Presented by : Paul Bourgois, Guillaume Delamare, Nathan Phan Tuan Linh

Learning to Jump From Pixels

Gabriel B. Margolis, Tao Chen, Kartik Paigwar, Xiang Fu, Donghyun Kim, Sangbae Kim, Pulkit Agrawal

Summary

Goal :

- Successfully navigate through **discontinuous terrain** (such as **gaps**)
- Synthesize **highly agile visually-guided locomotion behaviors**

Method : Depth-based Impulse Control (DIC)

- **High-Level** Controller : Model-Free DRL for **observation Processing** & Whole-Body **Trajectory Generator**
- **Low-Level** Controller : MPC & Whole-Body **Impulse Controller** (WBIC)

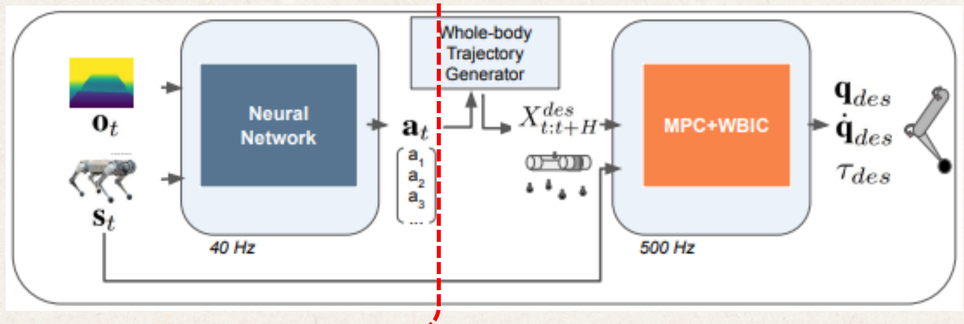
Key elements :

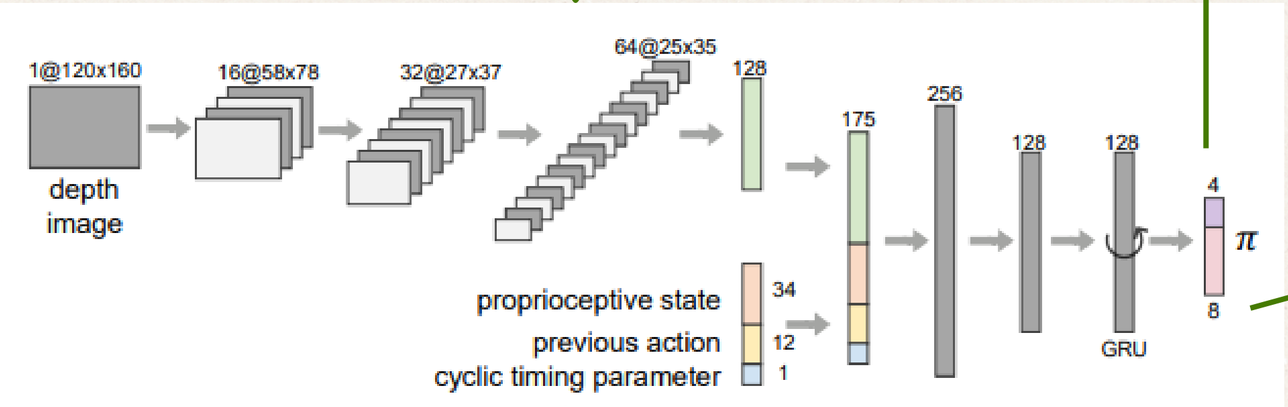
- Asymmetric-Information **Behavioral Cloning**
- Gait Adaptability
- Vision Input : Only Depth Images

High Level Control

Depth image

Proprioception





High level gait prediction network

- **Action space:**
  - 4D: $v_x, v_y, v_z, \dot{\gamma}$
  - Extra gait/contact parameters:
    - Fixed gait: none – trot/pronk schedule is hard-coded function of time.
    - Variable pronk: scalar $a_t^c \in [0,1]$ → thresholded to all-stance vs all-flight.
    - Unconstrained gait: vector $a_t^c \in [0,1]^4$ → per-leg contact state (stance vs flight).

