TEXAS
The University of Texas at Austin

WHAT STARTS HERE CHANGES THE WORLD

DEEP REINFORCEMENT LEARNING FOR ROBOTICS:

A SURVEY OF REAL-WORLD SUCCESSES

Chen Tang*, Ben Abbatematteo*, Jiaheng Hu*, Rohan Chandra, Roberto Martín-Martín, **Peter Stone**

TO APPEAR IN ANNUAL REVIEW OF CONTROL, ROBOTICS, AND AUTONOMOUS SYSTEMS

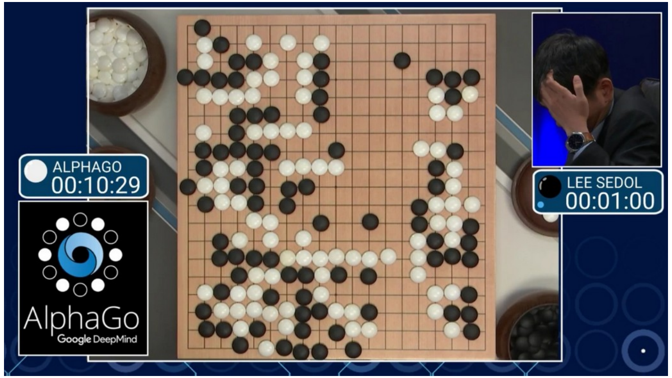
LARG

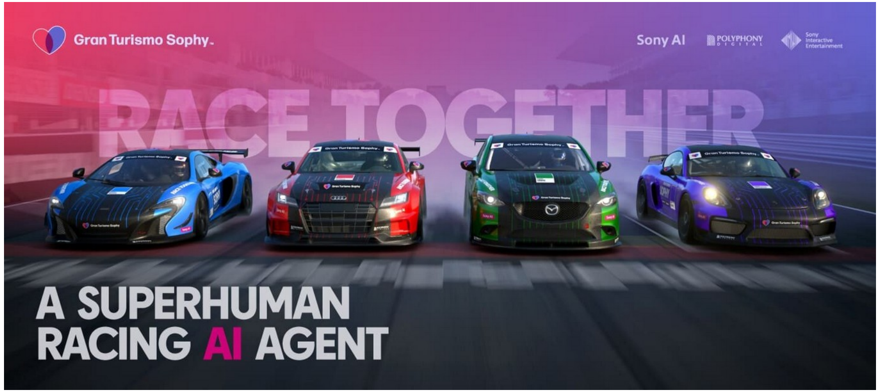
TEXAS Robotics

RobIn
ROBOT INTERACTIVE INTELLIGENCE LAB

Deep RL's major successes

- Board games, video games, recommendation systems, ...
- Suggest DRL's potential for controlling robotic systems
- But, real world robotics presents significant challenges





Deep Reinforcement Learning in Robotics

A Survey of Real-World Successes

- Goal: comprehensive evaluation of progress of DRL in real-world robotics
- Categorize papers based on
  - Robotic Competency
  - Problem formulation
  - Solution method
  - Level of real-world success
- Assess maturity across domains
- Identify general trends and key open challenges

