Learning Latent Dynamics for Planning from Pixels

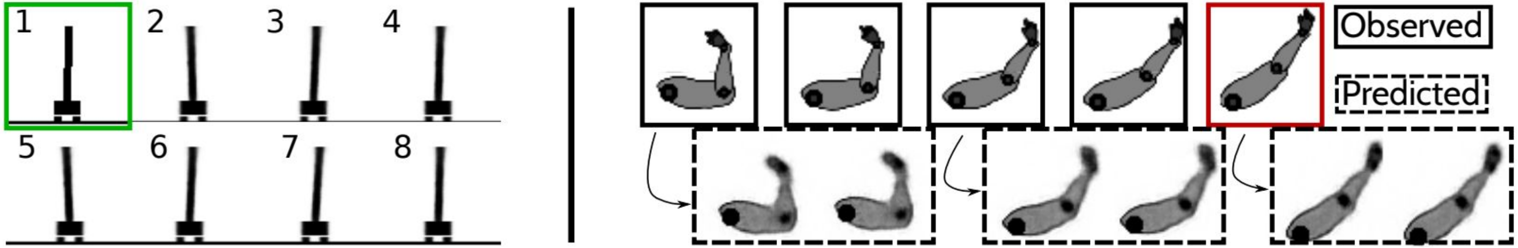
Danijar Hafner, Timothy Lillicrap, Ian Fischer, Ruben Villegas, David Ha, Honglak Lee, James Davidson

