Lecture 14
Shareability & Reproducibility

Song Chaoyang

Assistant Professor

Department of Mechanical and Energy Engineering

songcy@sustech.edu.cn

SUSTech

Southern University of Science and Technology

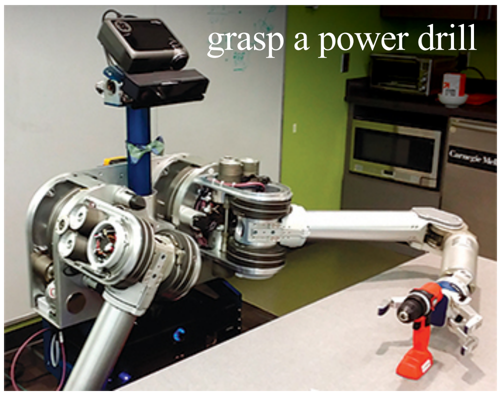
Task-based benchmarking for robotic science

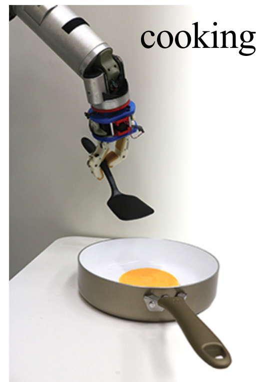
Towards experimental evidence that is shareable & reproducible

Benchmarking Tests

- **Physical**
  - Hand
  - Arm
- **Dexterity**
  - Motion
  - Finger
- **Functional**
  - Unimanual
  - Bimanual

(Huamán Quispe, et al, 2018)





Credit: Yale University

AncoraSIR.com

SUSTech
Southern University of Science and Technology

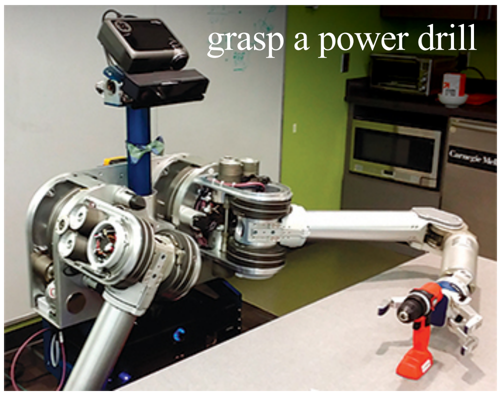
Task-based benchmarking for robotic science

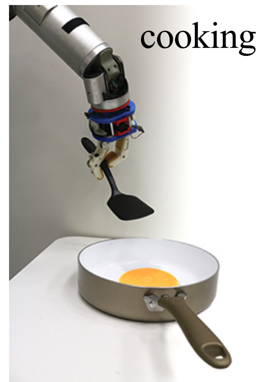
Towards experimental evidence that is shareable & reproducible

Benchmarking Tests

- Physical
  - Hand
  - Arm
- Dexterity
  - Motion
  - Finger
- Functional
  - Unimanual
  - Bimanual

(Huamán Quispe, et al, 2018)





Credit: Yale University

YCB object set

AncoraSIR.com

SUSTech
Southern University of Science and Technology

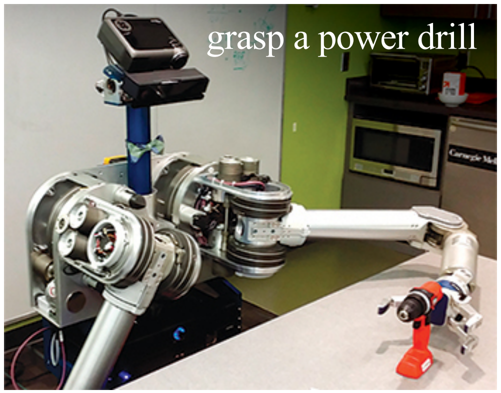
Task-based benchmarking for robotic science

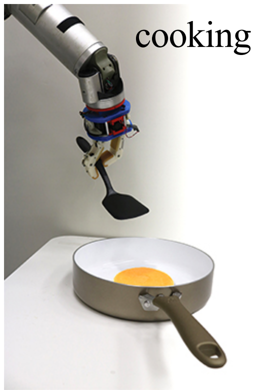
Towards experimental evidence that is shareable & reproducible