- **Whole-Body Trajectory Generator (WBTG):**
  - $a_t \rightarrow$ body trajectory + footstep locations + contact schedule $C(a_t)$

High Level Control

Training: Episodes and Reward

Episode initialization:

- Start in standing pose on flat ground.
- Gap positions and widths are randomized in each episode.

Termination conditions:

- Body height < 0.20 m.
- |roll| or |pitch| > 40 degrees.
- Any foot steps inside a gap.
- Max length: 500 steps, corresponding to 25 s simulated time.

Reward:

$$r_t = c_1(p^b_{t,x} - p^b_{t-1,x}) - c_2 max(0, \| v^b_t \|^2 - V_{thresh}) - c_3 \mid \alpha^b_t \mid - c_4 \mid \beta^b_t \mid - c_5 \mid \gamma^b_t \mid - c_6 \| \dot{q}_t \|$$

Forward Progress | Soft Speed Limit | Stability | Smoothness

Optimization:

- PPO + Adam, learning rate: 3×10^{-4},
- batch 25632 parallel environments
- ~60h simulation time; ~12 h wall-clock

High Level Control

Asymmetric-Information Behavioral Cloning: Heightmap to Depth

Challenges:

- **Partial observability**
- **Sensory variance**
- **Depth map rendering efficiency**

Stage 1 – Expert policy

- **Perfect heightmaps:** ✅full observability ✅pose-independent terrain.

Stage 2 – Student policy - realistic perception: DAgger-style Behavioral Cloning:

- **Roll out student in simulation**
- **At each visited state , Query expert action distribution**
- **Minimize KL divergence**