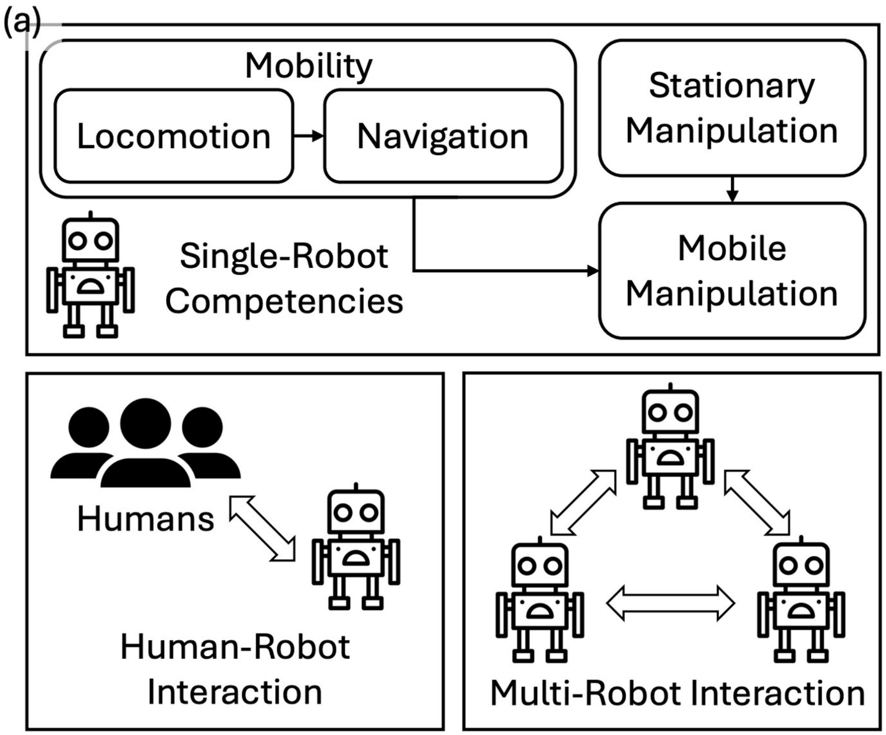
Taxonomy

- Level of Real-World Success
- Inspired by “Technology Readiness Levels”

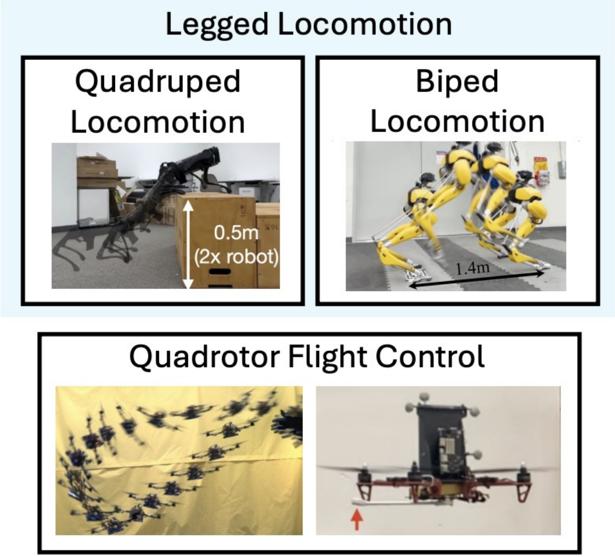
Level 5	Deployed on commercialized products
Level 4	Validated under diverse real-world conditions
Level 3	Validated under confined real-world conditions
Level 2	Validated under diverse lab conditions
Level 1	Validated under limited lab conditions
Level 0	Validated only in simulation environments

Taxonomy

- Competencies surveyed:
  - Mobility
  - Manipulation
  - Interaction with other agents



Locomotion



Quadruped	28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54
Biped	27, 55, 56, 57, 58, 59, 60, 61, 62, 63
Flight	64, 65, 66, 67, 68

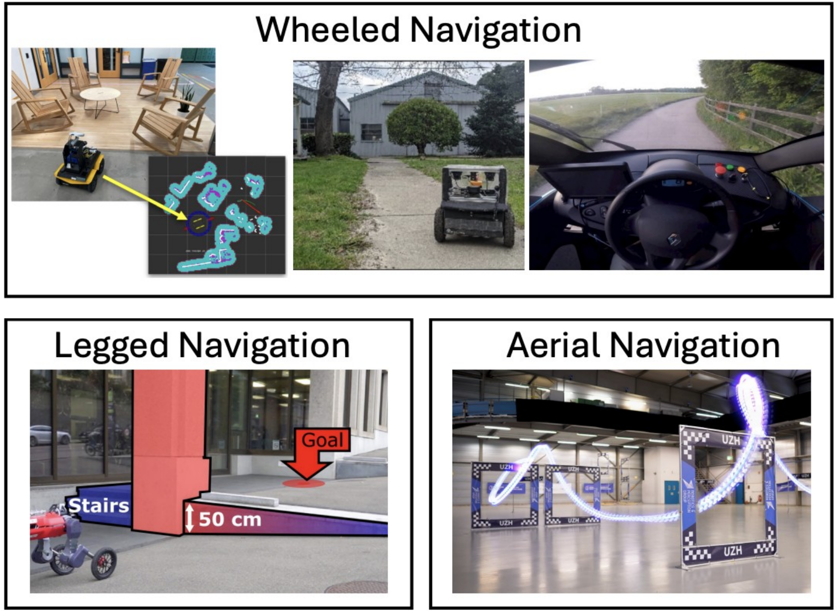
Limited Lab, *Diverse Lab*, *Limited Real*, and *Diverse Real*