Benchmarking Tests

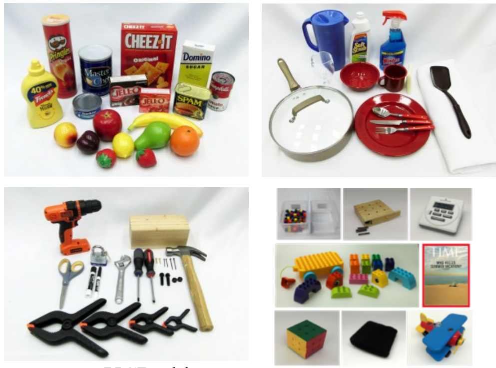
- **Physical**
  - Hand
  - Arm
- **Dexterity**
  - Motion
  - Finger
- **Functional**
  - Unimanual
  - Bimanual

(Huamán Quispe, et al, 2018)





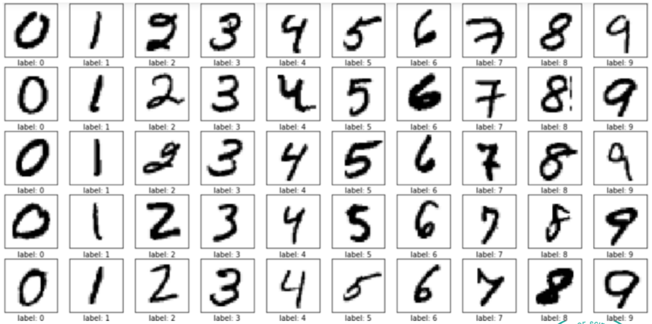
Credit: Yale University



YCB object set

Can we?

(or Should we?)



MNIST dataset

AncoraSIR.com

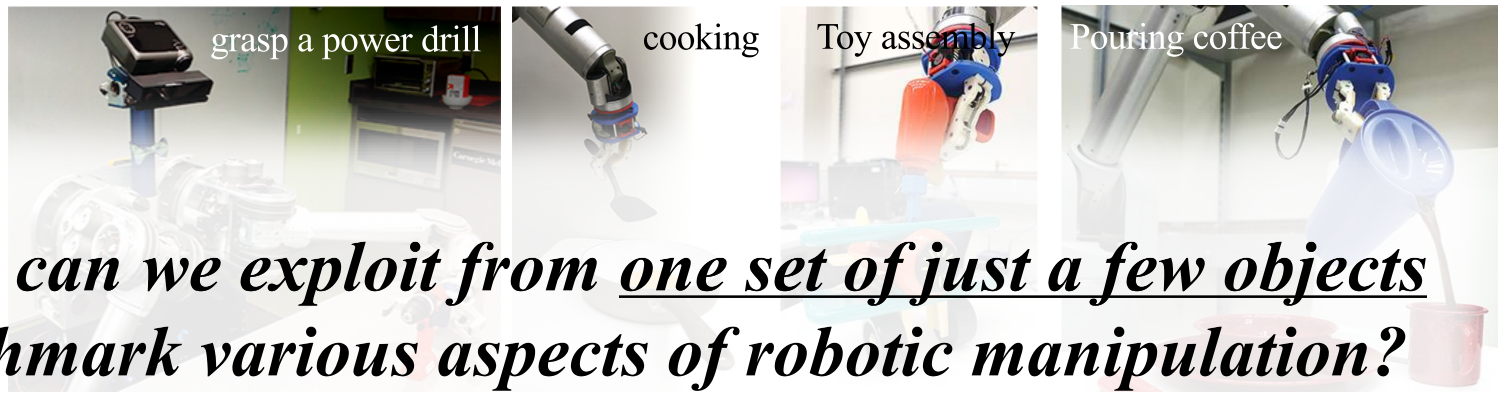
Task-based benchmarking for robotic science

