Team 01 CIT433027

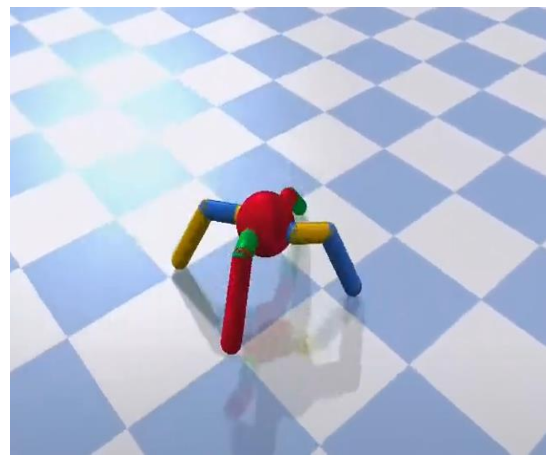
Learning What Matters: A Problem in Robotic Reinforcement Learning

Baran Ozer Serife Damla Konur

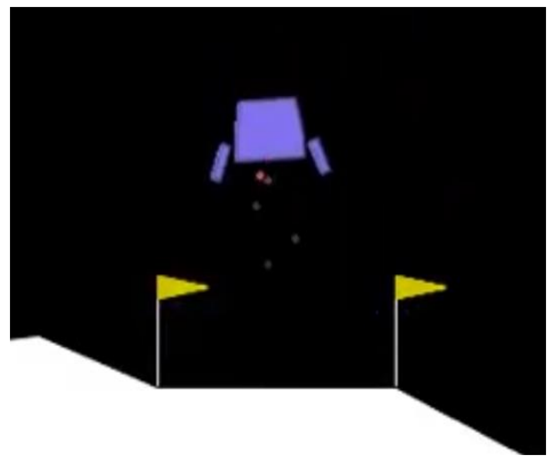
Technical University of Munich

{baran.oezer, konur.damla}@tum.de

Simulation Environments



AntBulletEnv-v0



LunarLanderContinuous-v3

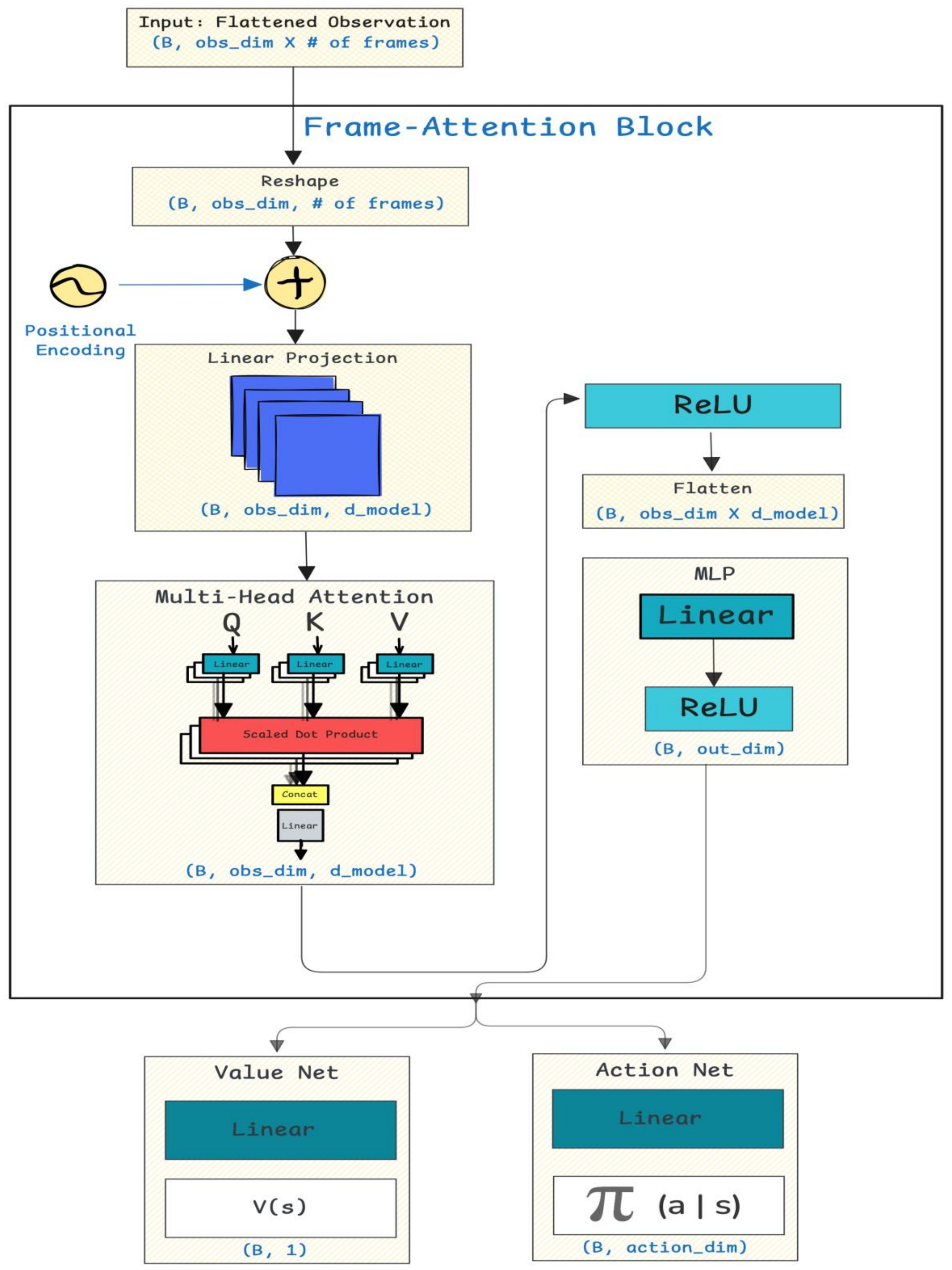
Challenge

- Equip Proximal Policy Optimization (PPO) with self-attention to ignore noise injected into observations

Curriculum

- Baseline establishment with vanilla PPO
- Obtain upper-bound performance
- Integrate frame-stacking
- Integrate self-attention mechanism
- Evaluate performance against baseline

Proposed Network Architecture



Attention Architecture Investigation

- Frame-stack size = 4 selected
- Progressively added and ablated self-attention mechanisms
