Sample Complexity Analysis of Transfer Learning for Deep Reinforcement Learning Models

Malek Ben Alaya

March 11, 2022

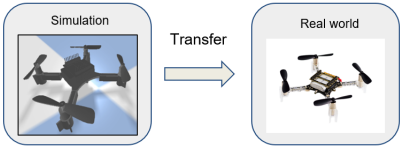
Motivation

- 'Project Phoenix': fly a quadcopter in the real world using deep Reinforcement Learning (RL).
- Deep RL methods:
  - Require a lot of data.
  - Have safety issues.
- Mitigate with Transfer Learning (TL).

Problem statement

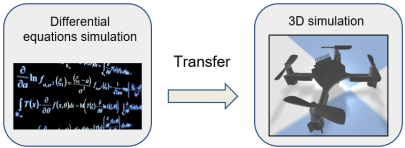
Problem:

- No guarantee TL is beneficial.
- Certain deep RL methods: inappropriate.



Solution:

- Test transfer in sim-to-sim.
- Poor deep RL methods in sim-to-sim: avoid them in sim-to-real.



Thesis contribution

Contribution of this work:
For a drone hovering task:

- Evaluate benefits of TL (sim-to-sim).
- Analyze sample complexity on 3D simulation.
- Conclude: most appropriate methods.

(Deep) Reinforcement Learning

Algorithms:

- **PPO:** on-policy, stochastic.
- **SAC:** off-policy, stochastic.
- **DDPG:** off-policy, deterministic.
- **TD3:** off-policy, deterministic.

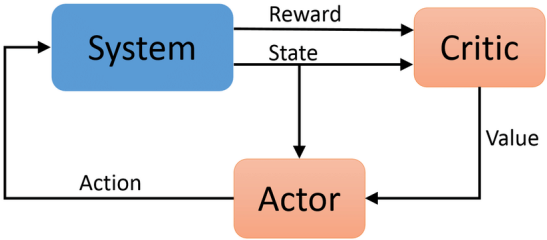
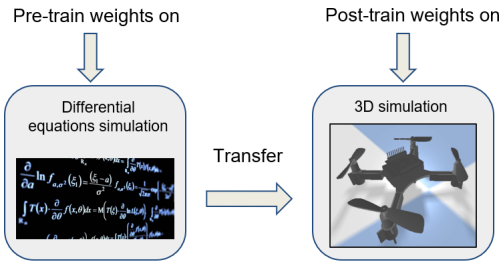


Figure: Block diagram of actor-critic approach.

Transfer Learning Approach

What knowledge do we transfer?

→ Weights of actor and critic networks.



Transfer Learning metrics

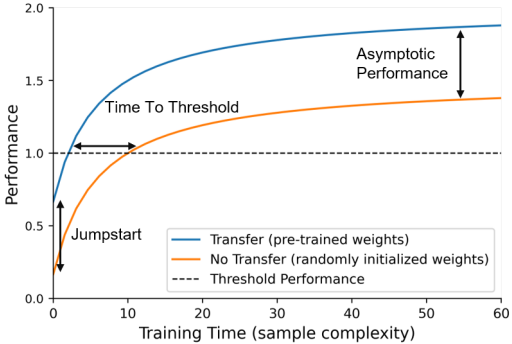
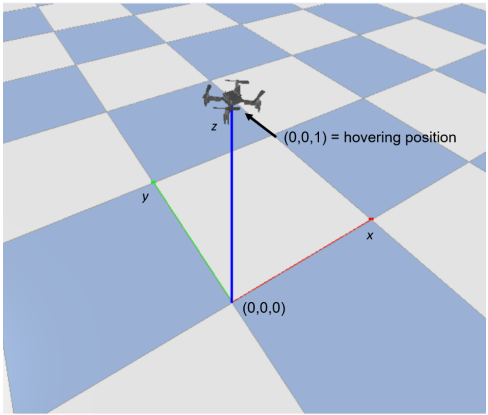


Figure: Metrics to evaluate TL benefits.

Drone Hovering Task

- No obstacles in the environment.
- Episode ends if:
  - Maximum episode length (500) is reached.
  - Constraint is violated: e.g., speed, position or orientation limit.

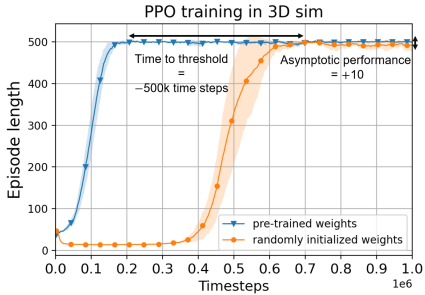


Experiment: Transfer Learning benefits

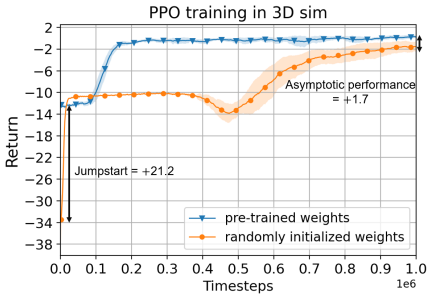
Aim: Investigate effects of TL in 3D sim:

- Initial actor and critic weights:
  - **Transfer:** initialized with pre-trained weights.
  - **No Transfer:** randomly initialized.
- **Transfer**/ **No Transfer:** train 5 instances.