Low Level Control

Model Predictive Control : 40Hz

Whole-Body Impulse Control : 500Hz

PD Controller : 40kHz

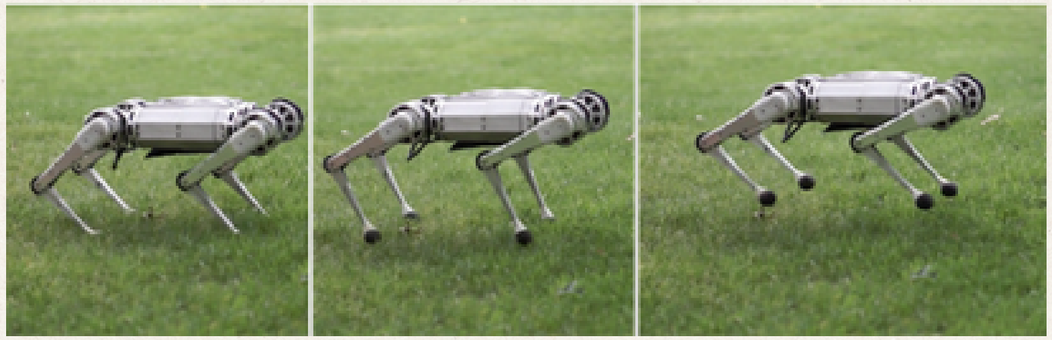
$X_{t,t+H}^{des}$
X_t

f_{des}
X_t^{des}
X_t

q_{des}
$\dot{q}_{des}$
τ_{des}

τ_{des}

Ground force reaction & Impulse



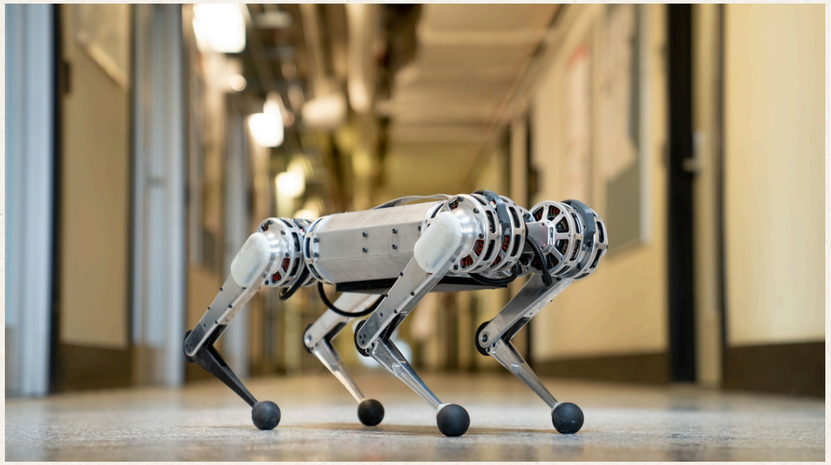
Experimental Setup

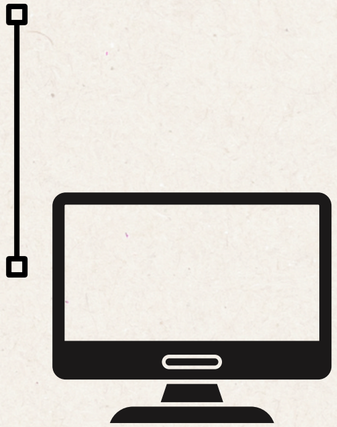
MIT Mini Cheetah

9kg, 28cm tall, 38cm long

+

Intel RealSense D435



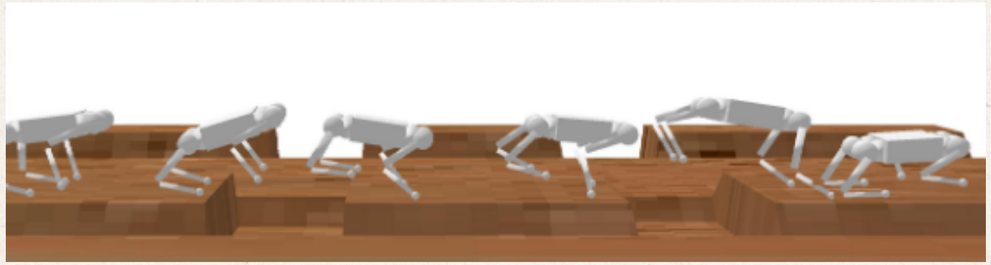