Locomotion

Key Takeaways

- RL has enabled mature quadrupedal locomotion
  - Bipedal, less so – dynamics are harder, higher DoF
  - Hardware accessibility matters
- Lots of zero-shot sim-to-real & privileged information
- Open questions:
  - Efficient & safe real-world learning
  - Integrating locomotion with downstream tasks

Navigation



Wheeled	73, 74, 75, 76, 78, 81, 82, 85, 88, 89, 90, 91, 92, 93
Legged	20, 83, 86, 87, 94, 95, 96, 97, 98, 99, 100
Aerial	7, 21, 101, 102, 103

Limited Lab, *Diverse Lab*, *Limited Real*, and *Diverse Real*

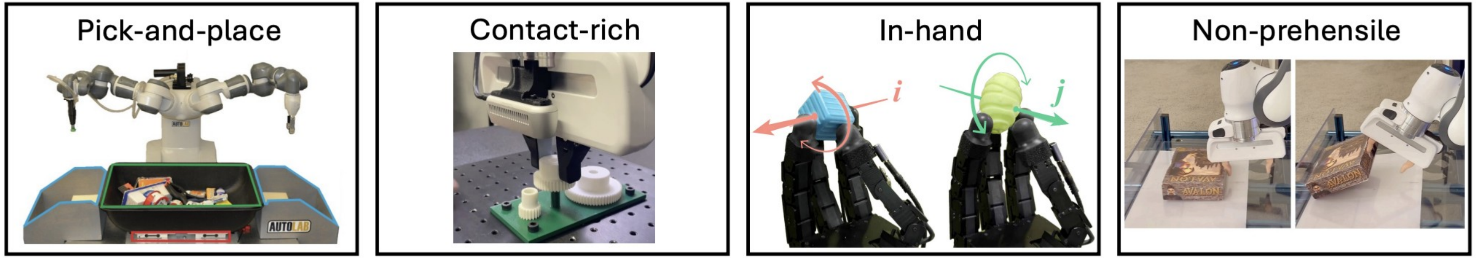
Navigation

Key Takeaways

- For indoor nav, end-to-end RL excels in simulation
- But, most successful real-world systems are modular
- *Agile* navigation: Jointly learning navigation and low-level control

- Open questions:
    - How much of the navigation stack should we learn?
    - How do we effectively jointly learn navigation & locomotion?
    - Safety critical applications (e.g., autonomous driving)

Manipulation



Pick-and-place	Grasping	108, 109, 110, 111, 112
	End-to-end Pick-and-place	54, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125
Contact-rich	Assembly	126, 127, 128, 129, 130
	Articulated Objects	122, 131, 132, 133
	Deformable Objects	134, 135, 136, 137
In-hand	—	138, 139, 140, 141, 142
Non-prehensile	—	109, 118, 143, 144, 145

