

9 Supplementary Material

9.1 Filters

The measured angle is filtered with a Kalman filter to smooth the results. However, it was found that the Kalman filter introduces inconsistent delay to the angular velocity, hence the angular velocity is instead directly computed from the Kalman filtered angle. This angular velocity is then smoothed using a non-causal filter, with a triangular window of size 9. This filter was found to produce the smoothest results, without adding any delay.

Following the conventions from [4], the kinematics model used was $x_{t+1} = Ax_t + w_t$ and the measurement model used was $z_t = Cx_t + v_t$. Here the process noise w_t and the measurement noise v_t are assumed to be 0 mean with covariance matrices Q and R respectively. The Kalman filter aims to estimate $x_t = [\alpha, \omega]^\top$ using the measurements z_t . The initial state x_0 is taken to be $[z_0, 0]^\top$ with covariance Σ_0 . The parameters used are listed in Table 3.

Parameter	Value
A	$\begin{bmatrix} 1 & \frac{1}{60} \\ 0 & 1 \end{bmatrix}$
C	$\begin{bmatrix} 1 & 0 \end{bmatrix}$
Q	$\begin{bmatrix} 3.25 \times 10^{-6} & 6.5 \times 10^{-5} \\ 6.5 \times 10^{-5} & 1.3 \times 10^{-3} \end{bmatrix}$
R	1×10^{-5}
Σ_0	$\begin{bmatrix} 1 \times 10^{-5} & 0 \\ 0 & 1 \times 10^{-5} \end{bmatrix}$

Table 3: Kalman filter parameters

9.2 Data Collection Methodology

Each object starts at rest on a horizontal plane. The gripper approaches the object at a global yaw angle ($\alpha_{approach}$) chosen from a set of angles $A_{approach}$. The object is then grasped by closing the gripper to a fixed width for each object. This width manually is chosen such that the objects remain stationary inside the gripper while being lifted. The object is then lifted to a fixed height above the table, where an end effector yaw angle perturbation is chosen from the set of angles $A_{perturb}$. The recording of both tactile measurements and the images is started after the object has been lifted. The object is held stationary for 0.5s. The gripper width is then loosened to start the gravitational pivoting. One of the two methodologies for gripper control will then be used. Once the object has finished rotating, another fixed delay of 1s occurs, where the object is held stationary in its steady-state position. The data recording is then finished. For ‘Angle Goal’, each parameter combination was run twice to increase the amount of data collected.

9.3 Parameters

9.3.1 Data Collection

Rotate to stop:

- $A_{approach} = \{-30^\circ, -15^\circ, 0^\circ, 15^\circ, 30^\circ\}$
- $A_{perturb} = \{-45^\circ, 0^\circ, 15^\circ, 30^\circ, 45^\circ, 60^\circ\}$

Angle Goal

- $A_{approach} = \{-15^\circ, 0^\circ\}$
- $A_{perturb} = \{0^\circ, 30^\circ, 45^\circ, 60^\circ\}$
- $A_{stop} = \{15^\circ, 30^\circ, 45^\circ\}$.

9.3.2 Experiments

- $A_{approach} = \{-30^\circ, 0^\circ, 30^\circ\}$
- $A_{perturb} = \{0^\circ\}$
- $A_{stop} = \{30^\circ, 45^\circ, 60^\circ\}$

9.3.3 Controller

The controller was hand-tuned to maximise the performance of the system with the **Oracle** baseline. The other controllers were directly transferred using the same parameters, listed in Table 4.

Parameter	Value
ϵ_α	1°
ω_{min}	20°/s
t_{wait}	0.75s
d	0.83s

Table 4: Controller Parameters

9.4 Unseen Class Experiment

We investigate the ability of the model to generalize between different classes of objects. To do this, we train and test on different classes of objects (boxes vs cylinders). The results of this training are displayed in Table 5. When training and testing on a single class, the performance of the networks is similar to training on both classes. This suggests that the network can generalize between objects of the same class. However, the performance of both the LSTM and MLP are decreased when testing on an unseen class. This is likely due to different shapes causing a large change in the resultant tactile information.

Methods	Train	Test	Angular Error (°)			Velocity Error (°/s)		
			IS	DR	SS	IS	DR	SS
MLP	Box	Box	1.94	10.53	20.42	5.24	36.36	5.78
LSTM	Box	Box	0.32	4.02	8.35	1.54	26.27	0.95
MLP	Box	Cyl	3.98	15.42	35.06	5.13	46.61	8.24
LSTM	Box	Cyl	3.38	11.99	18.83	2.949	44.46	0.72
MLP	Cyl	Cyl	1.27	7.69	14.86	8.86	41.31	7.93
LSTM	Cyl	Cyl	0.34	3.29	5.23	1.86	34.71	3.12
MLP	Cyl	Box	18.65	27.8	54.55	3.69	63.2	8.27
LSTM	Cyl	Box	10.17	24.43	26.59	4.35	73.17	3.08

Table 5: Results on training on an unseen class. IS, DR, SS represent the three sections of motion, Initial State, Dynamic Range and Steady State, respectively.