High Policy

Gap World Environment

Width between 4-30 cm

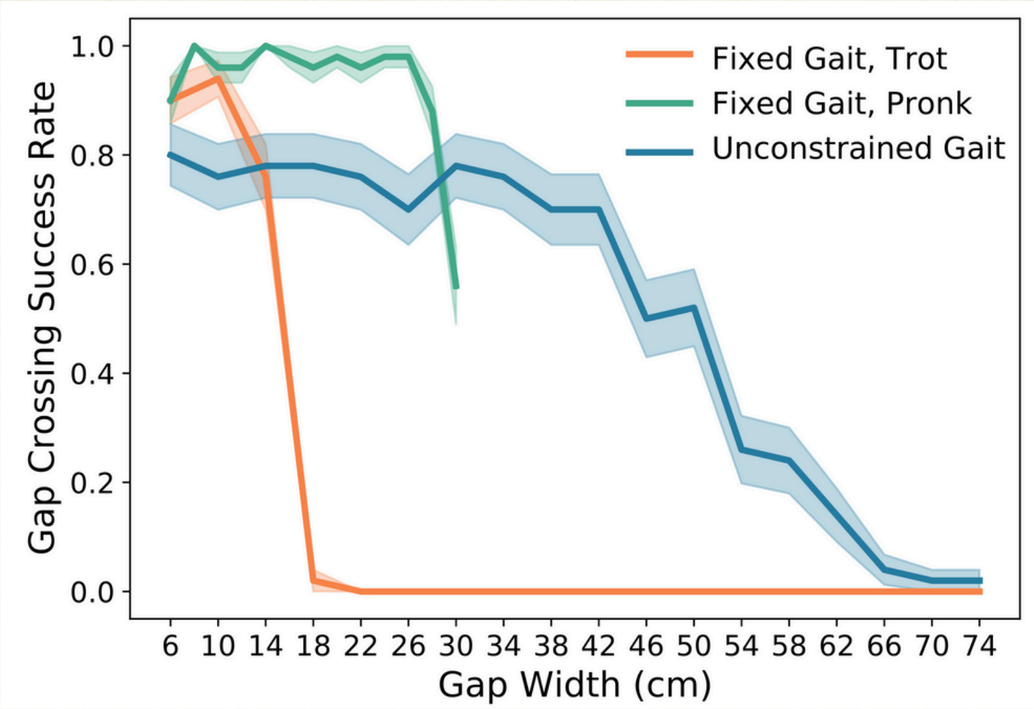
Space between gaps : 0.5-2 m



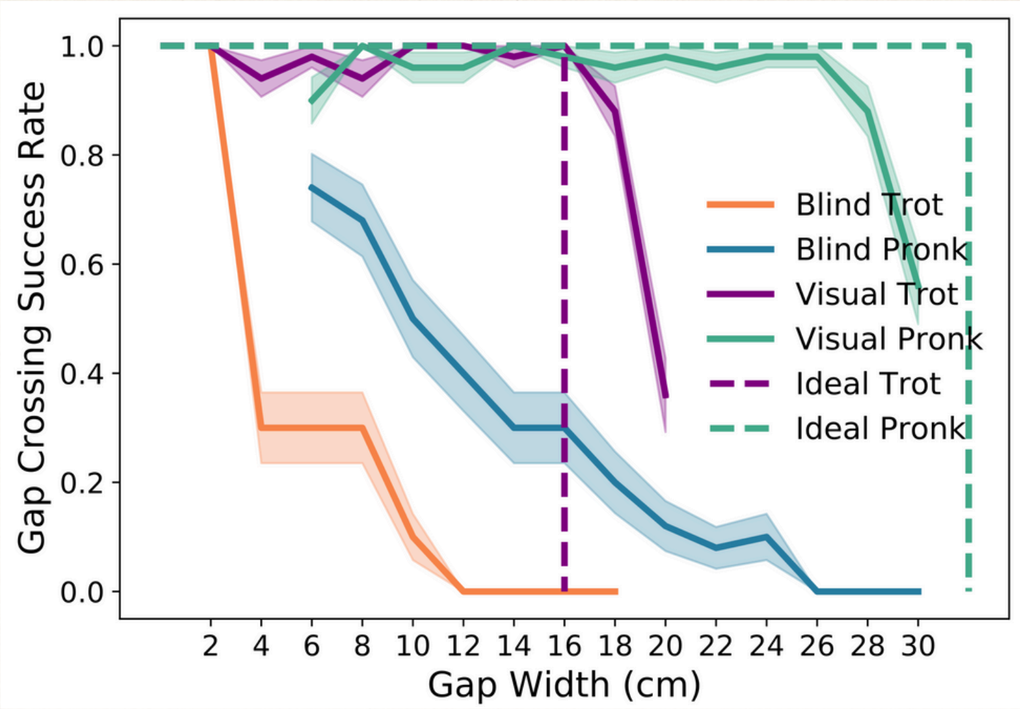
Results - Simulation

Fixed gaits (pronk & trot):

- **90+% success rate** for gaps up to **16cm wide (trot)** and **26cm (pronk)**



- **outperforms blind locomotion**
- **reaches theoretical limit** for gap size (Raibert heuristics)



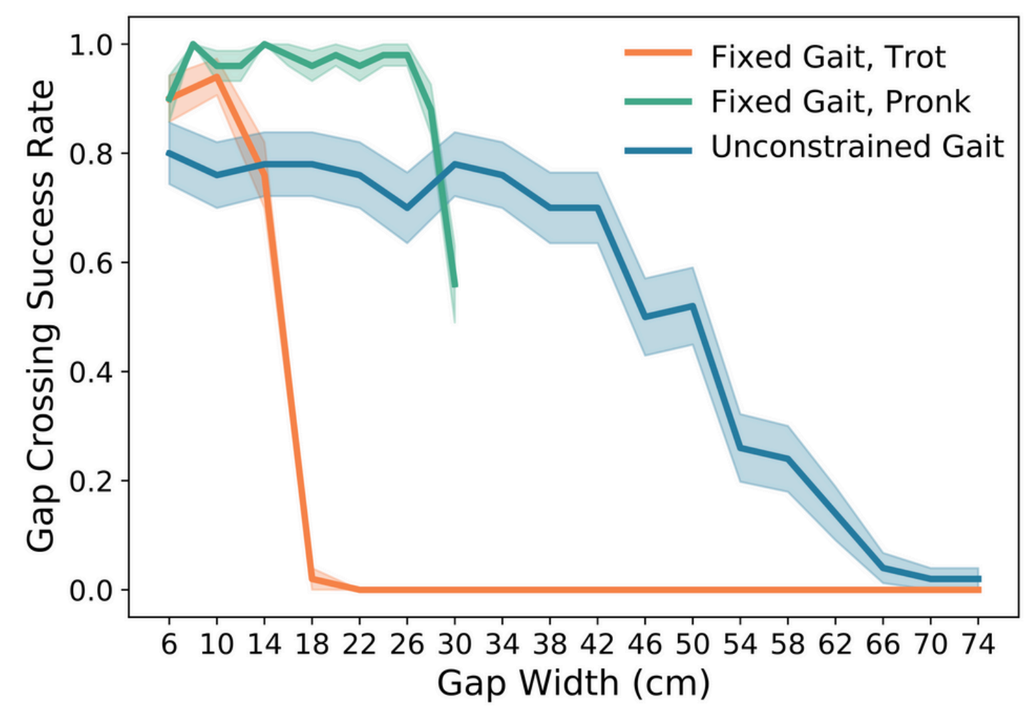
Ease of training:

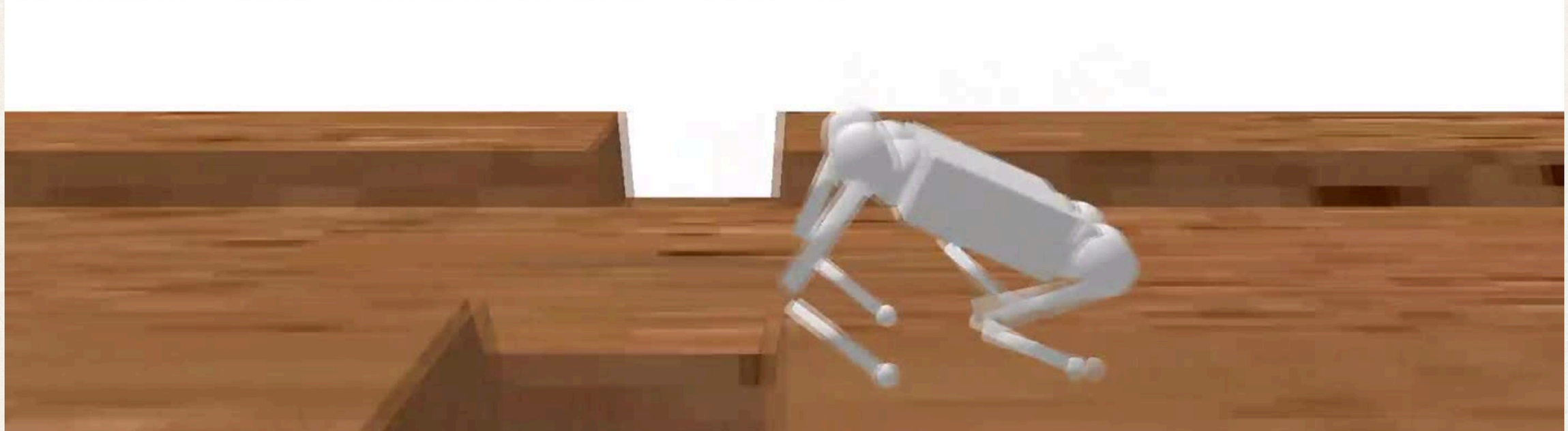
- **Single reward function** versus heavy reward tuning for model-based baseline (PMTG)

Results - Simulation

Unconstrained gait (vision-adaptive contact schedule)

- **less consistent** than fixed gait but **bigger gaps** crossed
- **emergence** of variable gaits
  - **<40cm gaps : variable-timing pronking gait**
  - **40-70cm gaps : bounding-like contact schedule** (60 cm gap in video)

