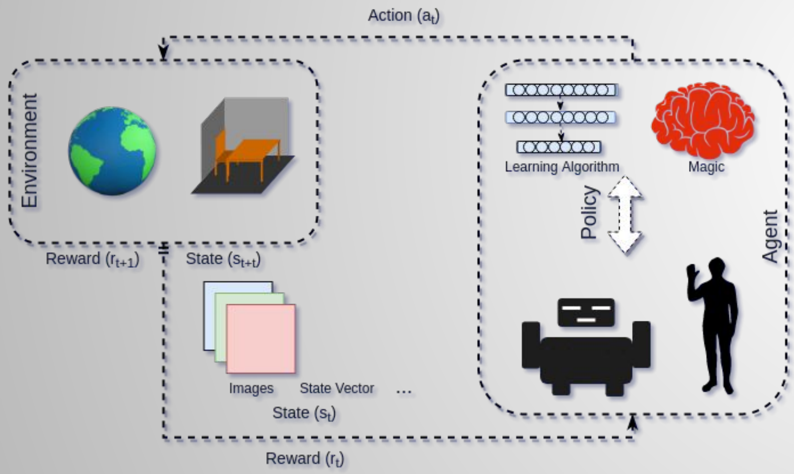
Imperial College
London

Deep Reinforcement Learning on Robots

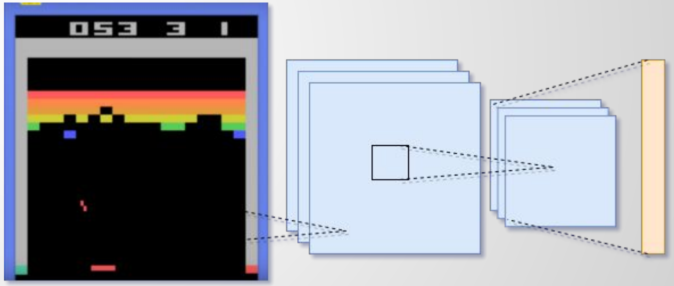
An Evaluation of Models, Demonstration and Rewards

Imperial College
London

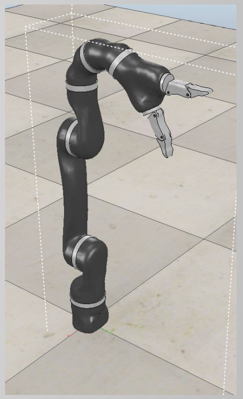
Deep Reinforcement Learning



Reinforcement Learning



DRL learns to play Atari Games



Mico Robot Arm

Imperial College London

DRL on Robotics: Challenges

Training on real robots

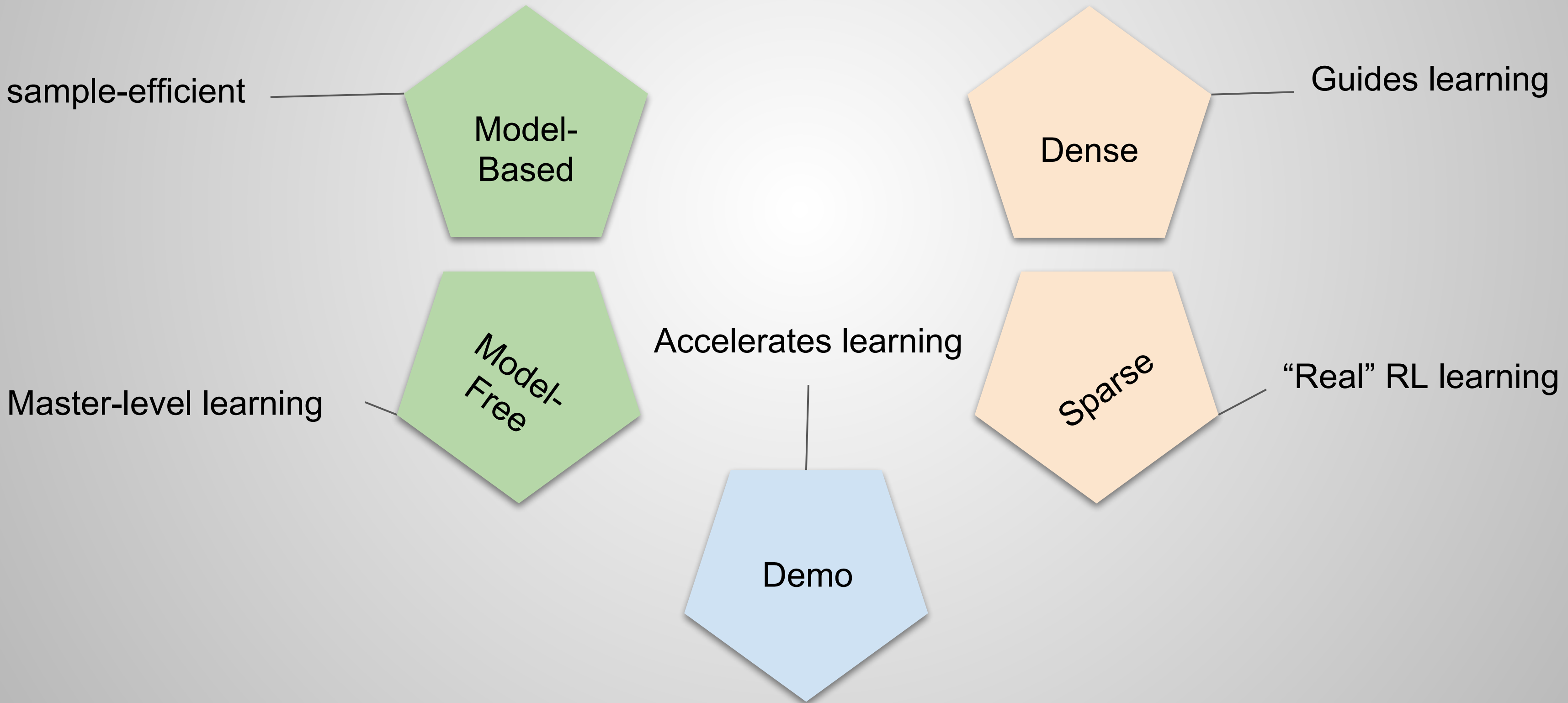
- Safety
- Interaction time/Sample efficiency

Training in simulation

- Transferring from simulation to real-world. Reality Gap.