9.5 Network Details and Training Details

All models are trained using the ADAM optimizer, using the sum of both L1 and L2 functions for both ω and α as the loss function. We normalized the target values of ω and α to be close to the range [0,1] so each component of the overall loss was given equal weighting. This was found to produce the smoothest result. While training the model, data is batched by cropping sequences to be the same length as the shortest sequence in the batch. The hyperparameters, including number of layers, of both models were found using a sweep and are reported in Table 6. The best model determined by the sweep was used for all experiments. All models are trained for 100 epochs. We report the accuracy of the model in terms of the mean absolute error (MAE).

9.6 Unseen Objects

The testing results on unseen objects are displayed in Table 7. The LSTM outperformed the MLP for all metrics, however the results have degraded compared to seen objects. In particular, the MLP failed to maintain a constant value close to 0 in the angle IS, angular velocity IS and angular velocity SS sections. The LSTM still maintains predictions close to 0 for these sections, displaying the benefit of the stability of the LSTM compared to the MLP, which is much noisier with its predictions.

To improve the generalization to unseen objects, we believe that a larger and more diverse object set will be required. This should consist of objects of a wide range of shapes, weights and surface

LSTM Parameter	Value	MLP Parameter	Value
LSTM Input Size	142	Input Size	2130
LSTM Hidden Size	500	Number Layers	4
LSTM Dropout	0.15	Dropout	0.15
LSTM Number Layers	3	Activation	Tanh
MLP Layers	2	Learning Rate	5×10^{-4}
MLP Hidden Size	500	Weight Decay	1×10^{-6}
Learning Rate	5×10^{-4}	Window Size	15
Weight Decay	1×10^{-6}		

Table 6: Hyperparameters used for the experiments

properties. We believe that this will allow the model to learn the effects of these properties on the forces received from the tactile sensor and so better account for them.

	Angular Error (°)			Angular Velocity Error (°/s)		
	IS	DR	SS	IS	DR	SS
LSTM	0.37±0.17	7.98±4.04	10.87±4.58	0.84±0.38	42.27±14.29	2.43±1.37
MLP	3.49±5.14	13.1±4.73	24.49±5.92	5.05±3.45	44.44±13.52	5.03±1.49

Table 7: Testing results on unseen objects. IS, DR, SS represent the three sections of motion, Initial State, Dynamic Range and Steady State, respectively.

9.7 Finetuning on real robot

The finetuning results on the training dataset are displayed in Table 8. Note that the finetune training is done following the unseen objects experiments. Although finetuning has increased error in some sections, the error within the critical ‘Dynamic Range’ section has decreased. This section is critical in detecting when a target angle has been reached and so this decreased error reflects the decreased target error of the finetuned models.

	Angular Error (°)			Angular Velocity Error (°/s)		
	IS	DR	SS	IS	DR	SS
LSTM	0.54±0.18	4.84±2.11	13.75±6.00	3.01±3.06	33.54±20.12	2.99±2.45
MLP	3.08±3.89	7.78±3.51	26.32±10.25	4.83±4.91	31.36±22.30	5.41±3.46

Table 8: Results of training finetuned models. IS, DR, SS represent the three sections of motion, Initial State, Dynamic Range and Steady State, respectively.

9.8 Finetuned Real-World Tracking Results

Table 9 shows the results for real-world tracking. Again these results were measured on unseen object models. While the error of real world tracking is similar to the training results, the variance is higher. This contributes to the tracking error being much higher than the average errors, with models failing at critical instances such as when the gripper width widens and the object begins to rotate faster. Such spikes in error will be smoothed out by the longer periods of slow rotations but are still present in the measurements of variance.

	Angular Error (°)			Angular Velocity Error (°/s)		
	IS	DR	SS	IS	DR	SS
LSTM	0.69±0.61	4.61±4.15	14.74±11.82	3.46±3.49	14.51±16.59	5.09±14.72
MLP	1.76±3.07	10.15±8.92	31.23±27.21	4.51±2.71	11.47±6.17	5.55±7.43

Table 9: Results of finetuned models with experimental setup. IS, DR, SS represent the three sections of motion, Initial State, Dynamic Range and Steady State, respectively.

9.9 Test Tracking Results

Figure 9 shows more examples of the LSTM performance on unseen data (as described in 6.3). The LSTM can accurately track the object, especially when the object follows a smooth trajectory. However, the LSTM also commonly suffers from steady state errors.