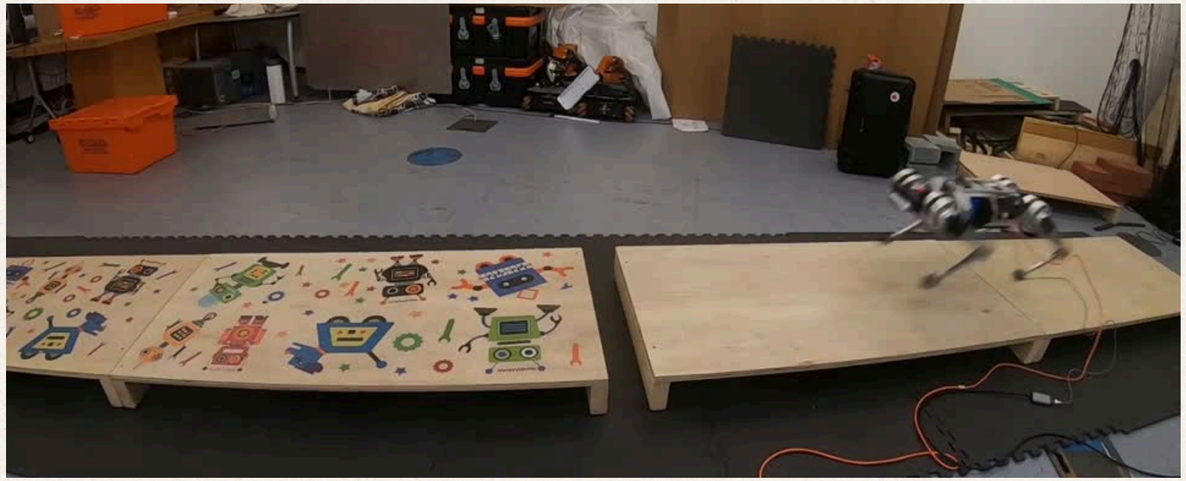




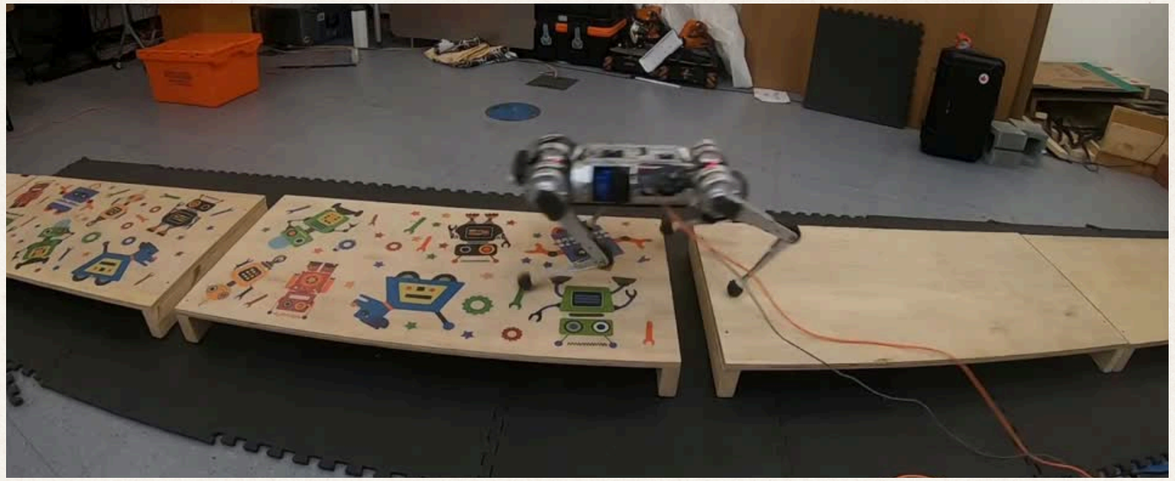
Results - Real world

Real time deployment:

- Onboard state estimation + depth images
  - **Max 16 cm** gaps



15 cm gap (pronking)



10 cm gap (trotting)

- **Sim-to-real gap**
  - **State estimation drift** due to sensor noise and imprecise contact timing
  - Low-level controller no-slip **assumtion fails**

- Ground-truth state info (motion capture) + terrain heightmap
  - **Max 26 cm** gaps

Related Work : 102 citations

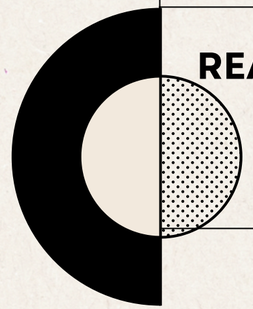
Walking with Terrain Reconstruction: Learning to Traverse Risky Sparse Footholds	• Manual design to adapt agile locomotion to **specific scenarios** • Trade-off between **model accuracy** and computational cost
Rapid Locomotion via Reinforcement Learning	• Teacher-Student Learning Policy : **Behavioral Cloning** : Train on non-observable data
Terrain-Aware Quadrupedal Locomotion via Reinforcement Learning	• **Knowledge of significant prior data** (Model, dynamic, manual tuning)
Legged Locomotion in Challenging Terrains using Egocentric Vision	• Locomotion from Egocentric **Depth**