Imperial College
London

Models, Rewards and Demonstrations

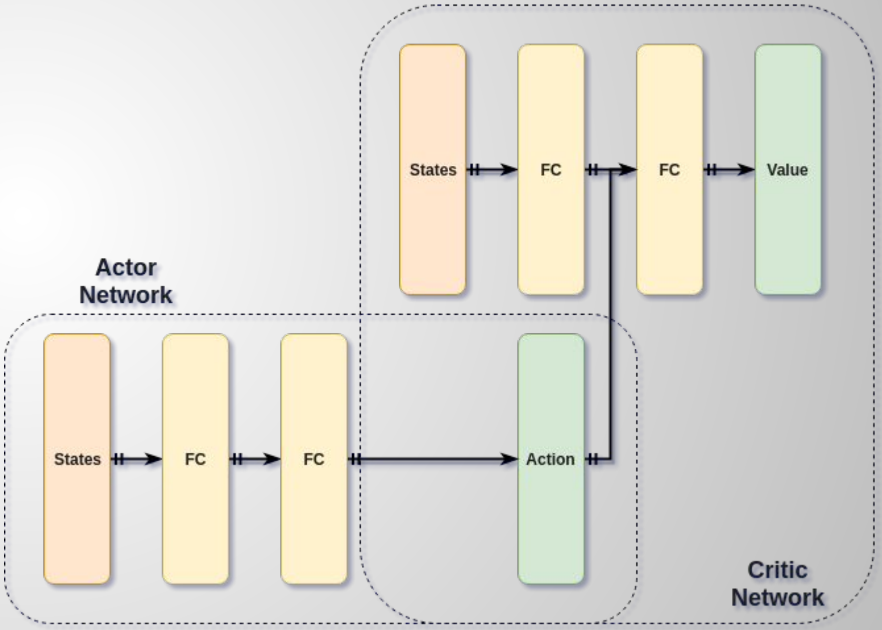


Imperial College
London

Agents: Model-Free

DDPGfD

- Actor-Critic
- Demonstration loaded to agent's experience replay

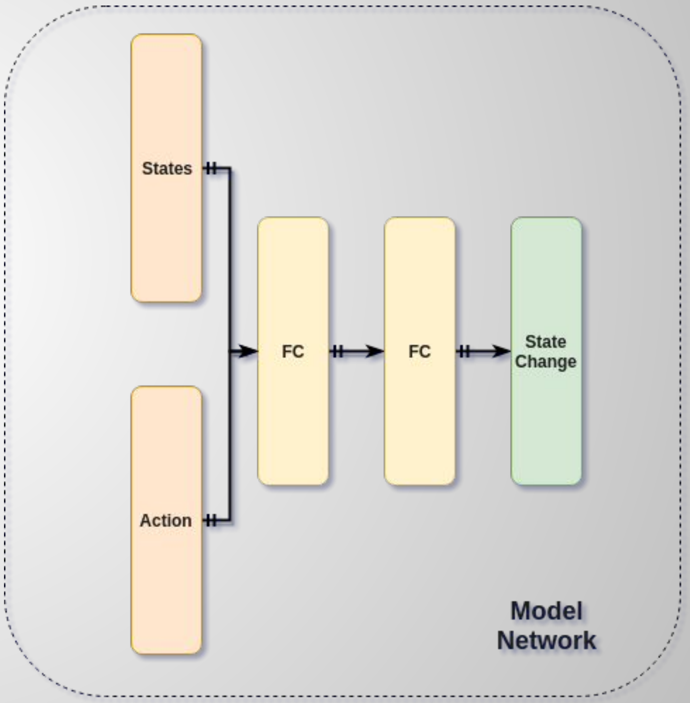


Imperial College
London

Agents: Model-Based

Neural Network Dynamics Model with Model Predictive Control(MPC)

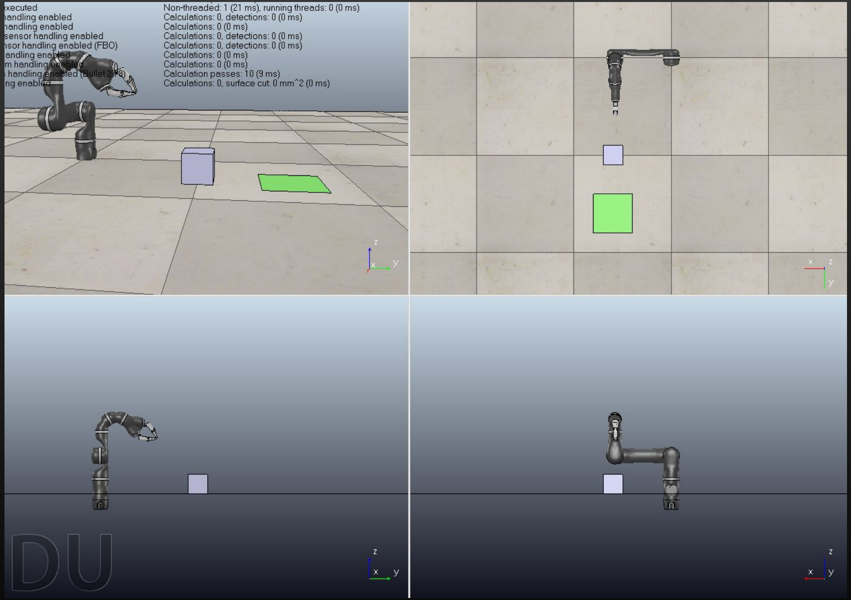
- Only Model Networks
- Sampling-based MPC
- Demonstration loaded to agent's experience replay



