Deep Reinforcement Learning For Robotics:

A Survey of Real-World Successes

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

2024. 09. 16 v3

로봇공학, 박지호 2025710799

Contents

History

버전	일자	작업 내역	페이지	작성자
v0.1	2025.11.10	문서 최초 작성	전체	박지호
v0.2	2025.12.01	문서 수정	.	.

Why This Survey

로봇 외 [9], 대부분 시뮬레이션 중심, 자체 연구 사례 중심[10]
특정 작업[11][12], 특정 기술 [13], [14]
딥러닝 이전[15], 시뮬레이션일 때 작성[16]

1. Real-World Success 중심 분석

- 현실 세계 성공한 DRL 연구 선별
- 성숙도(maturity) , 미해결 문제(Open Challenges) 평가

2. 새로운 DRL 분류 체계

Robot 역량 / 문제 식 형태 / Solution 접근 방식 / Success Level

3. 최신 DRL 발전 배경

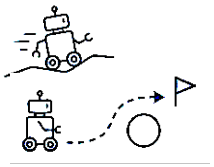
DRL 분야의 새로운 필요
simulation -> Real-world

분류

Robot 역량 / 문제 공식화 / Solution 접근 방식 / Success Level

Robot Competencies

- Mobility : 이동
  - Locomotion
  - Navigation
- Manipulation : 조작
- Interaction with other agents



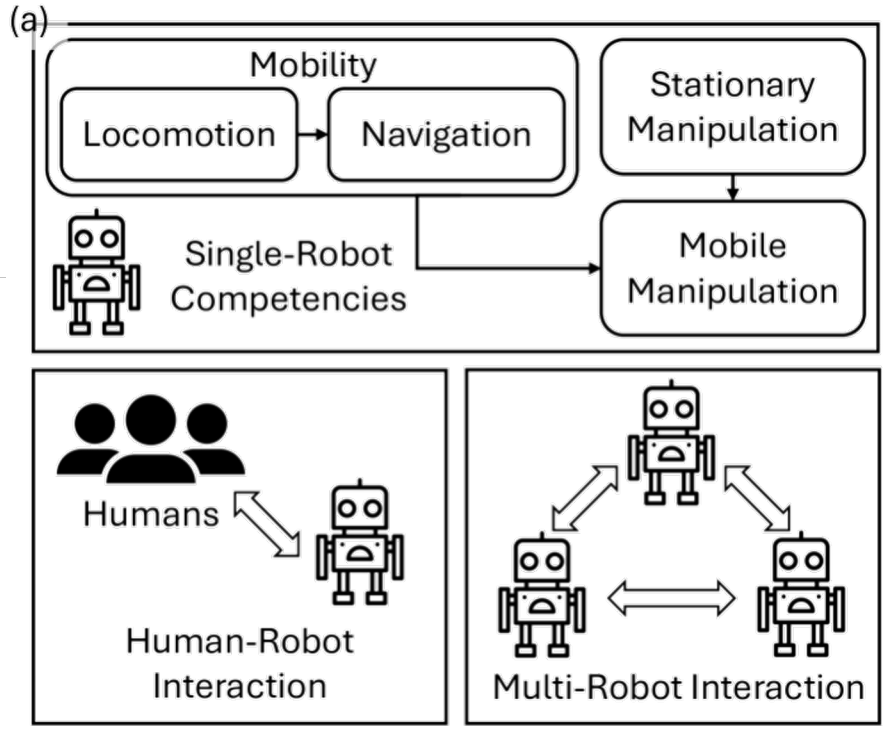


Fig 1. Robot competencies learned with DRL

분류

Robot 역량 / **문제 공식화** / **Solution 접근 방식** / Success Level

Problem Formulation

- Action Space
  - low-level (joint, motor commands)
  - mid-level (task-space commands)
  - high-level(extended to time)
- Observation space
  - high-dimensional sensor input
  - low-dimensional sensor input
- Reward function
  - sparse (reward signals)
  - dense (reward signals)