Towards experimental evidence that is shareable & reproducible

Benchmarking Tests

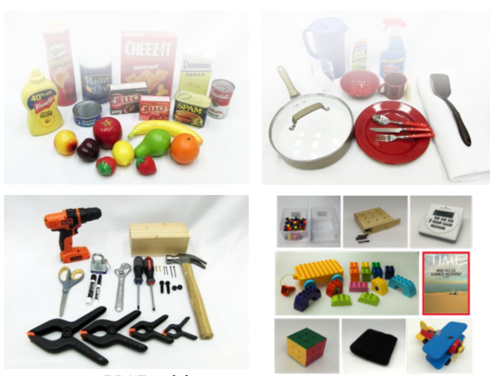
- Physical
  - Arm
- Dexterity
  - Motion
  - Finger
- Functional
  - Unimanual
  - Bimanual

(Huamán Quispe, et al, 2018)



How far can we exploit from <u>one set of just a few objects</u> to benchmark various aspects of robotic manipulation?

Credit: Yale University



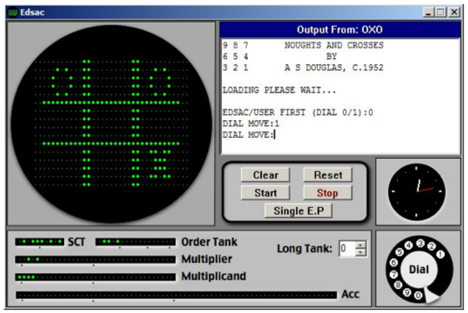
YCB object set

MNIST dataset

AncoraSIR.com

Game as a widely adopted benchmark for learning

Transferrable to robotic manipulation with the ease of accessibility and understanding



Sub-human
||
Par-human
||
High-human
||
Super-human
||
Optimal

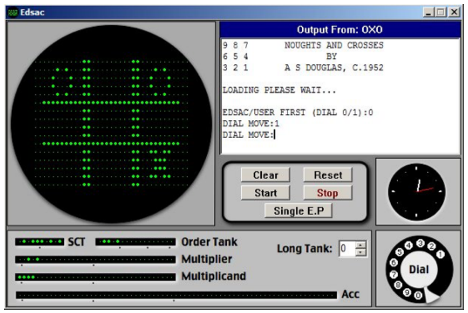
Game	Champion year[6]	Legal states ($\log_{10}$)[7]	Game tree complexity ($\log_{10}$)[7]	Game of perfect information?	Ref
Othello (reversi)	1997	28	58	Perfect	[8]
Draughts (checkers)	1994	21	31	Perfect	[9]
Chess	1997	46	123	Perfect	
Scrabble	2006				[10]
Shogi	2017	71	226	Perfect	[11]
Go	2016	172	360	Perfect	
2p no-limit hold 'em	2017			Imperfect	[12]
StarCraft	-	270+		Imperfect	[13]

AncoraSIR.com

SUSTech
Southern University of Science and Technology

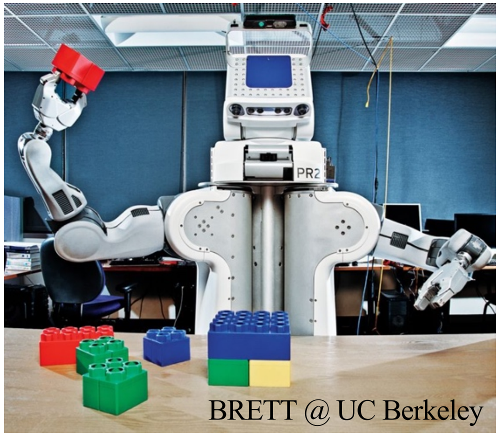
Game as a widely adopted benchmark for learning

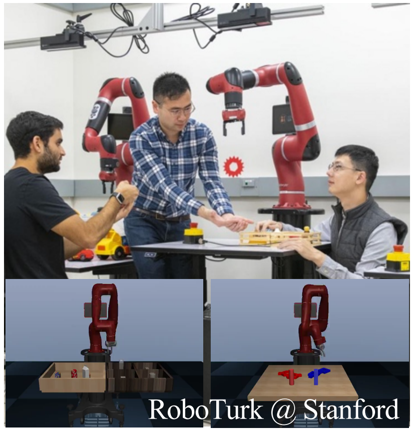
Transferrable to robotic manipulation with the ease of accessibility and understanding