Pros & Cons

SUMMARY:

DIC (DEPTH-BASED IMPULSE CONTROL):

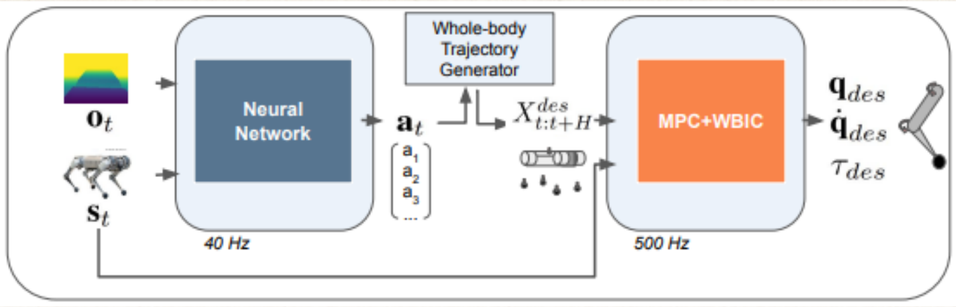
- **NEURAL NETWORK FOR HIGH-LEVEL CONTROL** (OBSERVATION PROCESSING AND WHOLE-BODY TRAJECTORY PLANNING)
- **CONVEX MPC + WHOLE-BODY IMPULSE CONTROLLER** (WBIC) **+ PD TORQUE CONTROLLER FOR LOW-LEVEL TRAJECTORY TRACKING**



PROS	CONS
HIERARCHICAL STRUCTURE WORKS WELL (minimal reward tuning)	**MULTI-STAGE TRAINING PIPELINE** (RL with priviled information then behavioral cloning)
SUPERIOR SIMULATION RESULTS THAN BASELINES (blind/model-free/model-based)	**BIG LOSS OF PERFORMANCE DURING SIM-TO-REAL TRANSFER** (imprecise state estimation & assuption violation)
REACHES THEORETICAL LIMITS FOR GAP CROSSING CAPABILITIES (Raibert heuristics)	**ONBOARD INTEGRATION CHALLENGES** (neural network running off-board due to computation load)
REAL-TIME DEPTH PERCEPTION FROM ONBOARD SENSORS (no need for external maps)	

Thanks ! Any questions ?

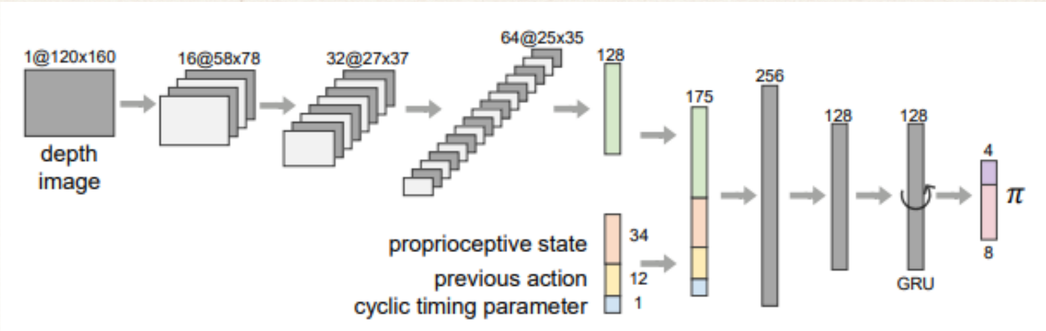
Overall Architecture



Hierarchical control

- High-level policy (NN): maps vision + proprioception → body velocity + gait schedule
- Low-level controller (MPC + WBIC): tracks desired whole-body trajectory, computes torques

High Level Control



High level gait prediction network

CNN for depth

- Depth image [160×120] → 3 conv blocks → 64 feature maps → flatten → 128-D feature vector

Feature concatenation

- 128D vision features
- 34D proprioceptive state
- 12D previous action
- 1D cyclic timing parameter (Scalar in [0,1] that advances periodically with the gait / jump cycle)
- → 175D combined feature