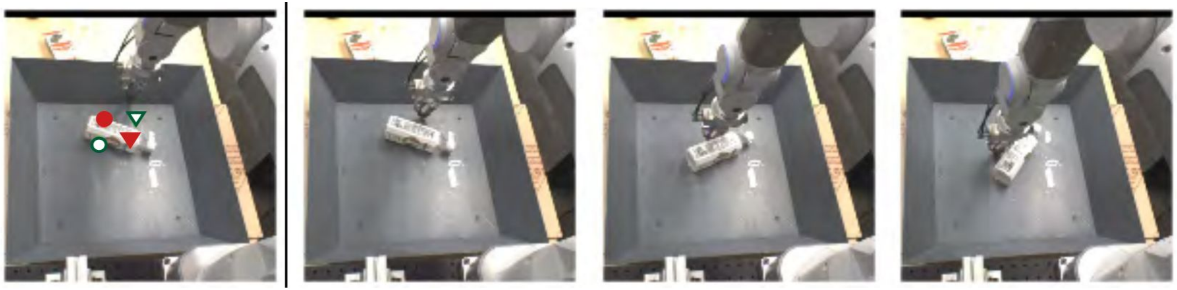
@danijarh

danijar.com/planet

Planning with Learned Models

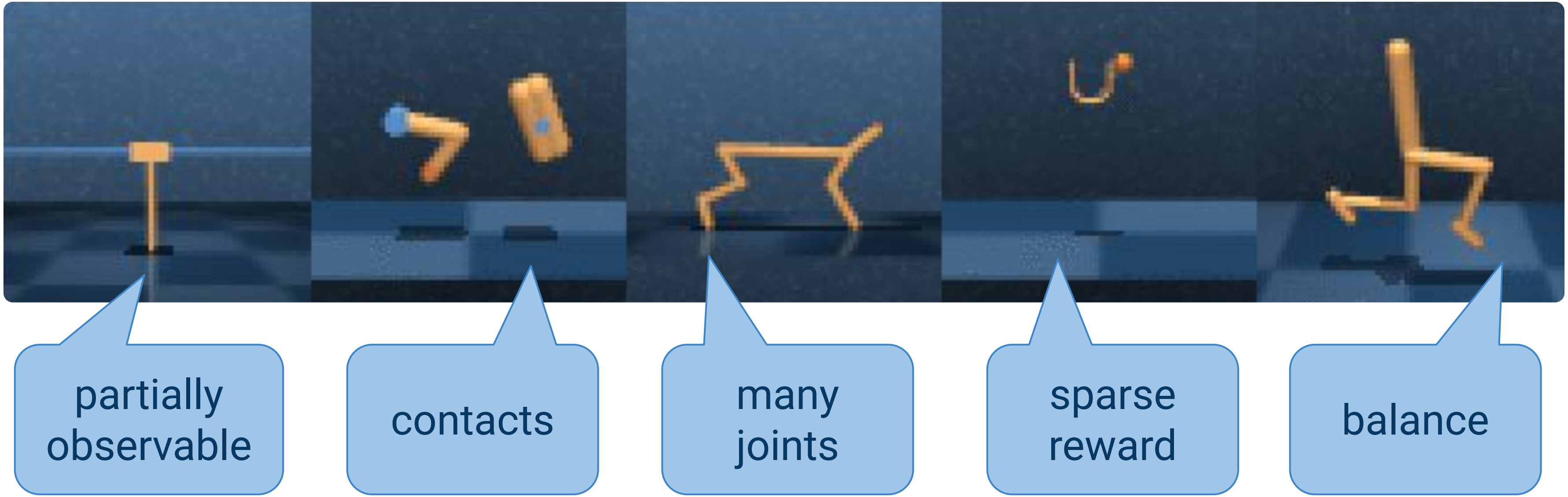


Watter et al., 2015, Banijamali et al. 2017, Zhang et al. 2017



Agrawal et al., 2016; Finn & Levine, 2016; Ebert et al., 2018

Visual Control Tasks



Some model-free methods can solve these tasks but need up to 100,000 episodes

Visual Control Tasks

partially observable

contacts

many joints

sparse reward

balance

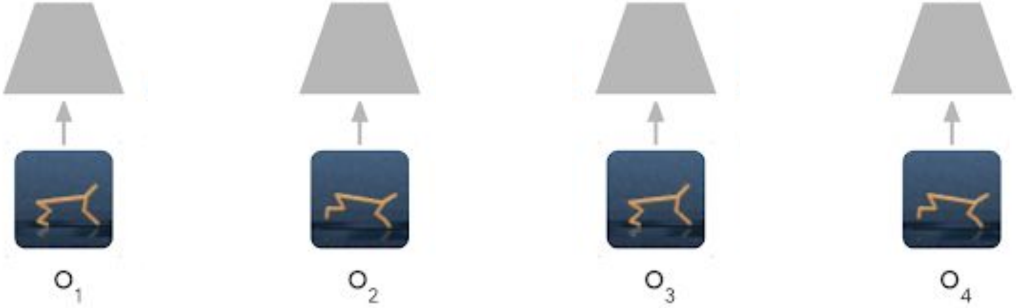
Some model-free methods can solve these tasks but need up to 100,000 episodes

We introduce PlaNet

1. Recipe for scalable model-based reinforcement learning
2. Efficient planning in latent space with large batch size
3. Reaches top performance using 200X fewer episodes

