | Progress |
|--------------------------|-------|----------|
| None                     | –     | 0.453    |
| Weighted Regression      | PC    | 0.345    |
|                          | PC+C  | 0.478    |
| <b>FIRST (Ours)</b>      | PC    | 0.573    |
|                          | PC+C  | 0.501    |

As shown in Table XI, weighted regression either degrades performance or provides only limited improvement over the no up-sampling baseline. In contrast, weighted sampling consistently improves LEGO assembly performance. We observe similar qualitative trends in other contact-rich tasks. This suggests that directly increasing the loss contribution of sparse contact-related samples can make optimization less stable,

potentially causing the policy to overfit to high-weighted segments or ignore the broader trajectory context. By contrast, per-batch up-sampling preserves the original regression objective while exposing the policy more frequently to critical pre-contact and contact transitions, leading to more stable policy learning.