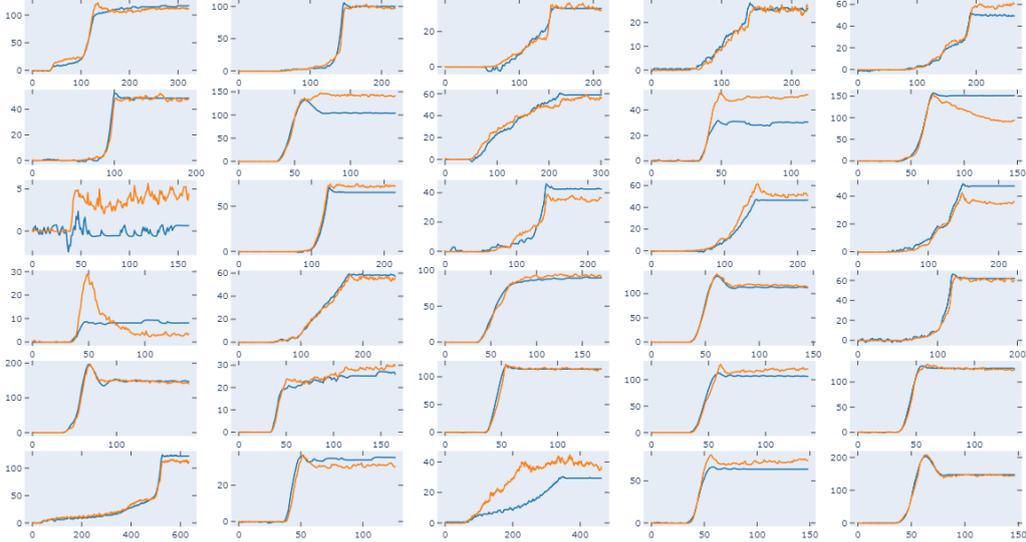


Figure 9: The LSTM tracking results on unseen data. The blue lines represent the ground truth sequences, and the orange lines represent the LSTM prediction. The y-axis represents the angular position. The scale of this axis varies between graphs.

9.10 MLP Window Size

Window Size	Angular Error ($^{\circ}$)			Angular Velocity Error ($^{\circ}/s$)		
	IS	DR	SS	IS	DR	SS
5	1.03	11.34	19.89	1.8	49.59	4.64
15	1.15	10.47	18.84	1.75	40.45	4.23
30	0.94	10.65	20.77	1.82	43.78	3.52
60	1.34	11.65	20.66	1.83	48.04	6.72
90	0.7	12.03	18.59	12.86	52.13	7.85

Table 10: Ablation study on the window size used for the MLP. IS, DR, SS represent the three sections of motion, Initial State, Dynamic Range and Steady State, respectively.

Table 10 shows the results for MLP models trained on an 80/20 random data split, comparing different window sizes. For Angular Error, we found that for a window size of 15, the error during Dynamic Range was lowest and had a comparable error to a window size of 90 during Steady State. We also found that a window size of 15 performed well in predicting velocity error, outperforming all other models during the Dynamic Range stage, which is the most important for our controller.

9.11 Other Recurrent Models with Unseen Objects

Table 11 shows the results when other recurrent model types are trained to generalize to unseen objects. The GRU had identical hyperparameters to the LSTM and the RNN had the same hyperparameters expect for a dropout of 0. We found that the RNN model outperforms the MLP baseline model consistently, but performed worse than both the LSTM and GRU models.

We also found that the LSTM and GRU models exhibit almost identical performance in all cases. The slightly lower angular error the GRU displays compared to the LSTM in the DR and SS regions comes entirely from the overperformance of the GRU on one object, Magnet. For all other

objects there was minimal discrepancy between the models. We believe these models could be used interchangeably for this task.

	Angular Error (°)			Angular Velocity Error (°/s)		
	IS	DR	SS	IS	DR	SS
LSTM	0.37±0.17	7.98±4.04	10.87±4.58	0.84±0.38	42.27±14.29	2.43±1.37
GRU	0.42±0.25	7.38±3.21	10.69±4.31	1.32±0.56	39.51±11.87	3.94±1.10
RNN	1.80±1.87	10.75±4.26	15.47±4.30	2.88±1.06	46.24±11.69	5.85±4.63

Table 11: Testing other recurrent models on unseen objects. IS, DR, SS represent the three sections of motion, Initial State, Dynamic Range and Steady State, respectively.

9.12 Object Details

	Size [mm]	Mass [g]	Class	3D-Printed?	Full?	Deformable
Toothpaste	167 × 58 × 12	52	Box		✓	✓
Earbud	134 × 51 × 29	27	Box		✓	
Breadboard	167 × 58 × 12	84	Box		✓	
Magnet	181 × 68 × 40	29	Box			
Deodorant	49 × 210	50	Cylinder			
Spray2	41 × 158	135	Cylinder		✓	✓
Shampoo	50 × 157	96	Cylinder	✓		
Spray1	37 × 142	46	Cylinder	✓		
Pill	55 × 116	26	Cylinder			
Toothbrush	217 × 29 × 21	33	Box		✓	

Table 12: Details of the object distributions used in our work. The dimensions of boxes are provided in the form $length \times width \times depth$, while cylindrical objects are dimensioned as $radius \times height$

Table 12 provides details on the properties of objects used in our work. All objects are required to be long, to allow for gravitational pivoting. The 3D-printed objects are to provide a different texture compared to the relatively smooth materials of the household objects. When objects that are partially full undergo rotation, their contents will move around such that there will be a large change in its centre of gravity. However, for this work, we only consider objects full such that the centre of gravity is relatively stable during rotation. Deformable objects include objects which can be squished such as the soft cardboard making up the toothpaste box. We tried to choose a range of objects to increase the ability for our system to generalize.