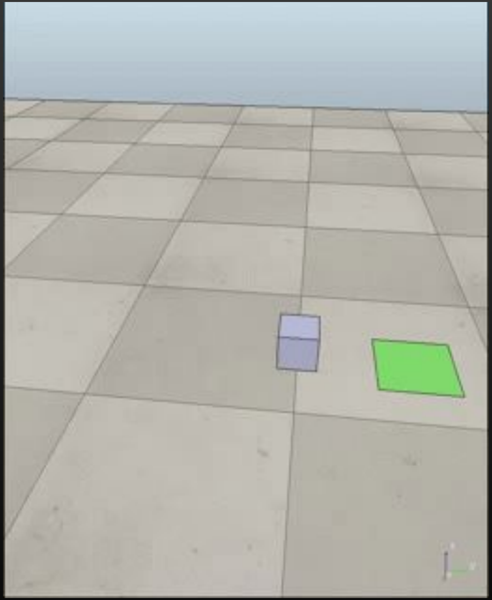
Tasks

Imperial College
London

Tasks: Push



Initial State

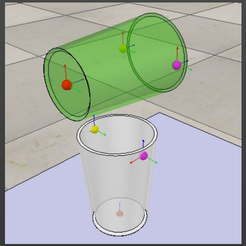


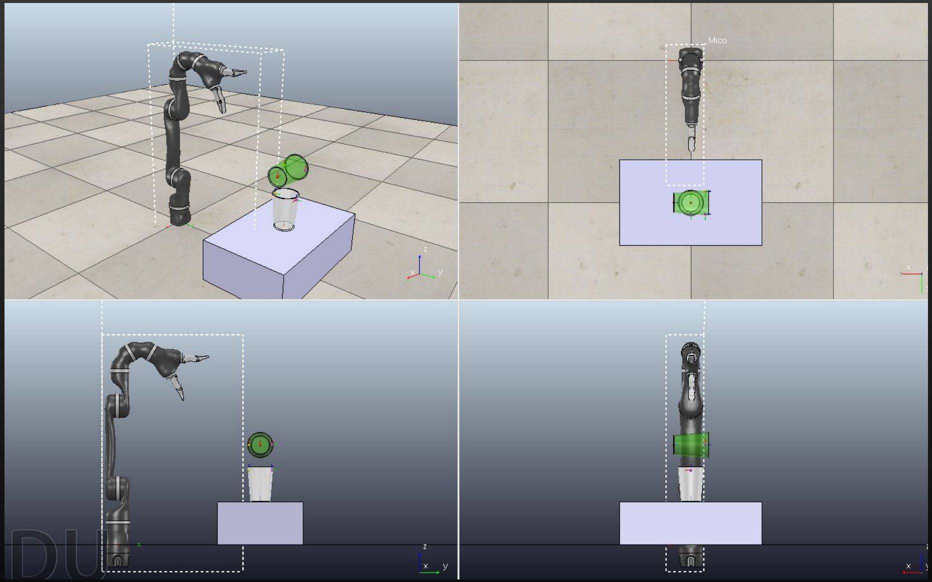
Demonstration
Collection

Imperial College
London

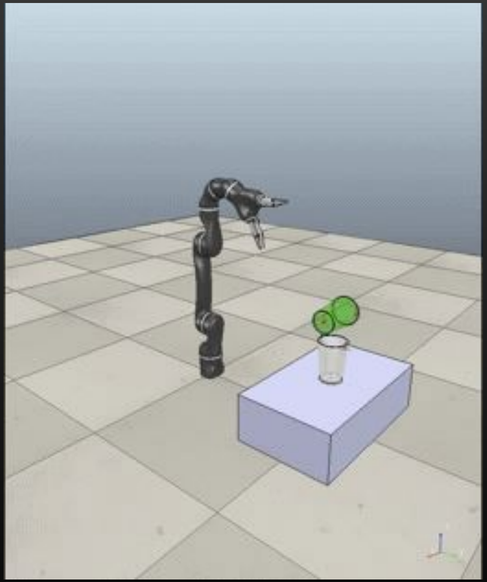
Task: Grasp

Surface points





Initial State



Demonstration
Collection

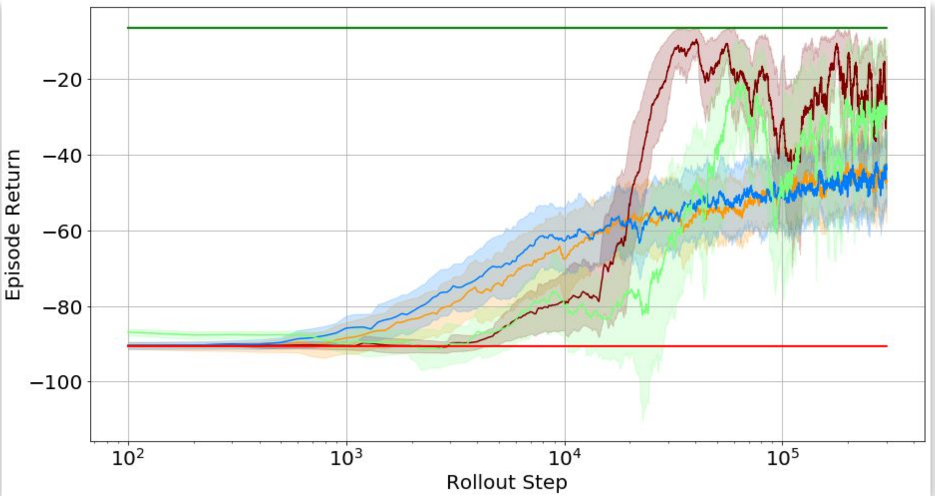
