Supplementary material

This document contains the supplementary material related to *SofToss: Learning to Throw Objects with a soft robot*.

A. Tables

Table SI
Neural network training. Here are collected the different hyperparameters changed and fixed in the training of the direct model of the task and the actuation network in the two distinguished actuation scenarios.

Direct model of the task			
Default values			
learning rate	0.001	Optimiser	Adam
batch's size	16	Loss function	MSE
number of epochs	4000	Training/test partition	0.95
Best combination of the changed parameters			
Hyperparameter	**Partial Actuation**	**Complete actuation**	
Number of units	64	64	
Activation function	ReLU	ReLU	
Normalisation	Rescaling	Z-score	
Actuation network			
Default values			
learning rate	0.001	Optimiser	Adam
batch's size	32	Loss function	MSE
number of epochs	500	Training/test partition	0.9
Best combination of the changed parameters			
Hyperparameter	**Partial Actuation**	**Complete actuation**	
Number of units	128	128	
Activation function	ReLU	ReLU	

Table SII
Qualitative analysis of the video recorded related to Deep reinforcement learning controller in the complete actuation scenario. All the recordings of the throws are reported in the supplementary material (Data file). We have considered the following four states of the object: (E) enter; (BI) bounced-in; (BO) bounced-out; (O) out.

Deep reinforcement learning controller - complete actuation scenario												
Target	Tossed object											
	PingPong			Lemon			Marker			Tomato		
A	O	O	O	O	O	O	O	O	O	O	O	O
B	BO	BO	O	BO	BO	BO	E	E	E	E	E	E
C	O	O	O	O	E	O	BI	E	E	BI	BI	BI
D	E	E	E	E	E	E	O	O	O	BI	BO	BI
E	E	E	E	BI	BI	E	BI	BI	BI	BO	BO	BO
F	BO	BO	BI	BO	BI	BI	BO	BO	O	BO	BI	BO
G	E	E	E	E	E	E	E	E	E	E	E	E
H	BO	BO	BO	E	E	E	E	E	E	E	E	E
I	BO	BI	BI	E	E	E	O	O	E	E	E	E
L	E	E	E	E	E	E	E	BI	BO	BO	BI	BI
$\sum$ **E**	12			17			13			12		
$\sum$ **BI**	3			4			5			8		
$\sum$ **BO**	8			4			3			7		
$\sum$ **O**	7			5			9			3		

Table SIII

Qualitative analysis of the video recorded related to Deep reinforcement learning controller in the partial actuation scenario. All the recordings of the throws are reported in the supplementary material (Data file). We have considered the following four states of the object: (E) enter; (BI) bounced-in; (BO) bounced-out; (O) out.

Deep reinforcement learning controller - partial actuation scenario												
Target	Tossed object											
	PingPong			Lemon			Marker			Tomato		
A	E	E	E	E	E	E	E	E	O	BI	BI	E
B	E	BO	E	E	E	E	E	E	E	BI	BO	E
C	O	BO	O	E	E	E	BO	BI	BI	BI	BO	O
D	O	O	O	O	O	O	O	E	E	E	E	E
E	E	O	BO	E	E	E	BO	BO	BO	BI	E	BI
F	E	E	BO	E	BI	BI	O	O	O	BI	BO	BI
G	E	E	E	E	E	E	E	E	E	E	E	E
H	E	E	E	E	E	E	O	O	O	BI	BI	BI
I	E	E	E	BO	BI	BO	O	O	O	O	BO	BO
L	E	E	E	BO	BI	BO	O	O	O	BI	BO	BO
$\sum$ **E**		20			19			10			9	
$\sum$ **BI**		0			4			2			12	
$\sum$ **BO**		4			4			4			7	
$\sum$ **O**		6			3			14			2	

Table SIV

Qualitative analysis of the video recorded related to non-real-time controller in the complete actuation scenario. All the recordings of the throws are reported in the supplementary material (Data file). We have considered the following four states of the object: (E) enter; (BI) bounced-in; (BO) bounced-out; (O) out.

Optimisation based controller - complete actuation scenario												
Target	Tossed object											
	PingPong			Lemon			Marker			Tomato		
A	E	E	E	E	E	E	E	E	E	BI	BI	BI
B	E	E	E	E	E	E	E	E	E	E	E	BI
C	BO	BO	BO	E	E	E	E	E	E	BI	BI	E
D	O	BO	O	BO	BO	BO	E	E	E	E	E	BI
E	E	E	E	E	E	E	E	E	E	BI	E	E
F	E	E	BO	BI	BO	BO	BO	O	BO	BO	BO	BI
G	E	E	E	E	E	E	E	E	E	E	E	E
H	E	E	E	E	E	E	E	E	E	BO	BI	BI
I	E	E	E	BO	BO	E	O	O	O	BI	BI	BO
L	BO	BO	BO	E	E	E	E	E	E	BO	BI	BI
$\sum$ **E**		20			22			24			10	
$\sum$ **BI**		0			1			0			15	
$\sum$ **BO**		8			7			2			5	
$\sum$ **O**		2			0			4			0	

Table SV

Quantitative analysis of the trajectories recorded related to Deep reinforcement learning controller in the complete actuation scenario. Thanks to the VICON system we recorded the trajectories of the object in the different trials. Each value represents the average of the distances, in millimeters, from the center of the box selected as a target, in the three trials performed.