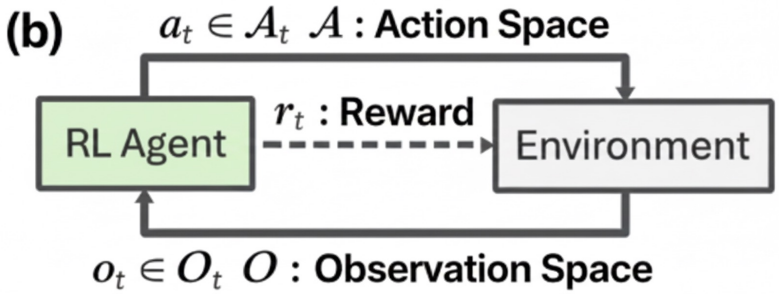


Fig 2. Problem formulation

분류

Solution Approach

1. Simulator usage
   - sim-to-real
     - zero shot
     - few shot
   - offline/real
2. Model learning
   - Model-free
   - Model-base

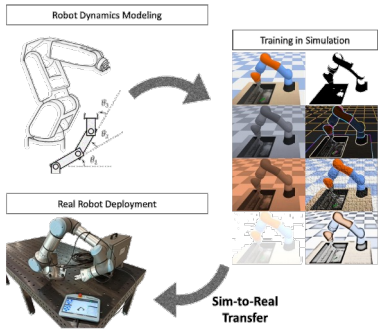


Fig 3. Conceptual view of sim-to-real

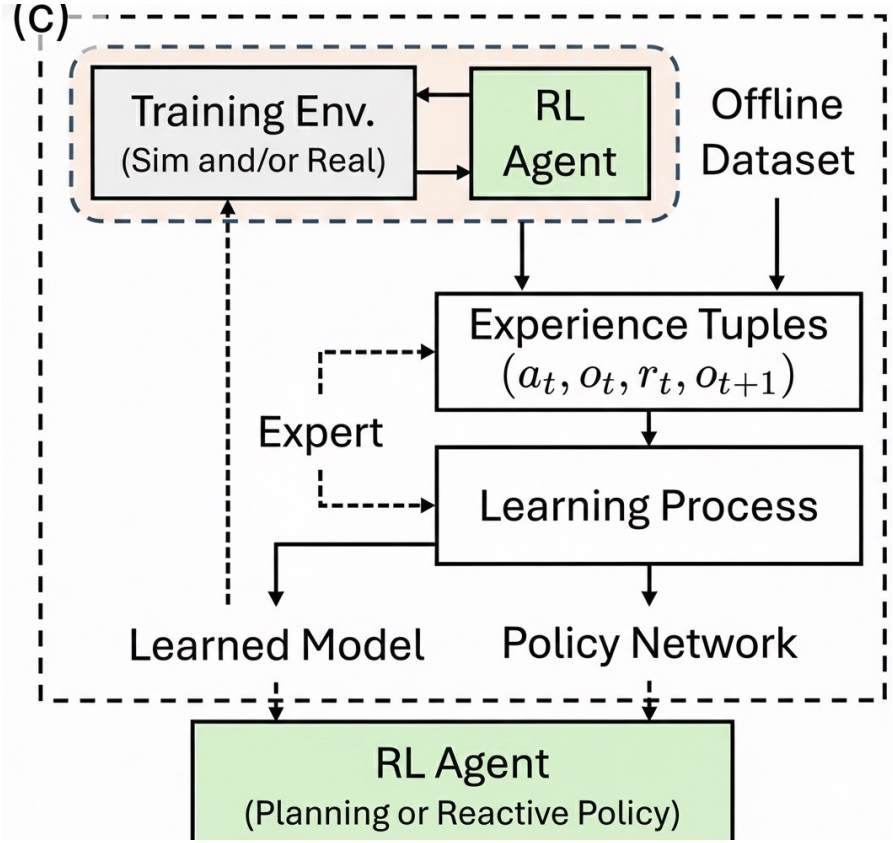


Fig 4. Solution approach

분류

Robot 역량 / 문제 공식화 / **Solution 접근 방식** / Success Level

Solution Approach

3. Expert usage
   - human demo, oracle, etc
4. Policy Optimization
   - planning
   - Offline RL
   - On-Policy RL
   - Off-Policy RL
5. Policy/Model Representation