Experiments and Evaluation

Imperial College
London

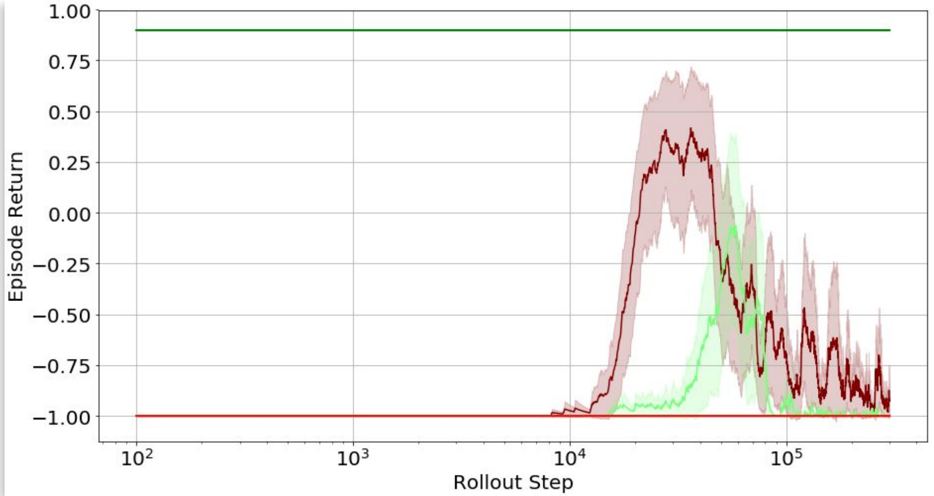
Push Task

Imperial College London

Training Statistics: Episode Return



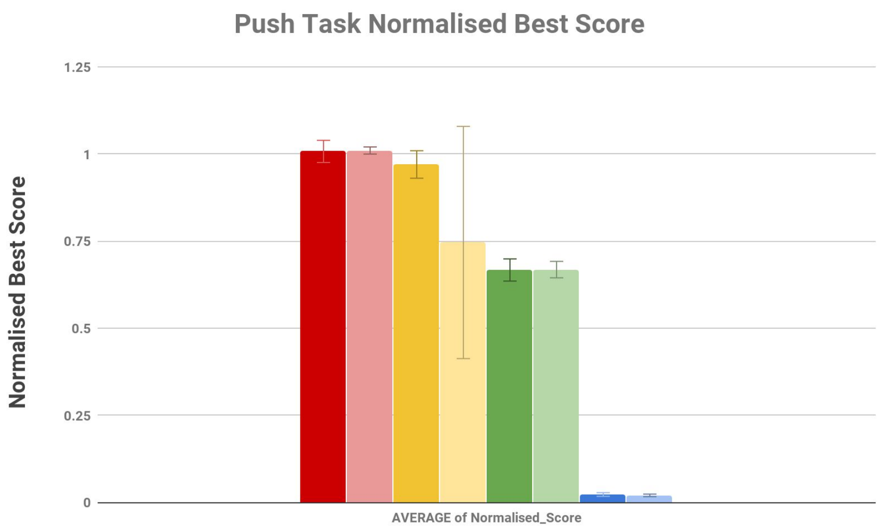
Dense

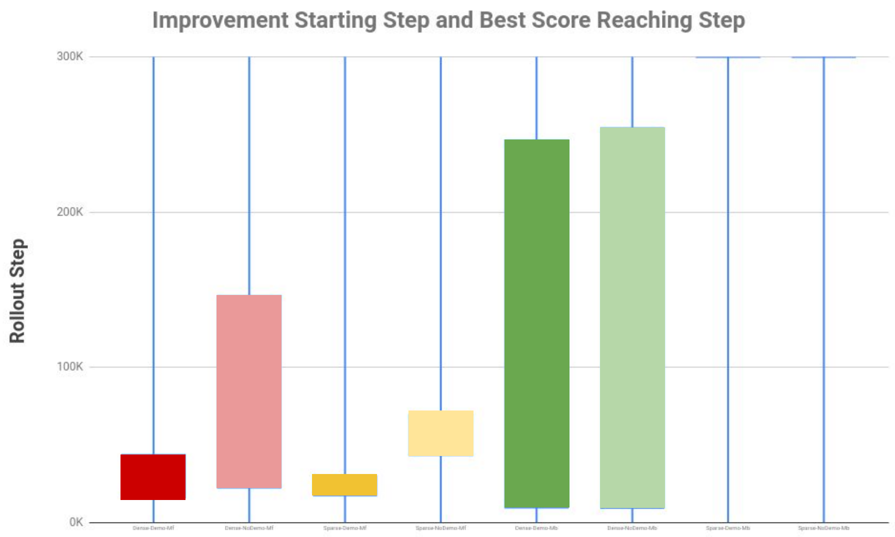


Sparse

Mf-Demo **Mf-NoDemo** **Mb-Demo** **Mb-NoDemo**

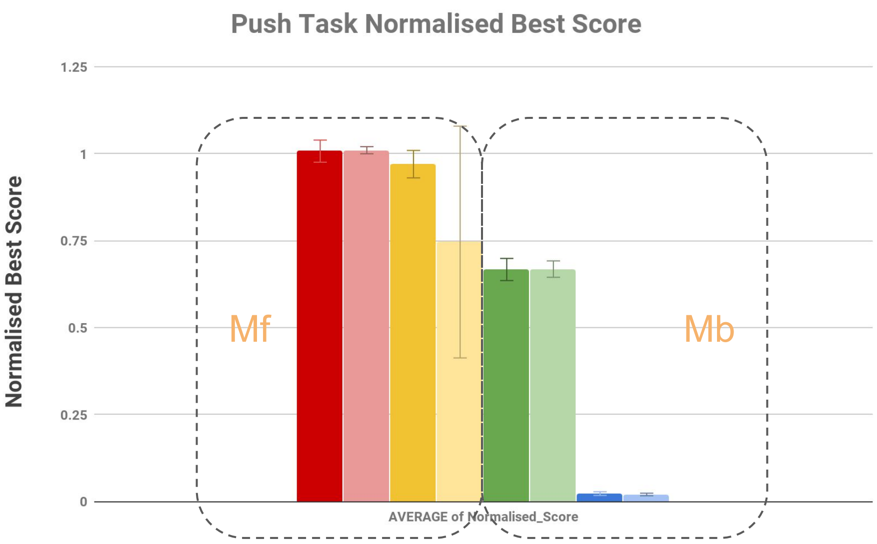