Latent Dynamics Model

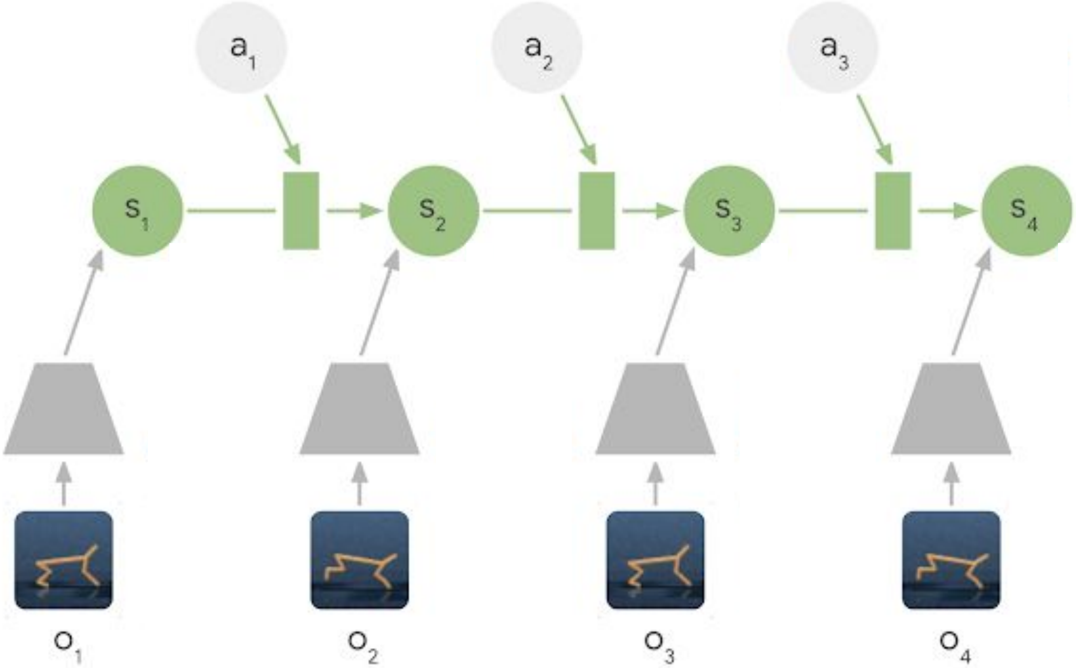
 encode images



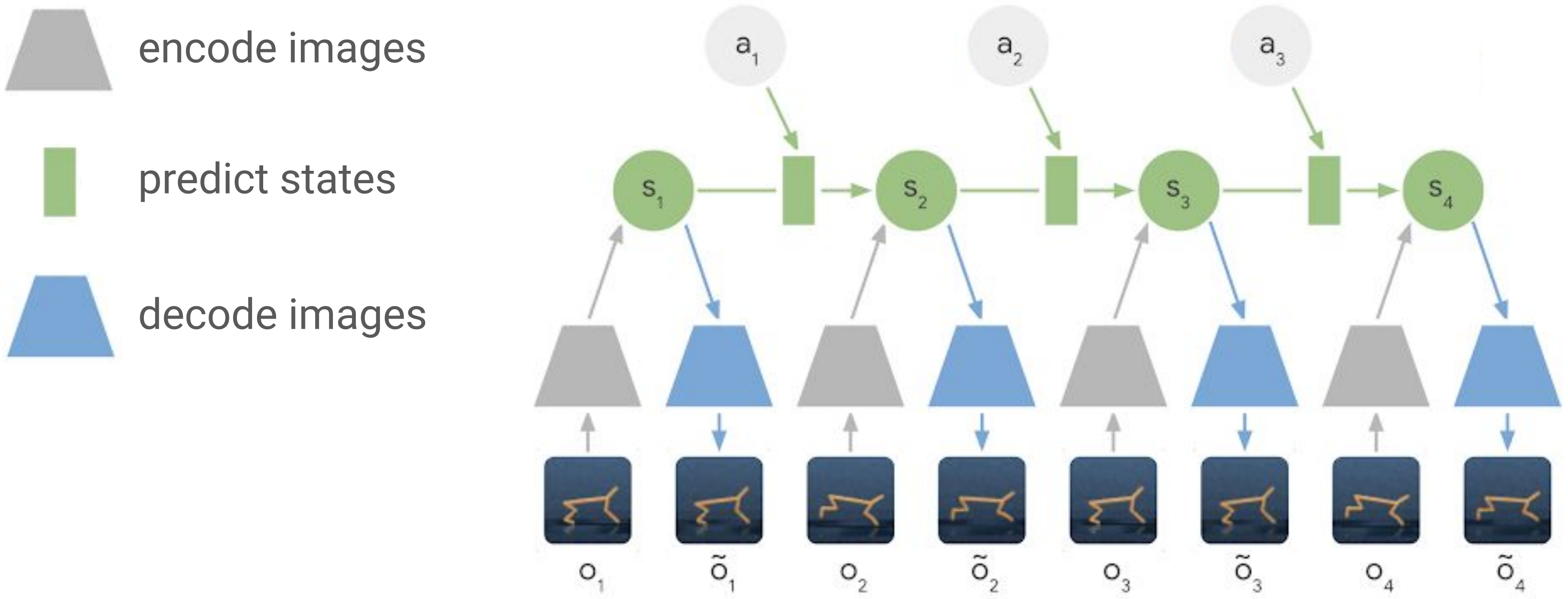
Latent Dynamics Model

encode images