  - Feature-wise
  - Bottlenecked
  - Hard-gated
  - Temporal
- Different variants replaced the default MLP in value and/or policy nets
- To see which design best preserves performence under noise

Feature-Attention

- Insert self-attention block inside every single observation frame
- Lets network learn correlations among raw features

Selective-Attention

- Insert self-attention block inside every single observation frame
- The output is squeezed through a bottleneck
- Forces network to compress useful features

Hard-Gated Attention

- Attended value overwrites the original feature
- Forces network to replace noisy dimensions instead of re-weighting

Frame-Attention

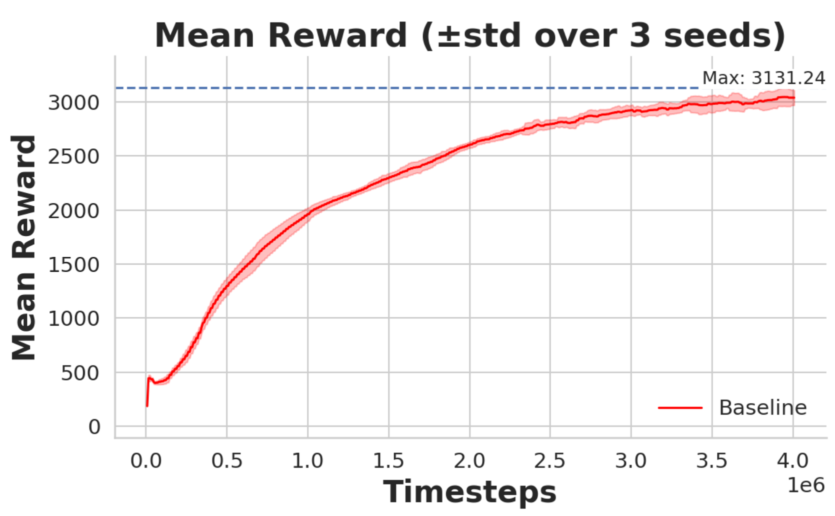
- Treats every stacked frame as a token
- Applies temporal attention across frame-stacked observations