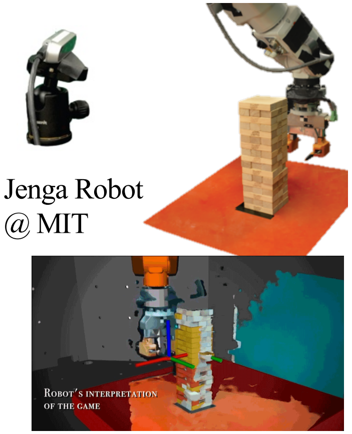
Sub-human
||
Par-human
||
High-human
||
Super-human
||
Optimal

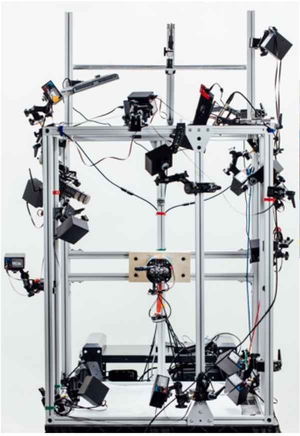
Game	Champion year[6]	Legal states ($\log_{10}$)[7]	Game tree complexity ($\log_{10}$)[7]	Game of perfect information?	Ref
Othello (reversi)	1997	28	58	Perfect	[8]
Draughts (checkers)	1994	21	31	Perfect	[9]
Chess	1997	46	123	Perfect	
Scrabble	2006				[10]
Shogi	2017	71	226	Perfect	[11]
Go	2016	172	360	Perfect	
2p no-limit hold 'em	2017			Imperfect	[12]
StarCraft	-	270+		Imperfect	[13]

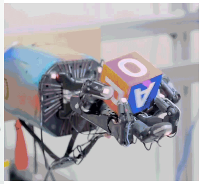

BRETT @ UC Berkeley


RoboTurk @ Stanford

Jenga Robot @ MIT







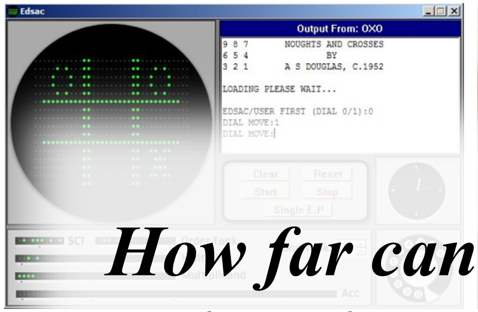
Dactyl @OpenAI

SUSTech
Southern University of Science and Technology

AncoraSIR.com

Game as a widely adopted benchmark for learning

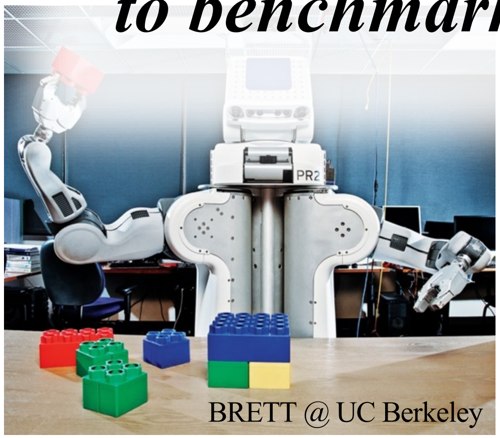
Transferrable to robotic manipulation with the ease of accessibility and understanding



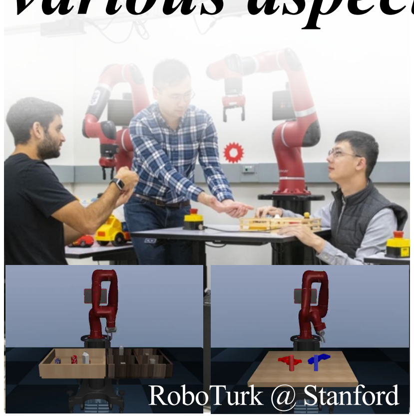
Sub-human
||
Par-human
||
High-human
||
Super-human