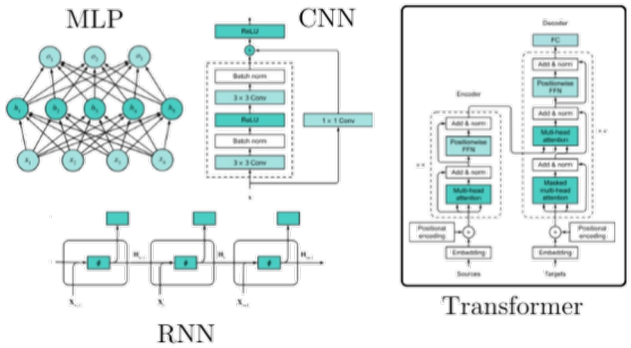


Fig 5. Popular archtectures

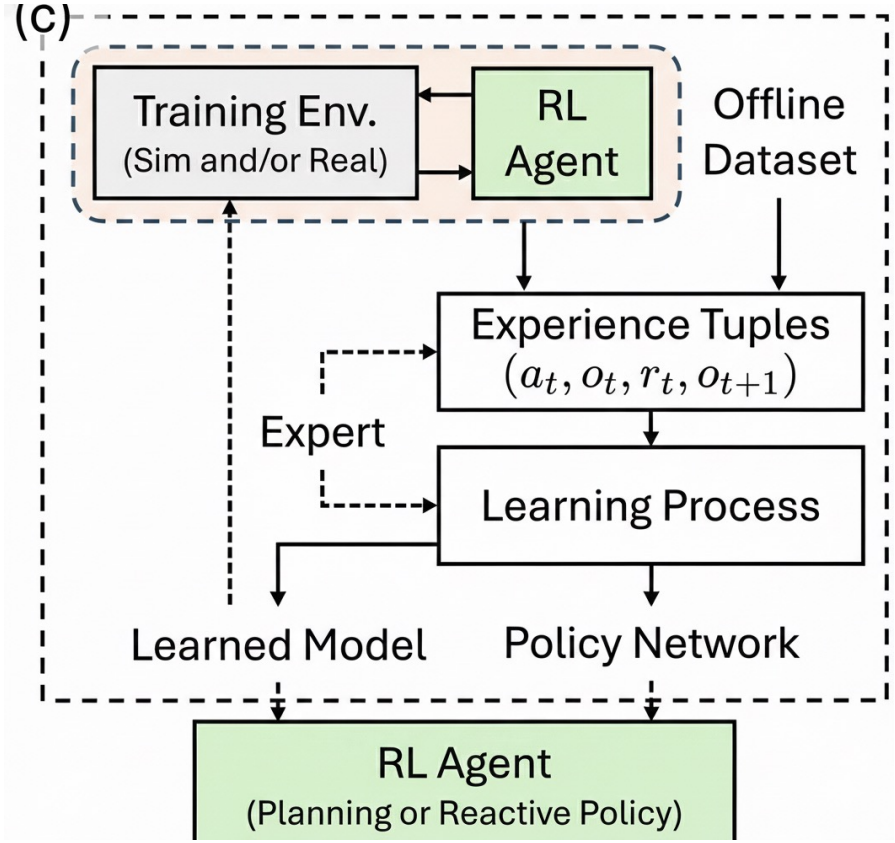


Fig 4. Solution approach

분류

Robot 역량 / 문제 공식화 / Solution 접근 방식 / **Success Level**

Real-World Success
Inspired by Technology Readiness Levels

(기술 성숙도)