Limited Lab, *Diverse Lab*, *Limited Real*, and *Diverse Real*

Manipulation

Key Takeaways

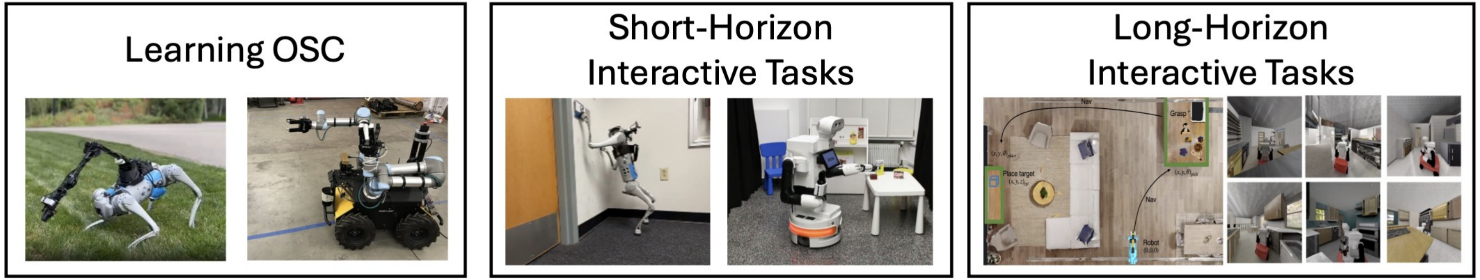
- RL is more successful when tasks are constrained, enumerable a priori
  - E.g., grasping, in-hand manipulation; cf. open-world pick-and-place
  - Allows for zero-shot sim-to-real & dense reward design
- Scaling to the open-world will require:
  - Scaling simulation assets & tasks
  - Multi-task, meta-, lifelong learning
  - Autonomous real-world learning (e.g, reward, resets)
  - Learning from human video
  - Leveraging demonstrations

Manipulation (cont.)

Key Takeaways

- Open questions:
  - How to integrate effective priors? Symmetry? Collision-avoidance?
  - How to put it all together?
    - Most works study one isolated subtask with specific action spaces
    - How do we integrate these abilities?

MoMa



Environment Perception & Object Interaction

Long-Horizon Reasoning & Partial Observability

WBC	152, 153, 154, 155
Short-Horizon	158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169
Long-Horizon	157, 170, 171

Limited Lab, *Diverse Lab*, *Limited Real*, and *Diverse Real*

MoMa

Key Takeaways

- Some initial successes, especially in short-horizon tasks, often sim-to-real
- Action space is critical, diverse morphologies
- Open questions:
  - Multi-tasking
  - Long-term memory
  - Safe exploration

HRI

Physical Human-Robot Interaction (pHRI)

Non-Collaborative

Collaborative

Shared Autonomy

Collaborative pHRI	173, 172, 174, 180
Non-collaborative pHRI	175, 176, 177, 178, 179
Shared Autonomy	181, 182, 183

Limited Lab, *Diverse Lab*, *Limited Real*, and *Diverse Real*

HRI

Key Takeaways

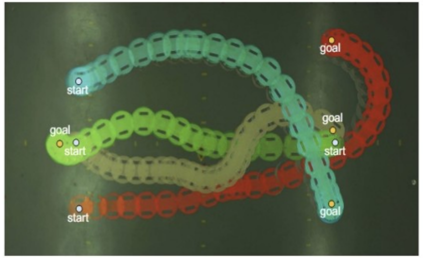
- Fewer successes than “single-robot” competencies
- Hard to collect human-like data
    - Non-Markovian
    - Limited rationality
    - Expensive

- Future directions:
    - Enable real-world learning alongside humans
    - Develop realistic human behavior simulation

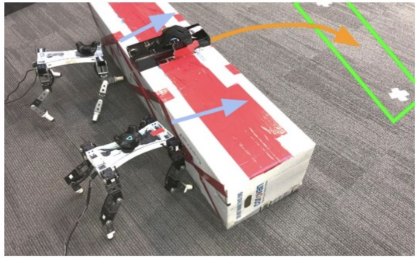
Multi-Robot

Multi-Robot Interaction Examples

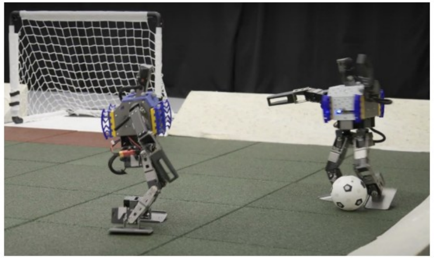
Collision Avoidance



Multi-Robot Manipulation



Robot Soccer



Multi-Robot Collision Avoidance	184, 185, 187, 188, 189
Multi-Robot Loco-Manipulation	190
Robot Soccer	191

Limited Lab, *Diverse Lab*, *Limited Real*, and *Diverse Real*

Multi-Robot

Key Takeaways

- Limited successes in cooperative “homogeneous” settings
  - E.g., collision-avoidance
- Challenges in complexity & scalability
- Critical areas:
  - Communication between agents
  - Convergence & stability
  - General, non-cooperative settings