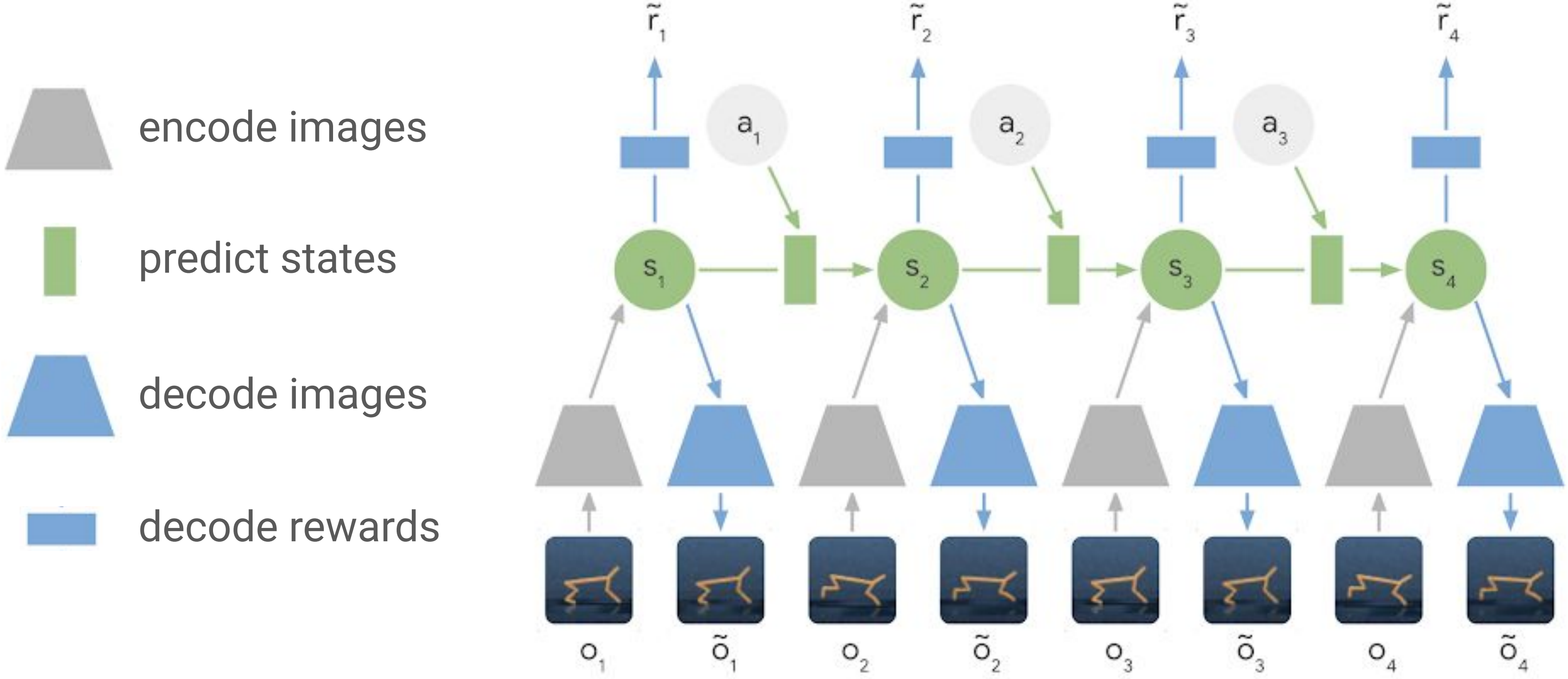
predict states



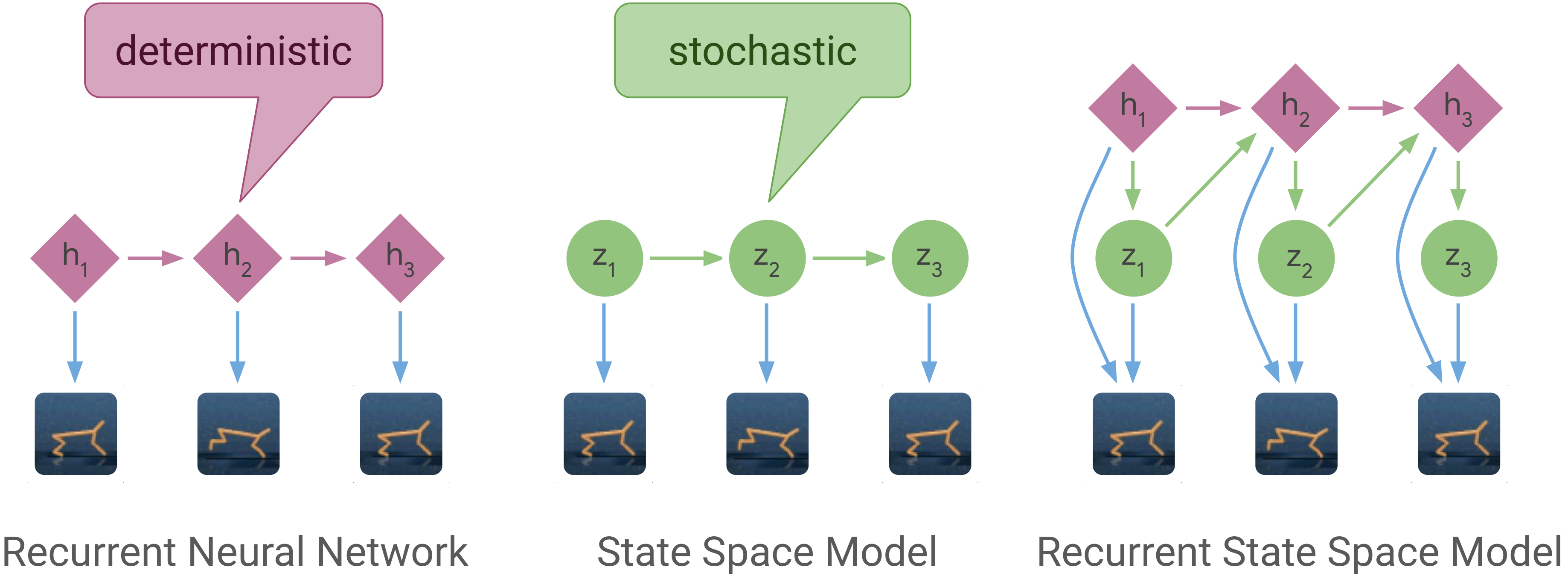
Latent Dynamics Model



Latent Dynamics Model