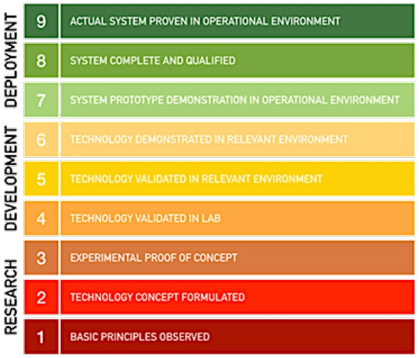


Fig 6. 기술 성숙도

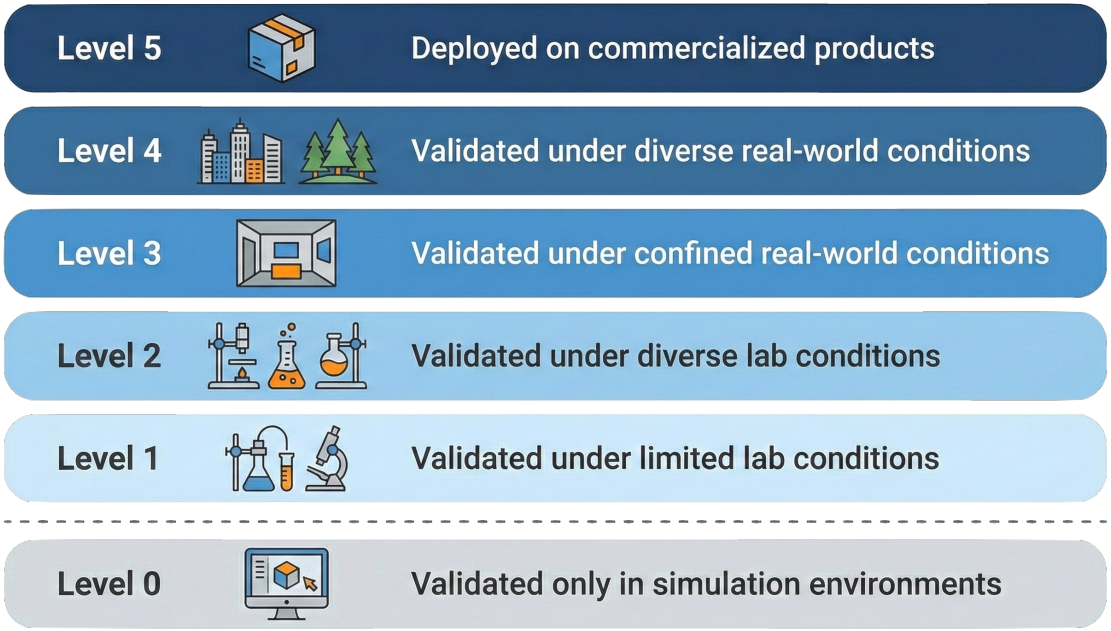


Fig 7. Real-World Success

Competency-Specific Review

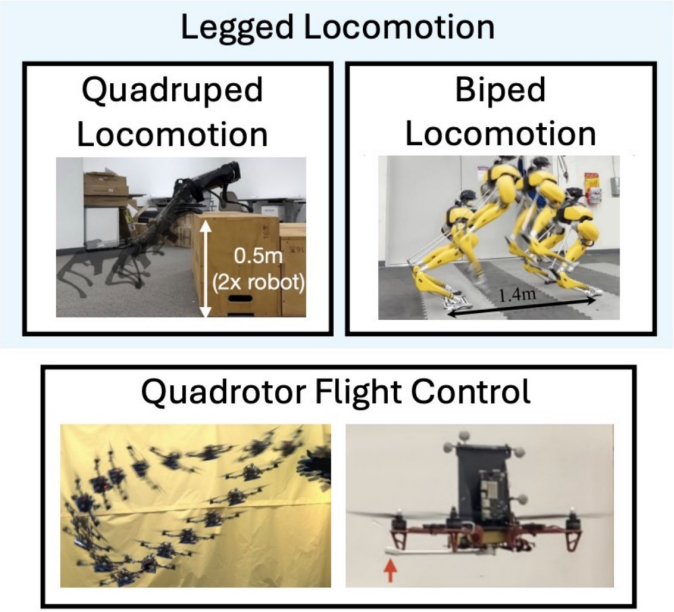
Focusing on a specific robot competency

- Locomotion
- Navigation
- Manipulation
- MoMa
- HRI
- Multi-Robot

Color legend
- Limited Lab : 제한된 실험실 환경
- Diverse Lab : 다양한 실험실 환경
- Limited Real : 제한된 실제 세계
- Diverse Real : 다양한 실제 환경

Limited Lab *Diverse Lab* *Limited Real* *Diverse Real*

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

Fig 8. Locomotions

Limited Lab *Diverse Lab* *Limited Real* *Diverse Real*

Locomotion

Key Takeaways

- **DRL 통한 quadrupedal locomotion의 구현**
  - Less Mature : 이족 보행 (DoF 높음, 동역학의 어려움)
- **다수의 zero-shot sim-to-real & privileged information**
  - Zero-shot, Sim-to-real : On-policy Model-Free
  - Privileged information
    - 특권 정보 가진 정책 훈련 후 Teacher-Student 증류

- **<u>Open questions:</u>**
  - 효율적, 안전한 real-world 학습
  - 이동과 다른 작업의 통합 (고차원, 복합적, 장기 목표)