Benchmark-Demo **Benchmark-Worst**

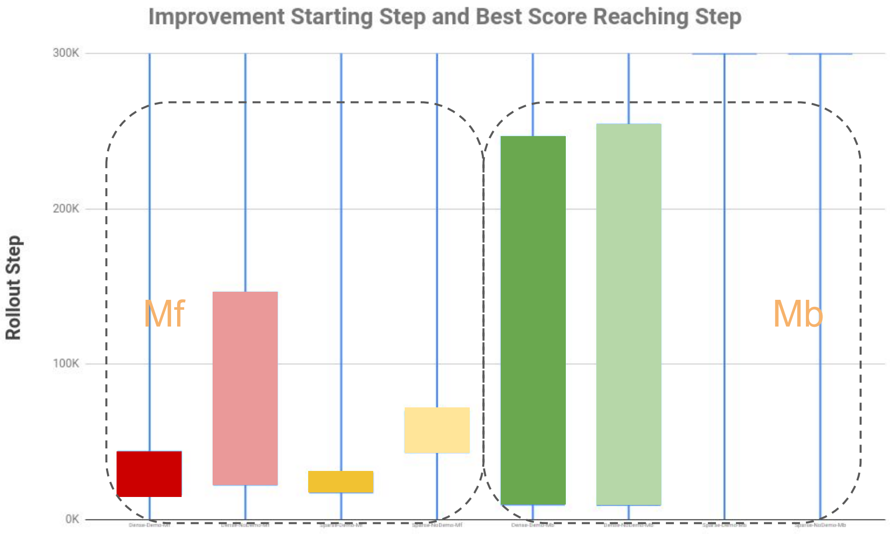




Mf-Dense-Demo
Mf-Dense-NoDemo
Mf-Sparse-Demo
Mf-Sparse-NoDemo
Mb-Dense-Demo
Mb-Dense-NoDemo
Mb-Sparse-Demo
Mb-Sparse-NoDemo

Imperial College
London



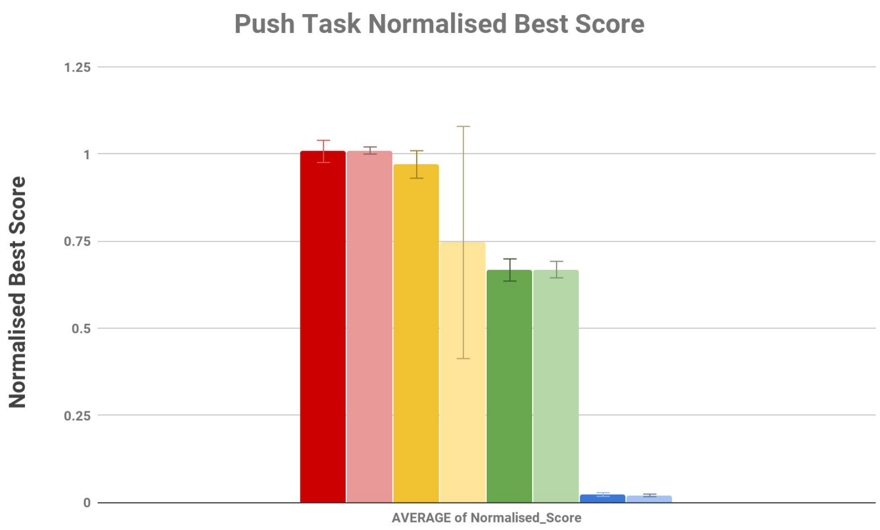


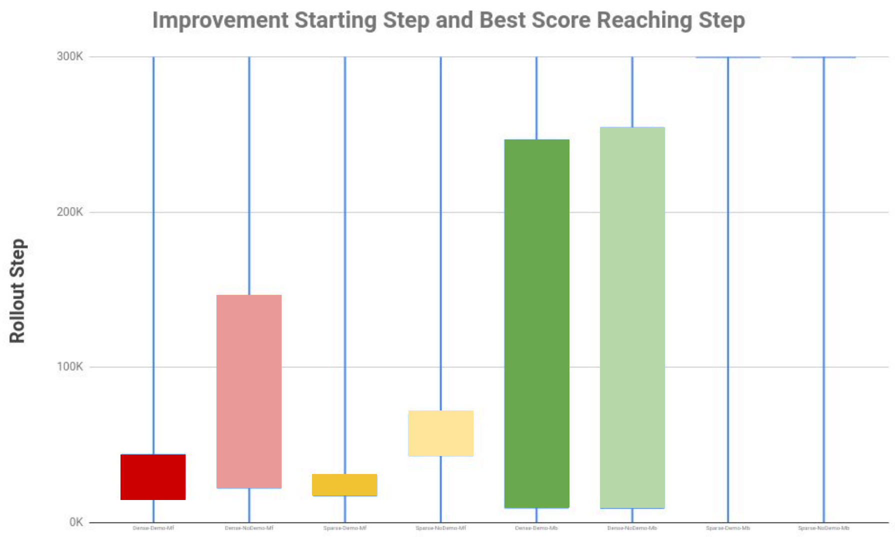
Mf-Dense-Demo Mf-Dense-NoDemo Mf-Sparse-Demo Mf-Sparse-NoDemo