Transfer Learning benefits: PPO

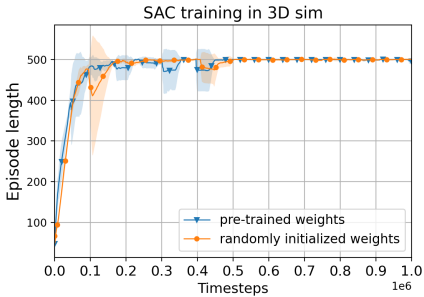


(a)

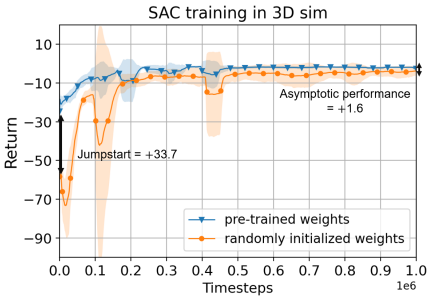


(b)

Transfer Learning benefits: SAC

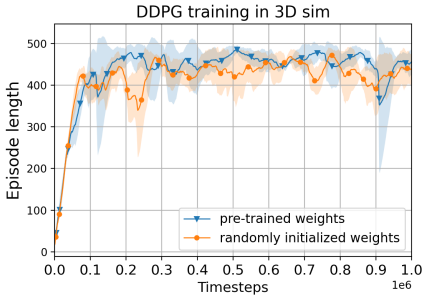


(a)

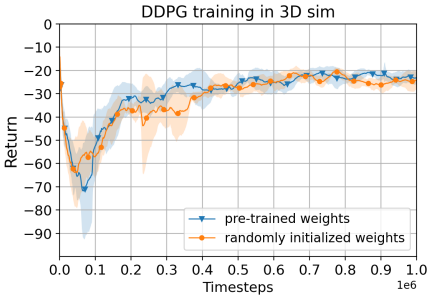


(b)

Transfer Learning benefits: DDPG

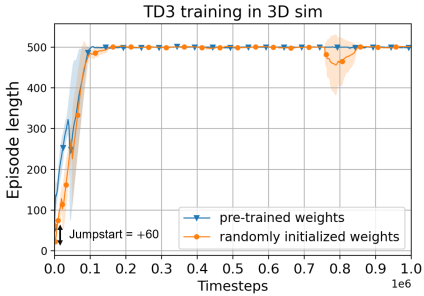


(a)

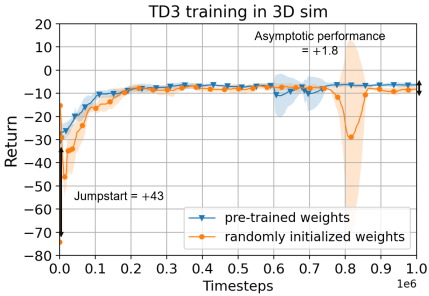


(b)

Transfer Learning benefits: TD3



(a)



(b)

Sample complexity analysis: PPO, SAC, TD3

Goal performance:

- Episode length: 500. $\rightarrow$ No constraint violation.
- Return : -2. $\rightarrow$ Magnitude of relative error $\approx 2\%$.

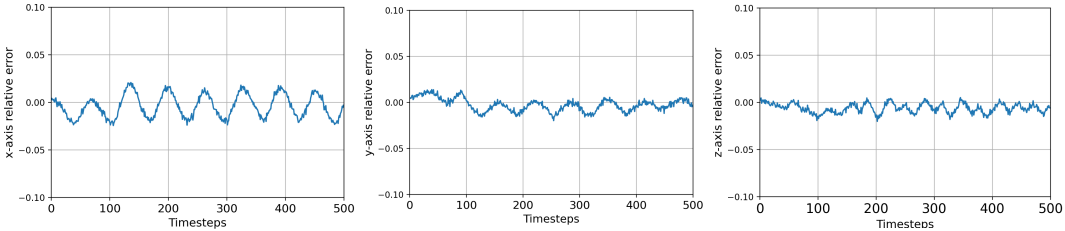
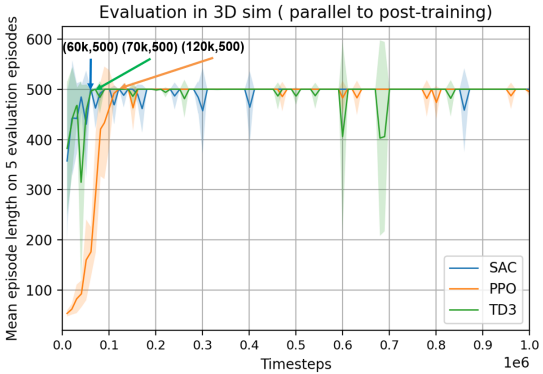


Figure: Relative error between goal position and current position (x,y,z).

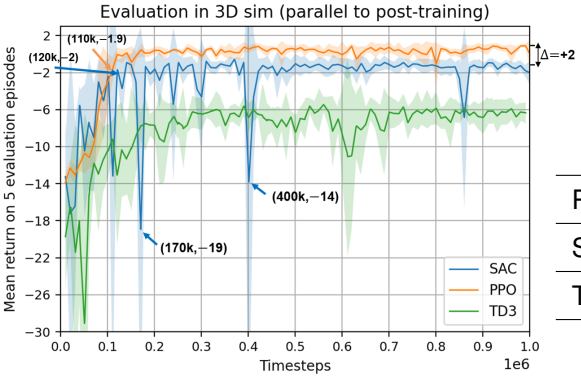
Sample complexity analysis



- Every 10k timesteps: 5 evaluation episodes.
- Performance drops $\rightarrow$ affect reliability.

	# Drops below 500	Average over drops
PPO	7	479
SAC	11	473
TD3	13	469

Sample complexity analysis



- SAC unstable: possible reason?
  $\rightarrow$ off-policy.

	# Drops below -2	Average over drops
PPO	0	-
SAC	14	-5.5
TD3	-	-

Conclusion

	Benefits from TL?	Solves the task ?	Time needed
DDPG	no	no	-
TD3	yes	no	-
SAC	yes	yes	120k
PPO	yes	yes	120k