Fig 9. Boston Dynamics, spot

Navigation

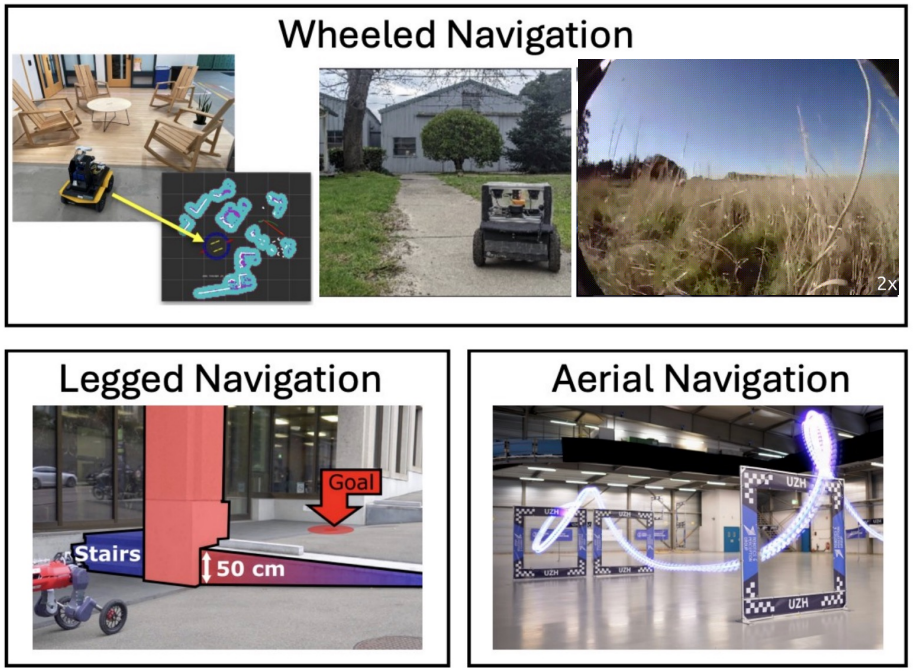


Fig 11. Navigation

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* *Diverse Real*

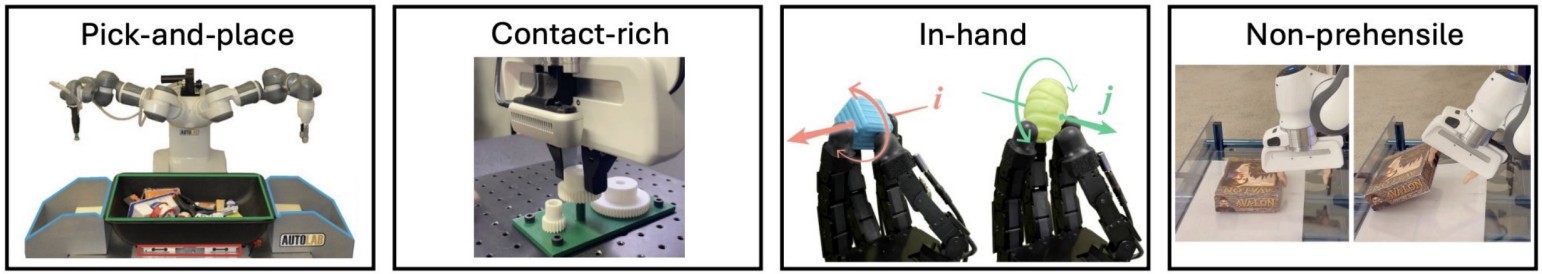
Navigation

Key Takeaways

- **실내 Nav 경우, end-to-end RL이 시뮬레이션 상 뛰어남**
  - Less Mature : 이족 보행 (DoF 높음, 동역학의 어려움)
- **real-world 경우, 모듈식이 가장 성공적**
  - 일반화, 설명 가능성, 안전성의 부재
  - 대부분 상용 시스템(classical stacks)
    - 유망한 접근 : local plan, semantic exploration
- **Agile Navigation**
  - Joint learning navigation & low-level control

- **<u>Open questions:</u>**