Evaluation Against Baseline

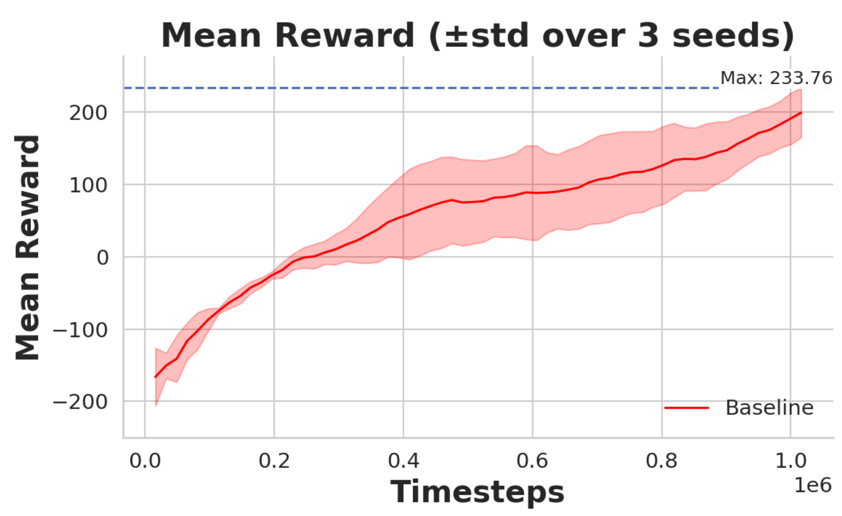
- **Policy Architecture:**
  - Frame Attention Policy
- **Observation Input:**
  - Frame-stack size = 4
- **Evaluation Protocol:**
  - 3 seeds per setting, identical PPO hyper-parameters
- **Ablations:**
  - Different noise types
  - Different number of noisy dimensions
- **Noise Injection:**
  - Ramp noise: Linearly increasing by 0.001 at each step, resetting to zero after episode termination
  - Uniform noise: Random noise sampled per feature from the range [-10, 10]
  - Gaussian noise: Gaussian noise sampled per feature with std=1