Deep reinforcement learning controller - complete actuation scenario				
Target	Tossed object			
	PingPong	Lemon	Marker	Tomato
A	179.27	139.67	122.36	87.94
B	92.72	67.41	30.43	38.02
C	47.63	61.54	63.50	61.52
D	47.70	66.56	81.69	63.36
E	50.12	69.49	68.26	71.15
F	111.49	128.45	119.22	144.12
G	32.99	39.26	36.01	46.08
H	41.38	49.81	30.78	52.27
I	97.23	9.35	37.66	53.75
L	51.73	36.13	62.42	41.61

Table SVI

Quantitative analysis of the trajectories recorded related to Deep reinforcement learning controller in the partial actuation scenario. THANKS TO THE VICON SYSTEM WE RECORDED THE TRAJECTORIES OF THE OBJECT IN THE DIFFERENT TRIALS. EACH VALUE REPRESENTS THE AVERAGE OF THE DISTANCES, IN MILLIMETERS, FROM THE CENTER OF THE BOX SELECTED AS A TARGET, IN THE THREE TRIALS PERFORMED.

Deep reinforcement learning controller - partial actuation scenario				
Target	Tossed object			
	PingPong	Lemon	Marker	Tomato
A	42.99	24.89	35.61	31.34
B	42.22	15.89	19.95	24.19
C	31.78	25.11	46.84	67.75
D	105.00	99.31	86.50	72.33
E	68.50	60.94	45.42	58.45
F	72.82	108.87	90.78	112.45
G	22.80	58.85	45.43	70.70
H	35.22	83.39	53.59	94.81
I	37.31	95.32	95.30	95.77
L	62.31	65.24	84.22	79.25

Table SVII

Quantitative analysis of the trajectories recorded related to Optimization-based controller in the complete actuation scenario. THANKS TO THE VICON SYSTEM WE RECORDED THE TRAJECTORIES OF THE OBJECT IN THE DIFFERENT TRIALS. EACH VALUE REPRESENTS THE AVERAGE OF THE DISTANCES, IN MILLIMETERS, FROM THE CENTER OF THE BOX SELECTED AS A TARGET, IN THE THREE TRIALS PERFORMED.

Non-real-time controller - complete actuation scenario				
Target	Tossed object			
	PingPong	Lemon	Marker	Tomato
A	80.99	46.31	45.73	39.92
B	66.24	51.37	29.97	32.76
C	44.02	48.15	36.34	95.66
D	130.04	75.59	76.10	58.40
E	52.81	38.74	38.80	43.79
F	111.80	135.28	119.18	147.09
G	62.93	55.88	42.64	74.09
H	24.15	85.46	49.70	52.46
I	81.22	75.70	98.33	93.06
L	77.15	17.37	38.59	32.76

Table SVIII

Errors of the RL agent in the direct model. DISTANCE IN MILLIMETERS OF THE LANDING POSITIONS ASSOCIATED WITH THE ACTUATION PATTERNS IDENTIFIED BY THE REINFORCEMENT LEARNING AGENT AND THE RESPECTIVE TARGETS.

Landing position errors in the Direct Model		
Target	Partial actuation	Complete actuation
A	12.38	42.29
B	27.32	20.30
C	20.13	19.07
D	29.49	39.76
E	35.16	37.33
F	28.96	27.16
G	28.07	49.50
H	48.18	24.72
I	2.43	19.17
L	11.74	26.23

Table SIX

Comparison of RL agents in object-specific direct models. ANALYSIS OF THE REINFORCEMENT LEARNING (RL) AGENTS WITHIN THE CONTEXT OF OBJECT-SPECIFIC DIRECT MODELS. THE PERFORMANCE OF THESE AGENTS IS EVALUATED ON THE MINIMUM, AVERAGE, AND MAXIMUM DISTANCES (IN MILLIMETERS) OF THE LANDING POSITIONS ACHIEVED BY THE CORRESPONDING ACTUATION PATTERNS. THE RL AGENTS' POLICIES ARE TRAINED AND TESTED USING THE ENVIRONMENT BUILT UPON THE OBJECT-SPECIFIC DIRECT MODEL.

Scenario	Partial				Complete			
Object	*Average*	Lemon	PingPong	Marker	*Average*	Lemon	PingPong	Marker
Min	2.43	15.96	13.48	14.45	19.07	4.23	6.14	1.84
Average	24.38	26.50	23.53	25.21	30.55	21.17	22.61	19.90
Max	48.18	43.36	48.12	42.84	49.50	40.34	40.64	34.86