  - Nav stacks 중 얼마나 많은 부분을 학습으로 대체 해야?
  - Nav & Locomotion을 어떻게 함께 효과적으로 학습?
  - Safety Critical 분야 (e.g. 자율주행)

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* *Diverse Real*

Manipulation

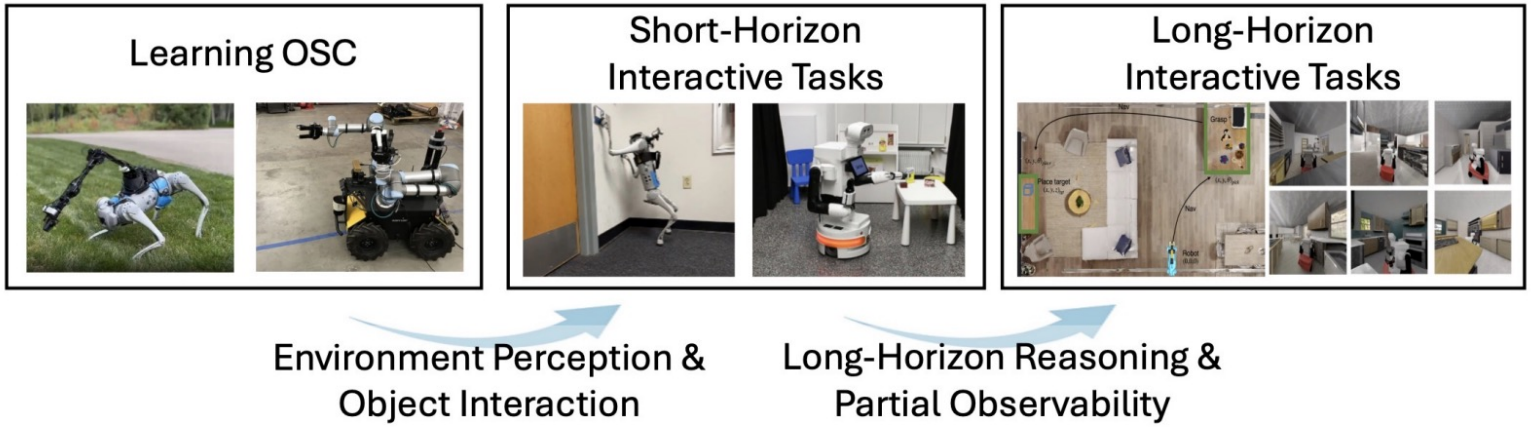
Key Takeaways

- **RL is more successful when task is Constrained, enumerable a priori**
  - Constrained(제약) : 정해진 물품, 정해진 환경 ..
  - Enumerable a priori (사전 열거 가능) : task/target/초기조건
  - grasping, in-hand manipulation
  - Allows 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

- **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?