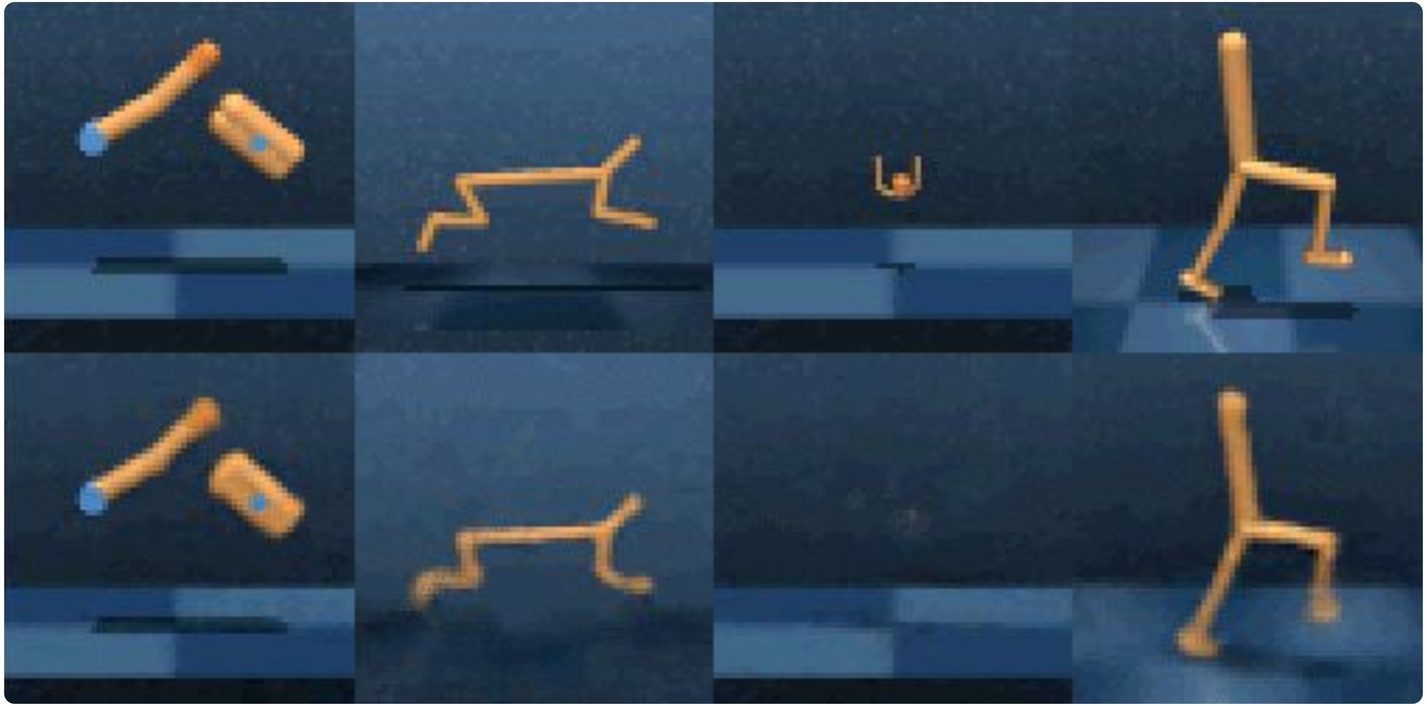


Recurrent State Space Model



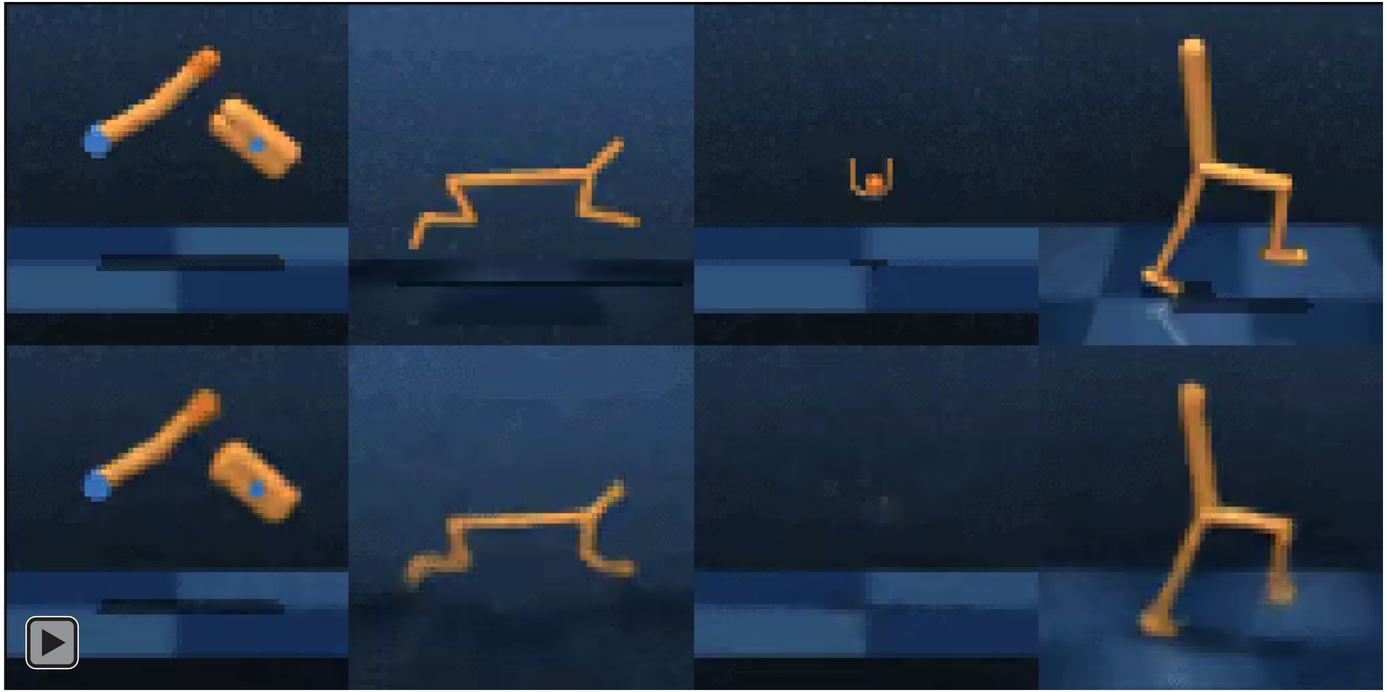
Unguided Video Predictions by Single Agent