Mb-Dense-Demo Mb-Dense-NoDemo Mb-Sparse-Demo Mb-Sparse-NoDemo

Imperial College London

Imperial College
London

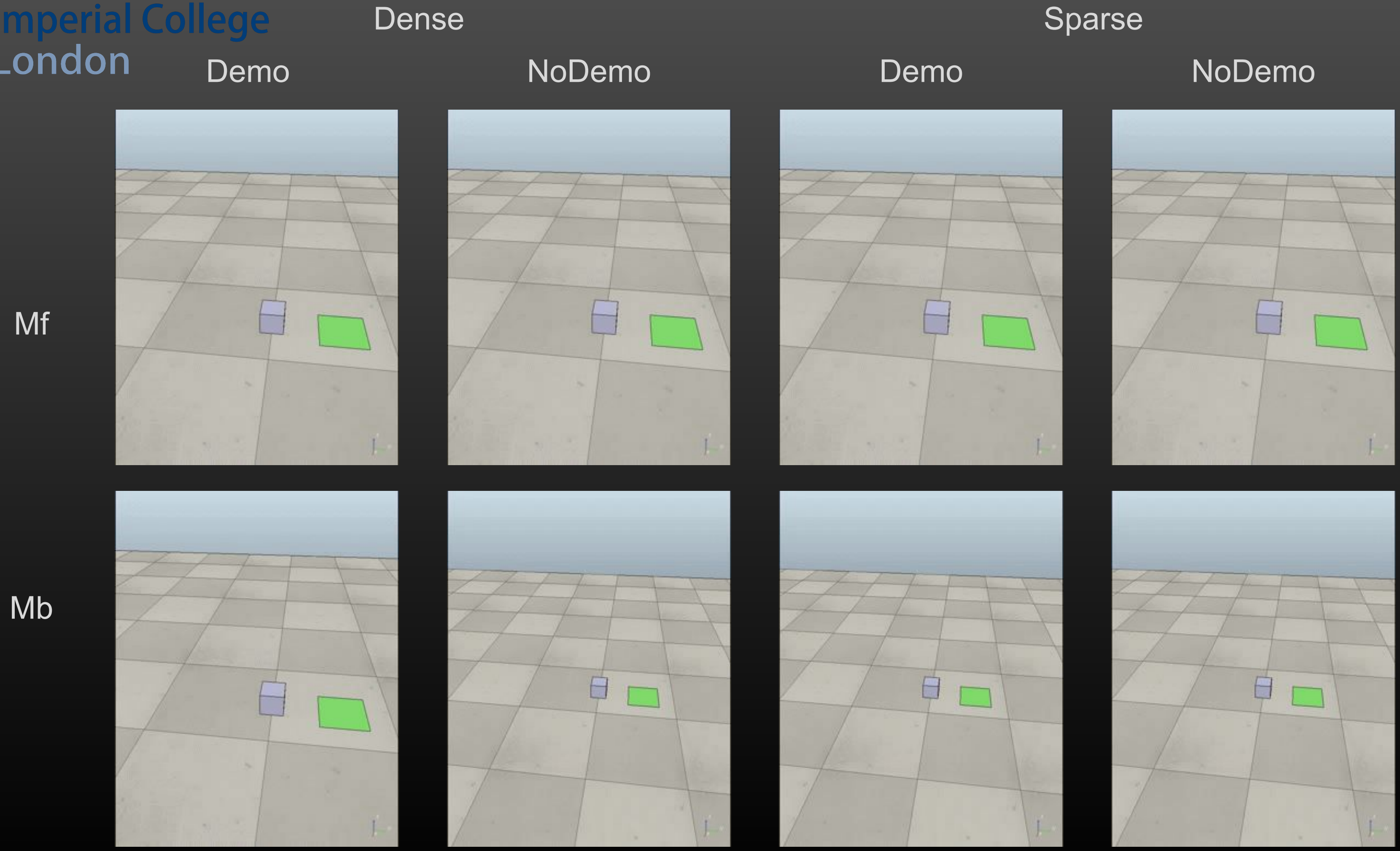




Mf-Dense-Demo Mf-Dense-NoDemo Mf-Sparse-Demo Mf-Sparse-NoDemo

Mb-Dense-Demo Mb-Dense-NoDemo Mb-Sparse-Demo Mb-Sparse-NoDemo

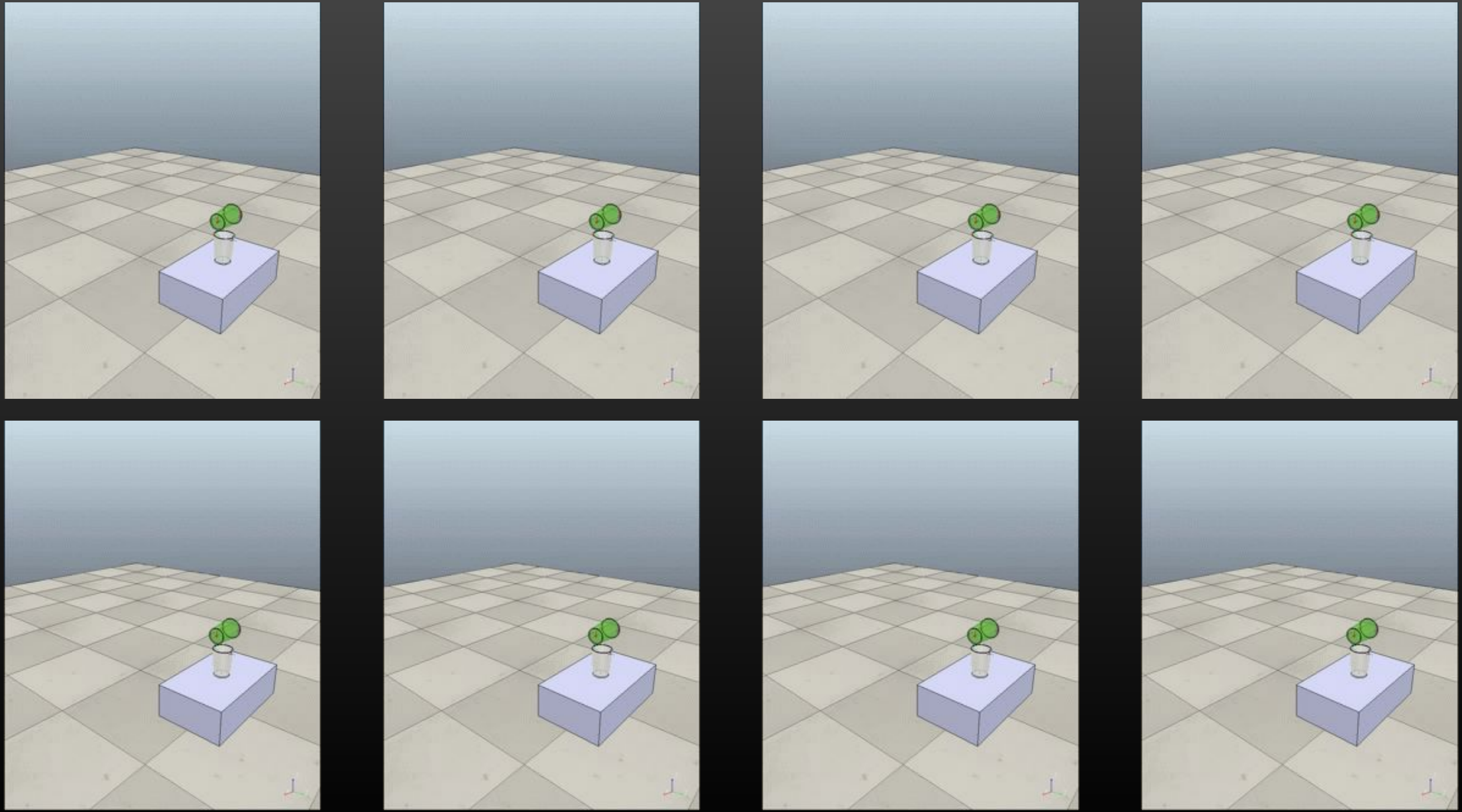
Imperial College London



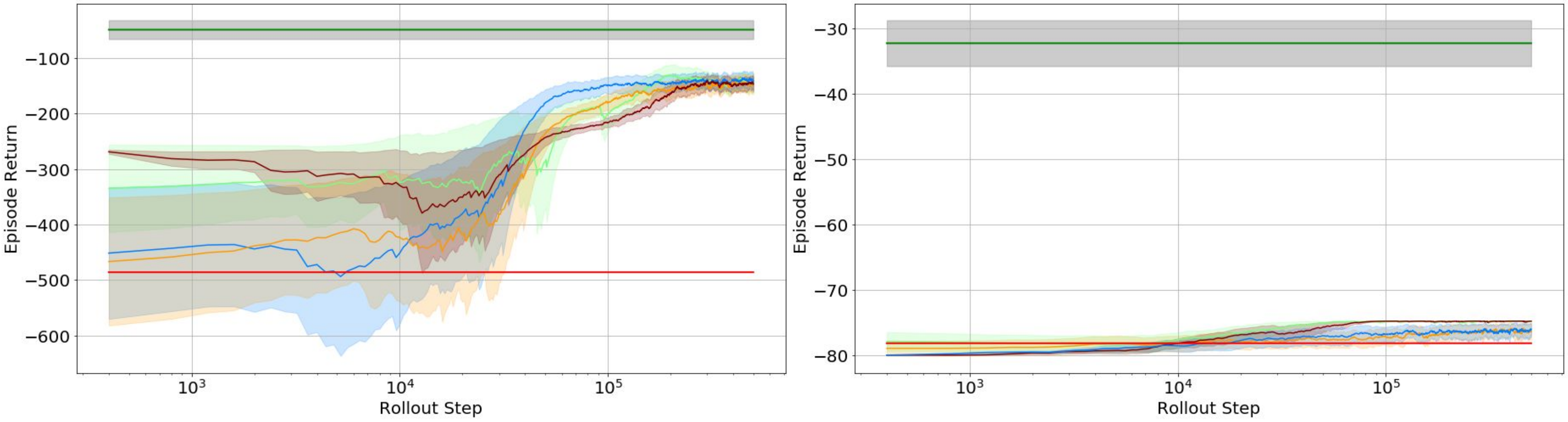
Grasp Task

Imperial College London

	Dense		Sparse	
	Demo	NoDemo	Demo	NoDemo
Mf				
Mb				



Training Statistics: Episode Return



Dense

Sparse

Mf-Demo

Mf-NoDemo

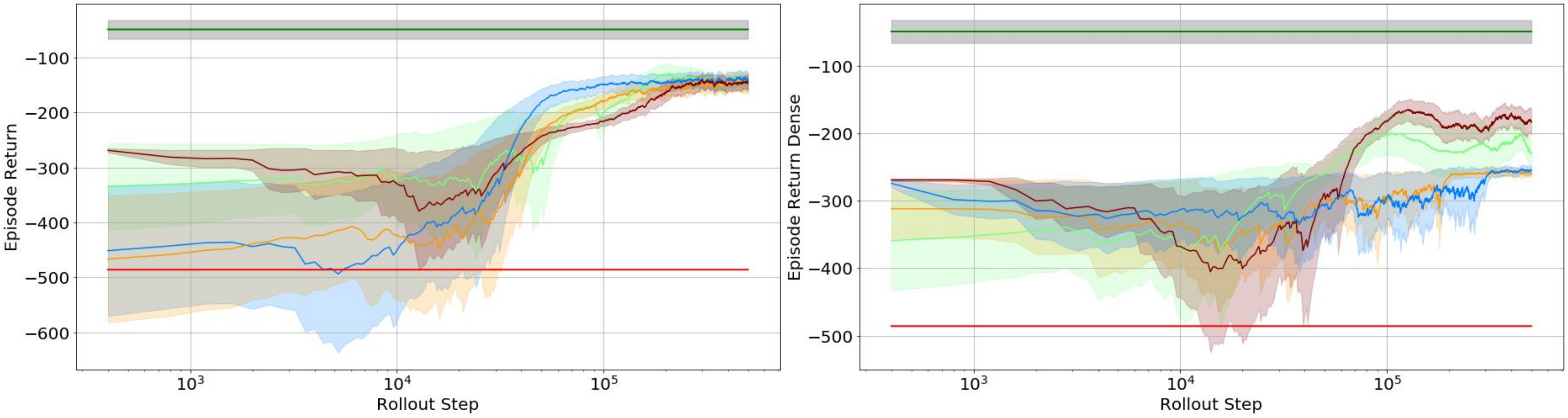
Mb-Demo

Mb-NoDemo