General Trends

- Mature domains:
  - Locomotion, some navigation & manipulation
- Less mature domains:
  - MoMA, HRI, Multi-robot
- Mature solutions are commonly sim-to-real
  - E.g., Locomotion, navigation, grasping, in-hand manipulation
  - Stable, straightforward to simulate
  - Dense, engineered reward functions
  - On-policy is feasible
  - Scalability?
- Without sim, human demos can mitigate exploration challenge

Key Future Directions

- Improving stability & sample-efficiency of RL algorithms
- Real-world learning
  - Gathering data: exploration, reward design, …
  - Sample-efficiency & transfer
- Long-horizon tasks
  - What skills should the robot learn?
  - How should they be combined?

Key Future Directions (cont.)

- Principled approaches for RL systems
  - Reward design, action space choice
  - Integration with classical model-based tools
- Benchmarking real-world success
  - Need standard platforms and test problems
- Leveraging Foundation Models
  - Avenue toward stronger generalization, language-conditioning
  - Possibility for reward design, simulation task & asset creation, etc.

Application	Action Space: Low-Level	Mid-Level	High-Level
Locomotion	27, 28, 29, 30, 31*, 32*, 33, 36*, 37, 38, 40, 41, 42, 43, 44, 45, 46, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 61, 62, 63, 64, 65, 68	31*, 32*, 34, 35, 47, 66, 67	36*, 60
Navigation	20, 90, 96*, 97*, 99, 100,	7, 21, 73, 74, 75, 78, 83, 85, 88, 89, 91, 92, 93, 94, 95*, 98, 101, 102, 103	76, 81, 82, 86, 87, 95*, 96*, 97*
Manipulation	113, 122, 127, 131, 138, 139, 140, 141, 142	54, 110, 111, 112, 114, 115, 116, 117, 118, 119, 120, 121, 124, 125, 126, 128, 129, 130, 132, 133, 134, 137, 143, 144, 145	108, 109, 123, 135, 136
MoMa	153, 154, 155, 163, 167, 169, 158, 162, 159, 166	152, 164, 160, 161, 171, 157, 170, 165	168
HRI	175, 176, 177, 178, 179, 180	173, 174, 181, 182, 183	172
Multi-Robot Interaction	184, 185, 187, 188, 190, 191	189	

Table 1: Categorizing Literature based on Problem Formulation

Application	Simulator Usage		
	Zero-shot Sim-to-Real	Few-shot Sim-to-Real	No Simulator
Locomotion	27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 40, 41, 42, 44, 45, 46, 47, 49, 50, 51, 52, 55, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68	43, 48, 56	53, 54
Navigation	20, 21, 73, 74, 75, 76, 78, 81, 82, 83, 85, 86, 87, 93, 94, 95, 96, 97, 98, 99, 100, 101, 103	7, 102	88, 90, 91, 92
Manipulation	108, 111, 123, 130, 133, 134, 135, 137, 138, 139, 141, 142, 143, 144, 145	116, 131	54, 109, 110, 112, 113, 114, 115, 117, 118, 119, 120, 121, 122, 124, 125, 126, 127, 128, 129, 132, 136, 140
MoMa	153, 154, 152, 155, 163, 167, 169, 162, 164, 161, 165, 159, 166, 171, 157	158, 170	160, 168
HRI	172, 173, 174, 175, 176, 177, 179	182	178, 181, 180
Multi-Robot Interaction	184, 185, 187, 188, 189, 190, 191		

Table 3: Categorizing Literature based on Solution Approach

TEXAS
The University of Texas at Austin
WHAT STARTS HERE CHANGES THE WORLD

Deep Reinforcement Learning in Robotics

A Survey of Real-World Successes

Chen Tang*, Ben Abbatematteo*, Jiaheng Hu*, Rohan Chandra, Roberto Martín-Martín, **Peter Stone**

TO APPEAR IN ANNUAL REVIEW OF CONTROL, ROBOTICS, AND AUTONOMOUS SYSTEMS

Contact: pstone@cs.utexas.edu

LARG TEXAS Robotics RobIn ROBOT INTERACTIVE INTELLIGENCE LAB