5 frames context and 45 frames predicted

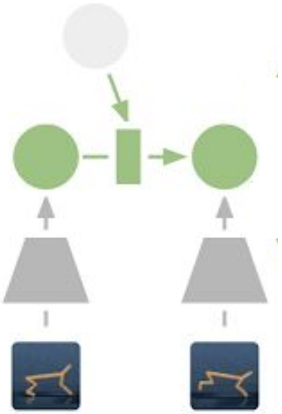


Unguided Video Predictions by Single Agent

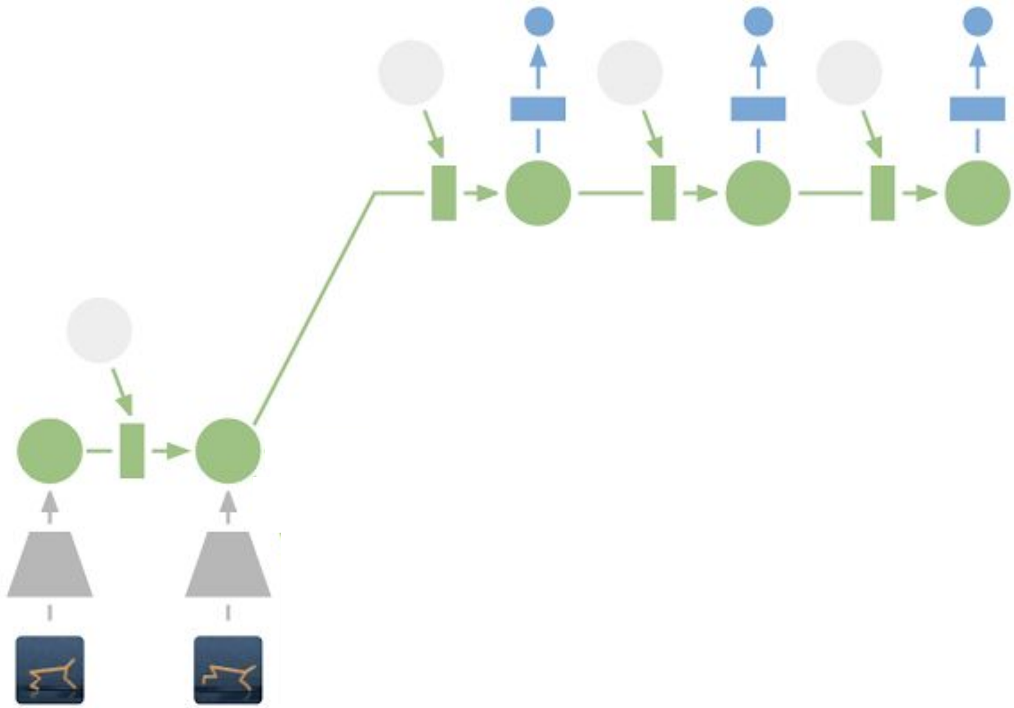
5 frames context and 45 frames predicted