Game	Champion year[6]	Legal states ($\log_{10}$)[7]	Game tree complexity ($\log_{10}$)[7]	Game of perfect information?	Ref
Othello (reversi)	1997	28	58	Perfect	[8]
Draughts (checkers)	1994	21	31	Perfect	[9]
Chess	1997	46	123	Perfect	
Scrabble	2006				[10]
Shogi	2017	71	226	Perfect	[11]
Go	[illegible]	172	360	Perfect	
[illegible]	[illegible]			Imperfect	[12]
StarCraft	-	270+		Imperfect	[13]

How far can we exploit from <u>a simple yet scalable game</u> to benchmark various aspects of robotic manipulation?

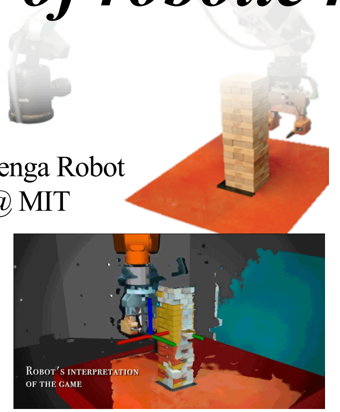


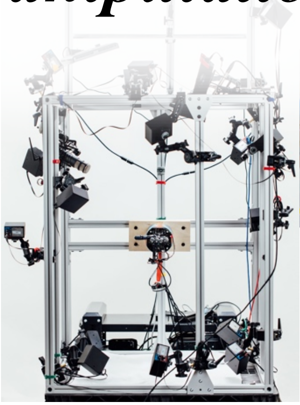
BRETT @ UC Berkeley

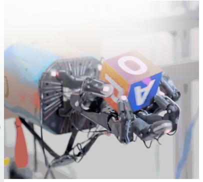


RoboTurk @ Stanford

Jenga Robot
@ MIT







Dactyl
@OpenAI

SUSTech
Southern University of Science and Technology

AncoraSIR.com

Benchmarking in Robotics

Objective performance evaluation of a robotic system or subsystem under controlled conditions

- Why?
  - It is a specific way of performing experimental evaluation
  - It enables a comparison of different systems on a common, predefined, setting
  - It provides a set of metrics (numerical scores / pass or fail / ranking / ...) together with a proper interpretation to perform an objective evaluation
  - It enables reproducibility and repeatability of experiments

AncoraSIR.com

SUSTech
Southern University of Science and Technology

Competition and Experiments

Can Competitions be treated as scientific experiments (despite their obvious differences)?

- *"Challenge and competition events in robotics provide an excellent vehicle for advancing the state of the art and evaluating new algorithms and techniques in the context of a common problem domain. [...] treat competitions and challenges as repeatable experiments."*
    - Monica Anderson, Odest Chadwicke Jenkins, and Sarah Osentoski Recasting Robotics Challenges as Experiments, IEEE Robotics & Automation Magazine, June 2011, 10-11

AncoraSIR.com

SUSTech
Southern University of Science and Technology

Experiments vs. Competitions

Competitions should aim at providing benchmarks by adopting a scientific approach

- *"Scientific" means able to increase scientific and technological knowledge by using rigorously experimental method*
- The experimental method suggest experiments to be designed to allow for:
  - Comparison
  - Reproducibility / repeatability
  - Justification / explanation

What Makes an Experiment

- **<u>Comparison</u>**: to know what has been already done in the field, to avoid the repetition of uninteresting experiments, and to get hints on promising issues to tackle.

- **<u>Reproducibility and repeatability</u>**: they are related to the idea that scientific results should be severely criticized to be confirmed; reproducibility is the possibility for independent scientists to verify the results of a given experiment by repeating it with the same initial conditions, instruments and techniques; repeatability is the property of an experiment that yields the same outcome from a number of trials performed at different times and/or in different places.