Baseline

Ant Bullet

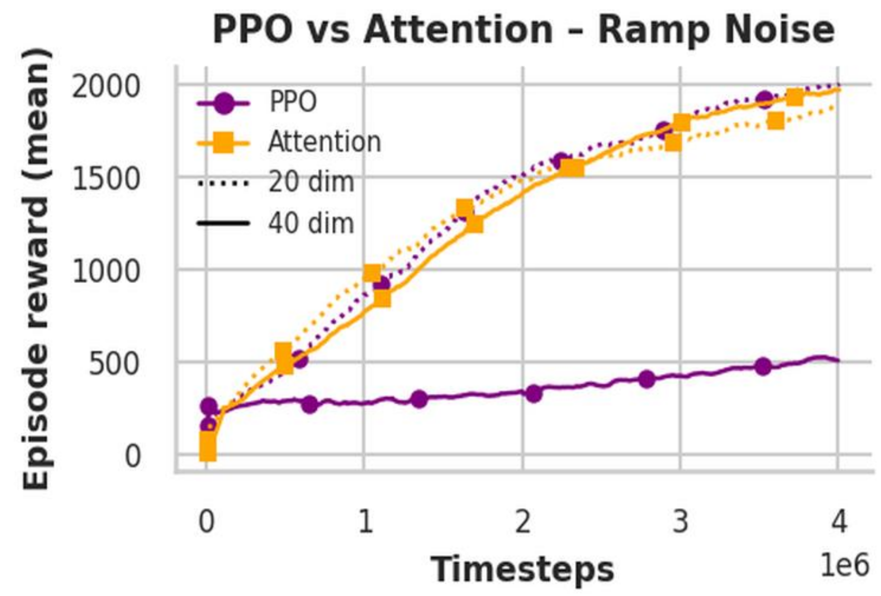


Lunar Lander

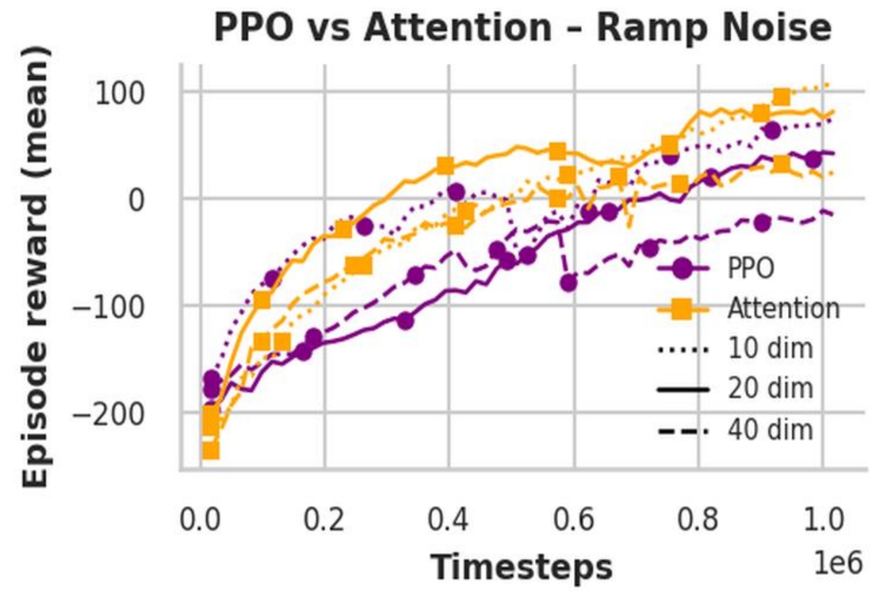


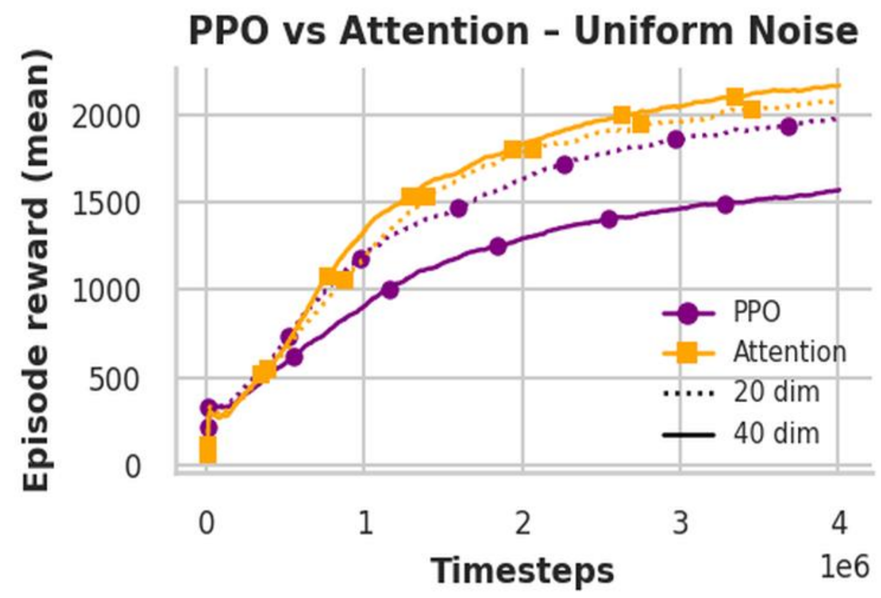
Results

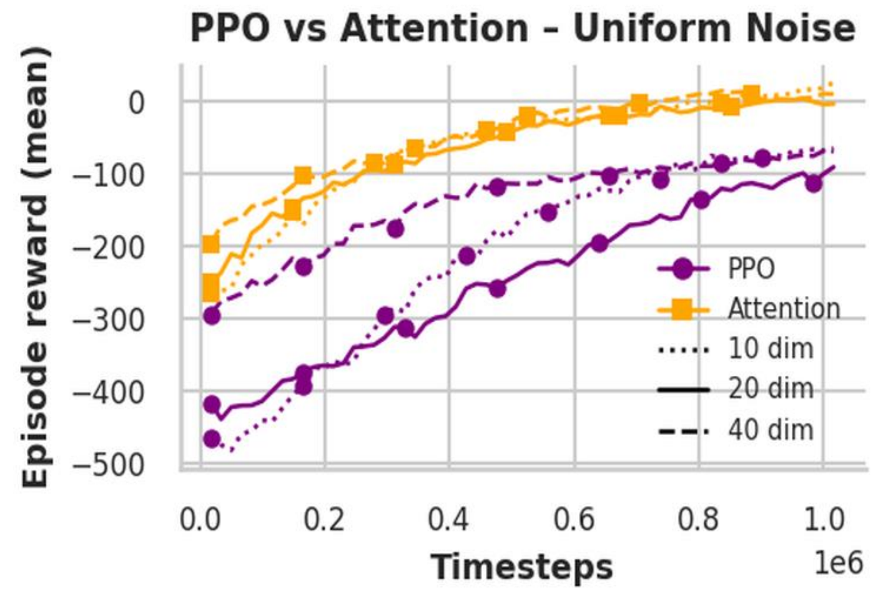
Ant Bullet

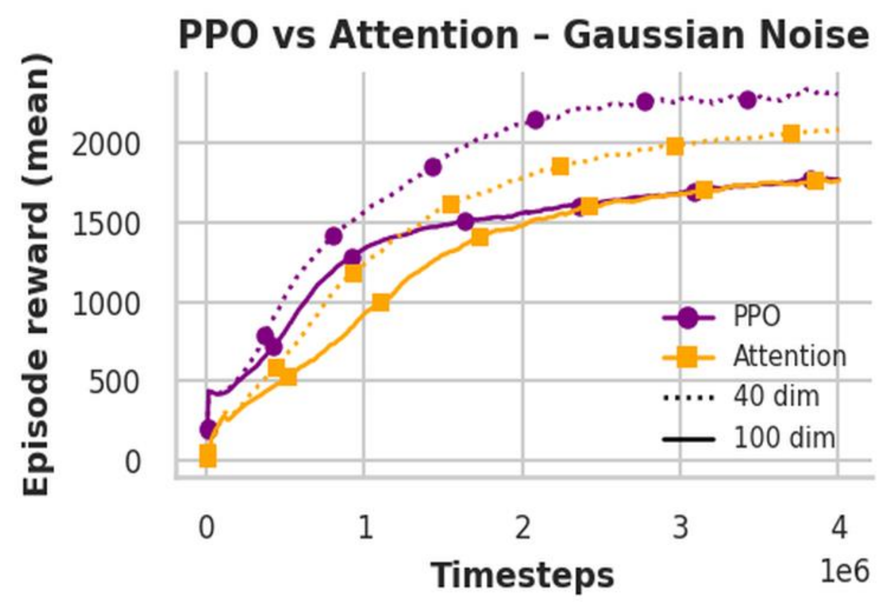


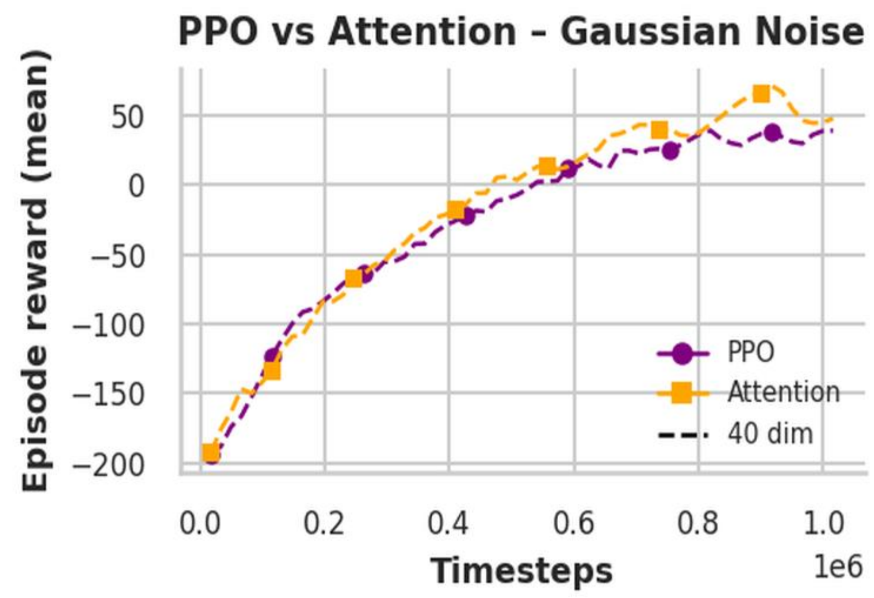
Lunar Lander











Conclusion / Success Factors

- Frame Attention improves PPO robustness under ramp and uniform noise
- Enables dynamic noise filtering
- Integrates cleanly into standard RL pipelines with minimal overhead

Future

- Improve performance under Gaussian noise
- Deploy on real robots with real-world sensor noise