Benchmark-Demo

Benchmark-Worst

Training Statistics: Episode Return



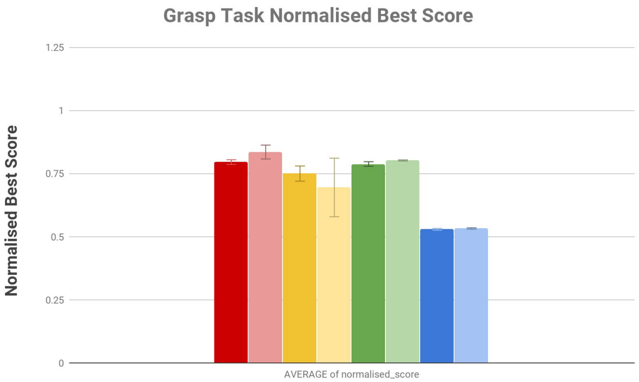
Dense

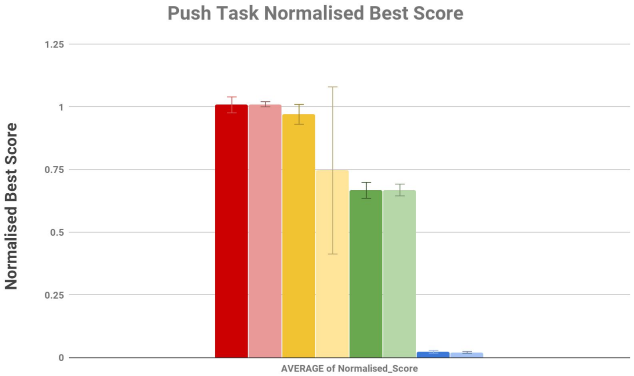
Sparse (With Dense Score)

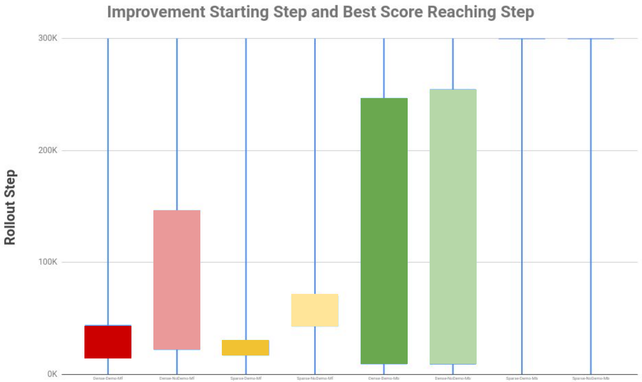
Mf-Demo **Mf-NoDemo** **Mb-Demo** **Mb-NoDemo**

Benchmark-Demo **Benchmark-Worst**

Imperial College
London





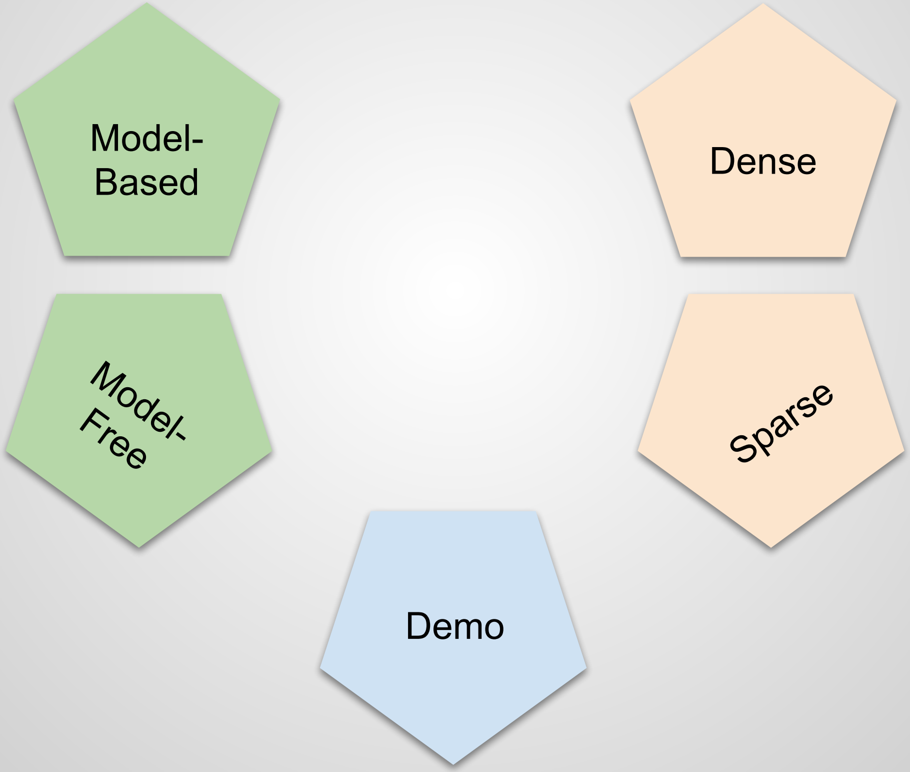


Mf-Dense-Demo Mf-Dense-NoDemo Mf-Sparse-Demo Mf-Sparse-NoDemo

Mb-Dense-Demo Mb-Dense-NoDemo Mb-Sparse-Demo Mb-Sparse-NoDemo

Conclusion

Imperial College London



Imperial College
London

Future Work

Imperial College
London

Thank you very much!

Imperial College
London

Appendix

Model

Model-Based	Model-Free
+ Sample-efficient + Stable training process	+ Master-level learning + Operates under sparse rewards
- Typically worse asymptotic performance - Requires dense reward function	- Needs a lot more training data - Training can be unstable

Rewards

Dense Reward	Sparse Reward
+ Provides a strong reward signal	+ Intuitive to define + Real autonomous learning
- May be impractical to define in complex tasks - May lead to sub-optimal solutions	- Hard to learn

Imperial College
London

Demonstration

+	Accelerates learning
-	May be hard to collect