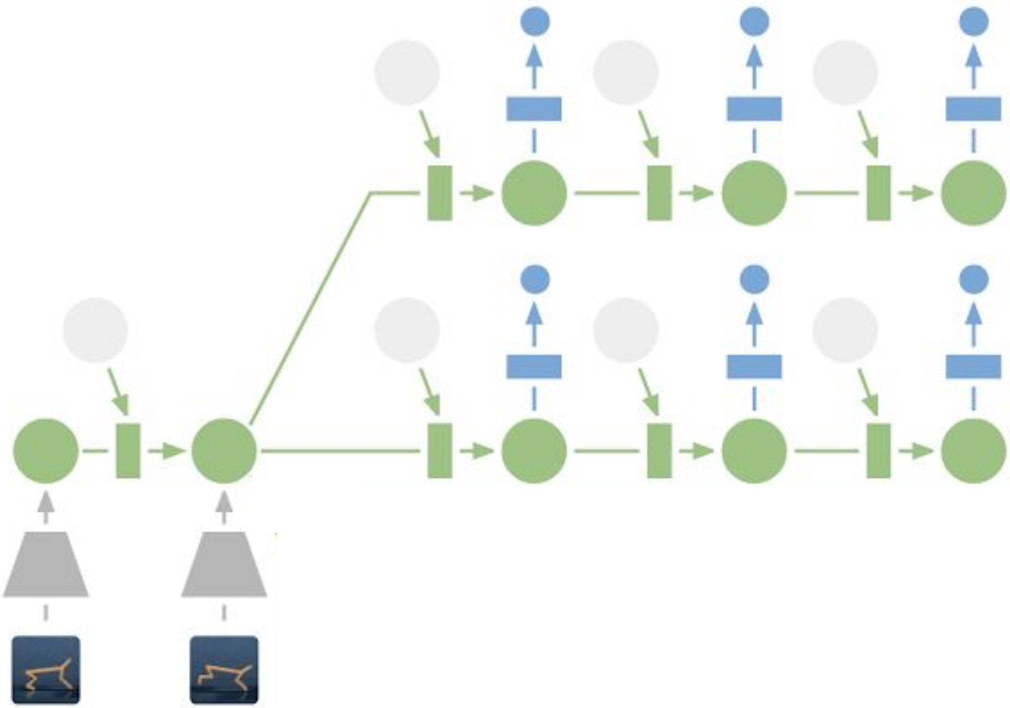
Planning in Latent Space



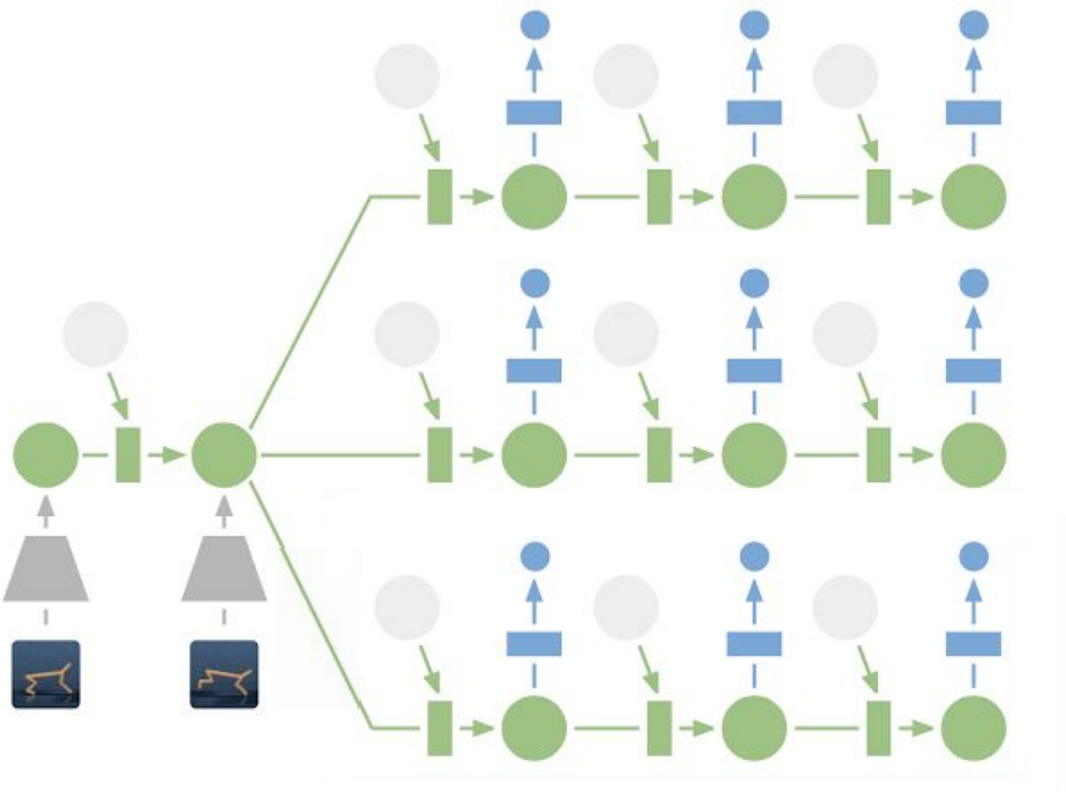
Planning in Latent Space



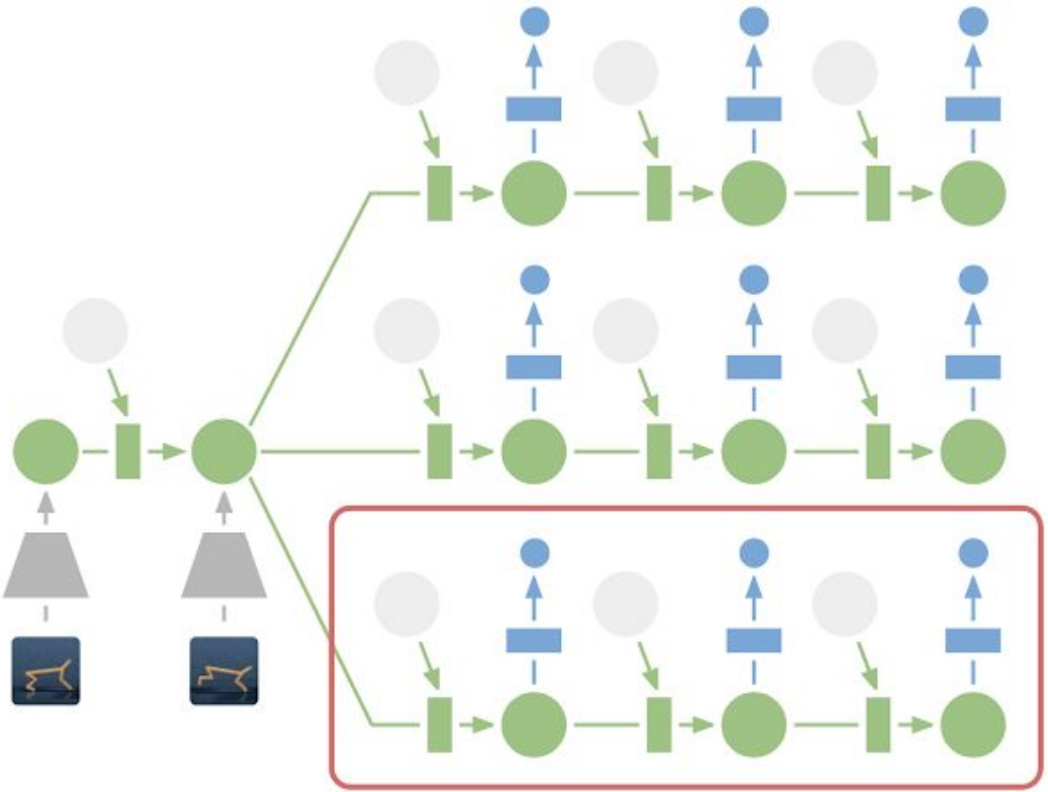
Planning in Latent Space



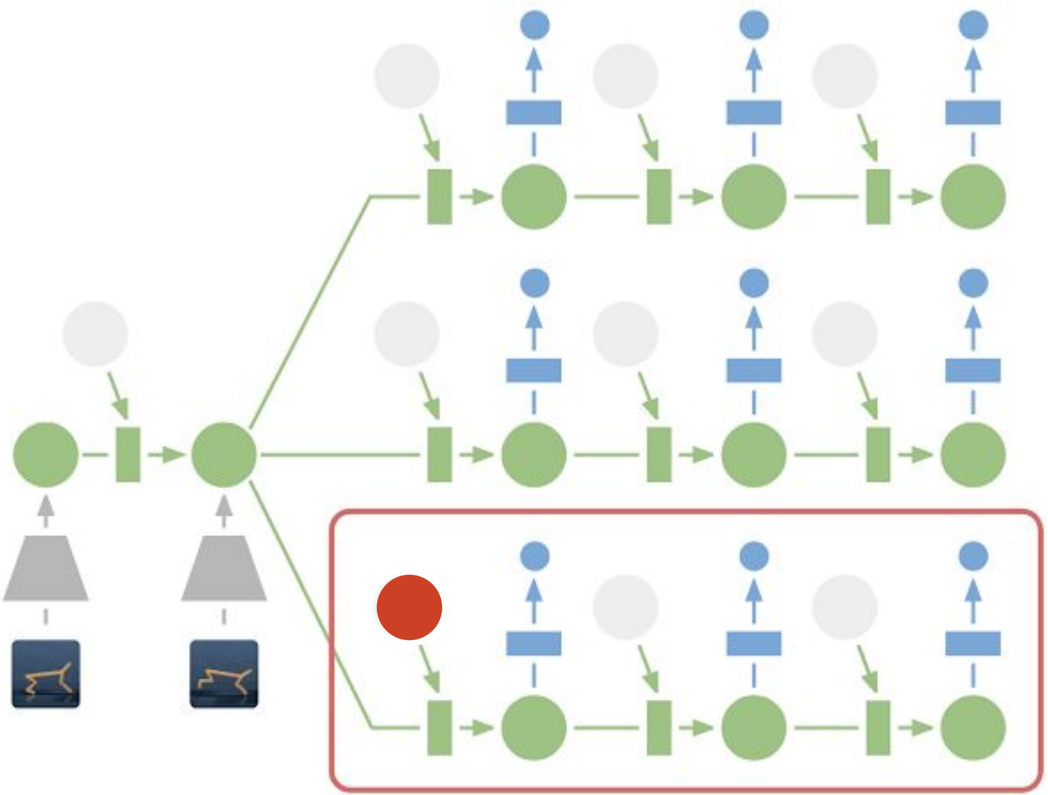
Planning in Latent Space



Planning in Latent Space



Planning in Latent Space

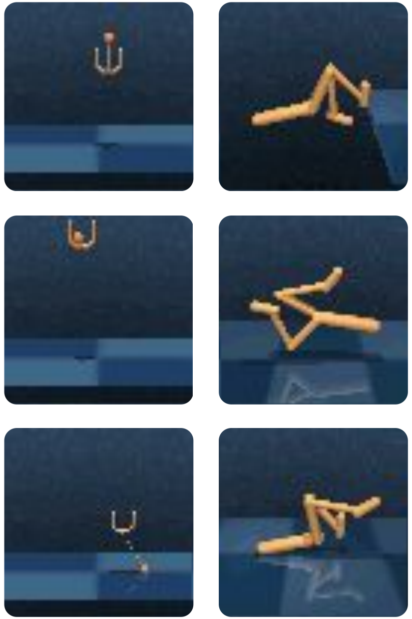


Comparison to Model-Free Agents

Method	Modality	Episodes	Cartpole Swing Up	Reacher Easy	Cheetah Run	Finger Spin	Cup Catch	Walker Walk
A3C	proprioceptive	100,000	558	285	214	129	105	311