Mobile Manipulation



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* *Diverse Real*

Mobile Manipulation

Key Takeaways

- **Some initial successes, especially in short-horizon tasks, often sim-to-real**
- **Action space is critical, diverse morphologies**

- **<u>Open questions:</u>**
  - Multi-tasking
  - Long-term memory
  - Safe exploration

Human-Robot Interaction

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* *Diverse Real*

Human-Robot Interaction

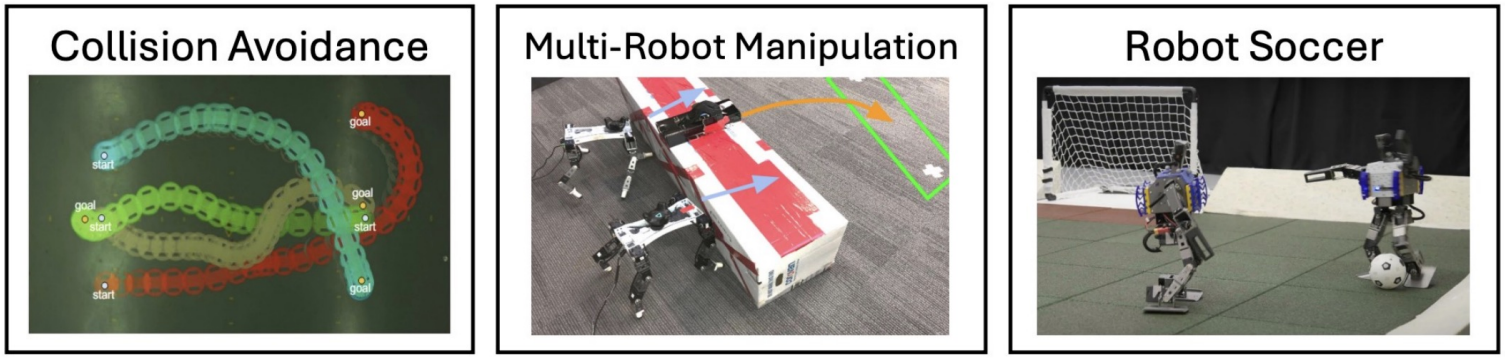
Key Takeaways

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

- **<u>Future directions:</u>**
  - Enable real-world learning alongside humans
  - Develop realistic human behavior simulation

Multi-Robot Interaction

Multi-Robot Interaction Examples



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

Limited Lab *Diverse Lab* *Limited Real* *Diverse Real*

Multi-Robot Interaction

Key Takeaways

- **Limited successes in cooperative "homogeneous" settings**
  - E.g, collision-avoidance
- **Challenges in complexity & scalability**

- **<u>Critical areas:</u>**
  - 통신 : Communication between agents
  - 수렴성 : Convergence & stability
  - 비협조, 일반. 상황 : General, non-cooperative setting

General Trends

Key Takeaways

- **Real-World Success**
  - Mature domains
    - Locomotion
    - Some (Navigation, Manipulation)
      - grasping, assembly, in-hand, non-prehensile
- **Less mature domains**
  - MoMA, HRI, Multi-robot,
  - Some(Nav, Mani)
    - pick and place
- **Mature solutions : commonly Zero-Shot Sim-to-real**
  - Domain : Locomotion, Navigation, (grasping, in-hand) manipulation
  - Dense, engineered reward functions
  - On-policy is feasible
- **Without sim, human demos can mitigate exploration challenge**