Temporal processing

- Fully connected layer: 175 → 256
- GRU

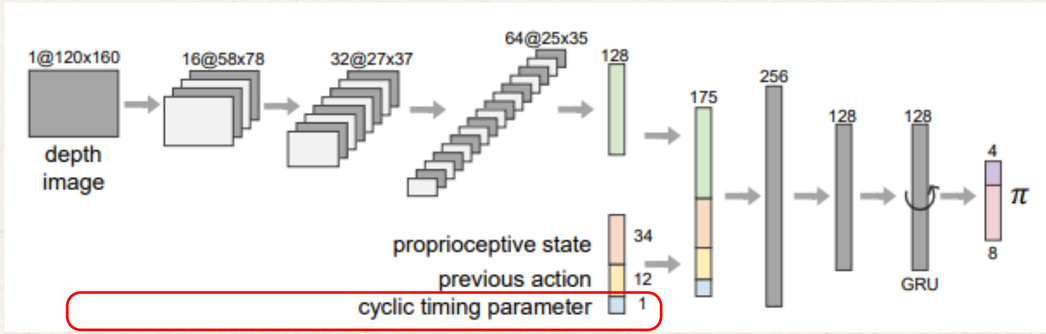
Output head

- Linear layers → final action vector (4–8D depending on gait mode)

High Level Control

Cycling timing parameter

- Scalar in [0,1] that advances periodically with the gait / jump cycle
- Concatenated with other inputs before the FC+GRU stack
- Role (inspired by Policies Modulating Trajectory Generator):
  - Acts as a global “clock” for the NN
  - Helps the policy distinguish:
    - “prepare” phase (crouch)
    - “push-off” phase
    - “flight” phase
    - “landing / recovery” phase
  - Makes it easier to learn consistent, rhythmic decisions for contact scheduling and velocity modulation

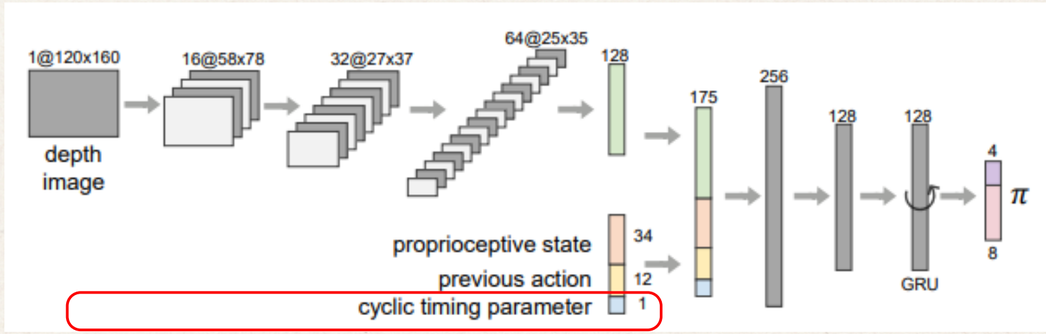


High level gait prediction network

High Level Control

Cycling timing parameter

- Scalar in [0,1] that advances periodically with the gait / jump cycle
- Concatenated with other inputs before the FC+GRU stack
- Role (inspired by Policies Modulating Trajectory Generator):
  - Acts as a global “clock” for the NN
  - Helps the policy distinguish:
    - “prepare” phase (crouch)
    - “push-off” phase
    - “flight” phase
    - “landing / recovery” phase
  - Makes it easier to learn consistent, rhythmic decisions for contact scheduling and velocity modulation



High level gait prediction network

Results - Training and neural network

Behavioral cloning:

- Behavioral cloning from heightmaps to depth images
  - Advantage over learning directly from depth images
  - Faster training

Recurrent architecture:

- Higher final performance than without
  - memory/generalization of unobserved terrain regions given the observation history

Input	T, 10cm	T, 20cm	P, 20cm	P, 30cm	VP, 30cm
Heightmap (MLP)	1.0	1.0	1.0	0.7	1.0
Depth Image (RNN)	0.6	0.3	0.9	**0.9**	0.7
Heightmap (MLP) $\rightarrow$ *Depth Image (MLP)*	1.0	0.9	0.1	0.0	0.0
Heightmap (MLP) $\rightarrow$ *Depth Image (RNN)*	**1.0**	**1.0**	**1.0**	0.4	**1.0**

Gap crossing success rate for RL policies (Trotting (T), Pronking (P), or Variable Pronking (VP)) trained on various maximum gap widths