D4PG	pixels	100,000	862	967	524	985	980	968
PlaNet (ours)	pixels	1,000	821	832	662	700	930	951
Data efficiency gain PlaNet over D4PG (factor)			250	40	500+	300	100	90

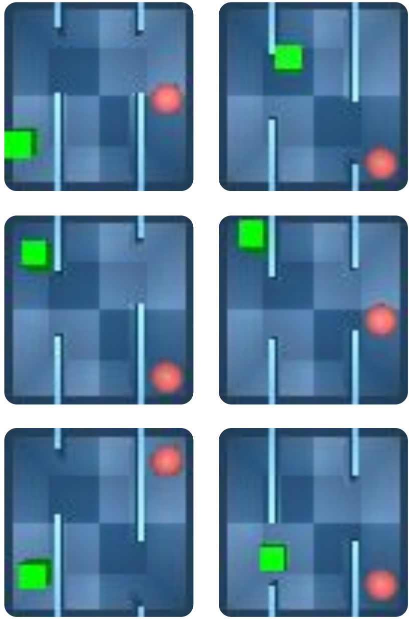
Training time 1 day on a single GPU

Enabling More Model-Based RL Research



Explore dynamics
without supervision

Distill the planner to save
computation



Value function to extend
planning horizon

Learning Latent Dynamics for Planning from Pixels

Website with code, videos, blog post, animated paper:

danijar.com/planet

33