Key Future Directions

Key Takeaways

- **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 strong generalization, language-conditioning
  - Possibility for reward design, simulation task & asset creation, etc

Additional Table

Problem Formulation

➢ **Table 1 : Categorizing Literature base on Problem Formulation**

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*

Additional Table

Problem Formulation

➢ **Table 1 : Categorizing Literature base on Problem Formulation**

Application	Action Space: Low-Level	Action Space: Mid-Level	Action Space: High-Level
Manipulation	113, 122, 127, 131, 138, 139, 140, 141, 142, 143, 144	54, 110, 111, 112, 114, 115, 116, 117, 118, 119, 120, 121, 124, 125, 126, 128, 129, 130, 132, 133, 134, 137, 145, 146, 147	108, 109, 123, 135, 136
MoMa	155, 156, 157, 165, 169, 171, 160, 164, 161, 168	154, 166, 162, 163, 173, 159, 172, 167	170
HRI	177, 178, 179, 180, 181, 182	175, 176, 183, 184, 185	174
Multi-Robot Interaction	186, 187, 189, 190, 192, 193	191	

Additional Table

Problem Formulation

➢ **Table 2 : Categorizing Literature base on Problem Formulation**

Application	Observation Space		Reward Function	
	High-dim	Low-dim	Sparse	Dense
Locomotion	35, 36, 43, 44, 45, 49, 50, 54, 61	27, 28, 29, 30, 31, 32, 33, 34, 37, 38, 40, 41, 42, 46, 47, 48, 51, 52, 53, 55, 56, 57, 58, 59, 60, 62, 63, 64, 65, 66, 67, 68	56	27, 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, 55, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68
Navigation	73, 74, 75, 76, 78, 81, 82, 83, 85, 86, 87, 88, 89, 91, 92, 94, 95, 96, 97, 98, 99, 101, 102	7, 20, 21, 90, 93, 100, 103	78, 96 *	7, 20, 21, 73, 74, 75, 76, 81, 82, 83, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96 *, 97, 98, 99, 100, 101, 102, 103

Additional Table

➢ **Table 2 : Categorizing Literature base on Problem Formulation**

Application	Observation Space: High-dim	Observation Space: Low-dim	Reward Function: Sparse	Reward Function: Dense
Manipulation	54, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 124, 125, 128, 133, 134, 135, 136, 137, 141, 142, 145, 146, 147	122, 123, 126, 127, 129, 130, 131, 132, 138, 139, 140, 143, 144	54, 108, 110, 111, 112, 114, 115, 117, 118, 122, 124, 128, 129, 134	109, 113, 116, 118, 119, 120, 121, 123, 125, 126, 127, 130, 131, 132, 133, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147
MoMa	165, 169, 166, 162, 170, 167, 161, 173, 159, 172	155, 156, 154, 157, 171, 160, 164, 163, 168	170, 173, 172	155, 156, 154, 157, 165, 169, 171, 160, 164, 166, 162, 163, 167, 161, 168, 159
HRI	175, 176, 179, 180, 183	174, 177, 178, 181, 182, 184, 185	174	175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185
Multi-Robot Interaction		186, 187, 189, 190, 191, 192, 193	186, 187	189, 190, 191, 192, 193

Additional Table

Solution Approach

➢ **Table 3 : Categorizing Literature base on Solution Approach**

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

Additional Table

Solution Approach

➢ **Table 3 : Categorizing Literature base on Solution Approach**

Application	Simulator Usage: Zero-shot Sim-to-Real	Few-shot Sim-to-Real	No Simulator
Manipulation	108, 111, 123, 130, 133, 134, 135, 137, 138, 139, 141, 142, 143, 144, 145, 146, 147	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	155, 156, 154, 157, 165, 169, 171, 164, 166, 163, 167, 161, 168, 173, 159	160, 172	162, 170
HRI	174, 175, 176, 177, 178, 179, 181	184	180, 183, 182
Multi-Robot Interaction	186, 187, 189, 190, 191, 192, 193		