- **<u>Justification and explanation</u>**: it is not sufficient to collect as many precise data as possible, but it is also necessary to look for an explanation, namely all the experimental data should be interpreted in order to derive the correct implications that lead to the conclusion.

AncoraSIR.com

SUSTech
Southern University of Science and Technology

Benchmarking Competitions

Competitions often lack scientific grounding

- They do not apply the “scientific method” to allow comparison, reproducibility and repeatability, justification and explanation
- As for justification and explanation, they produce a ranking, but few insights on the motivations for this ranking
- Their results cannot be used as benchmarking tools

- The Benchmarking through Competition Challenge
  - *“Designing competitions to make them more scientifically grounded and suitable as benchmarks”*

AncoraSIR.com

Non-Robotic Scientific Competitions

- Scientific Competitions treated as (paper) experiments:
  - Machine Learning and Pattern Recognition (e.g., Kaggle)
  - Computational Intelligence in Games (e.g., CIG)
  - Information Retrieval (e.g., TREC)
  - Computer Vision (e.g., PETS)

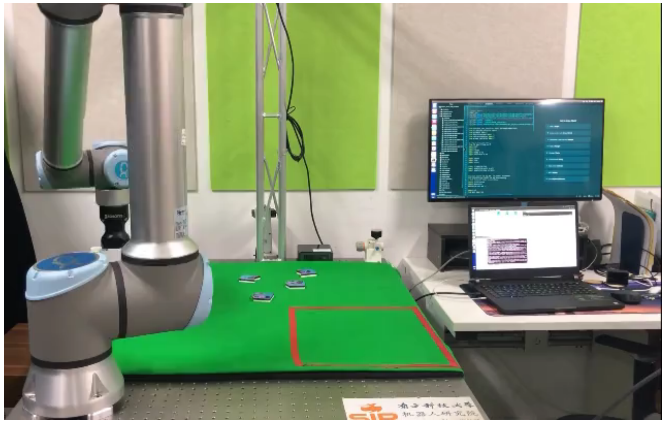
- Most of them have nowadays reached the level of
  - Defining proper metrics to measure significant aspects of the scientific result (e.g., F-measure)
  - Having different testbeds/tasks/scores (with different features) used to avoid overspecialization (e.g., background subtraction)
  - Investigating general features of the tasks and testbed used to design new competitions from an application perspective

AncoraSIR.com

Project 5: Autonomous Robot Manipulation

General Guideline

- Design a competitive and autonomous picking robot
  - Design a robot system to autonomously play jigsaw competitively;
  - Each team will present your robot system design in the final class
    - Writing a full paper describing your robot system in technical details;
    - Submit a video demo of your robot system design;
    - Submit a poster to present your robot system design;
    - Open-source your codes by uploading on GitHub;
    - Compete live and win! (Good luck)



- 30%: final project marking, including
  - 10% final paper
  - 10% final video demo
  - 5% final poster
  - 5% code submission

1 base frame piece

4 fragment pieces

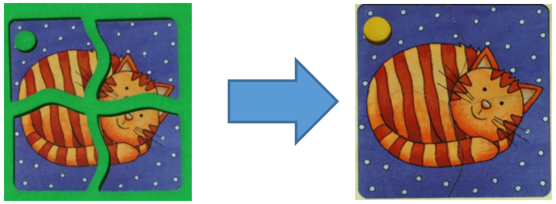
Full assembly of the jigsaw

AncoraSIR.com

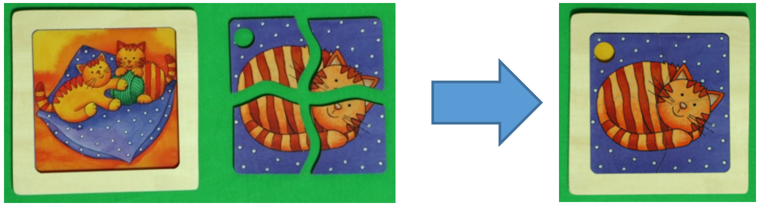
SUSTech Southern University of Science and Technology

Detailed Task Description

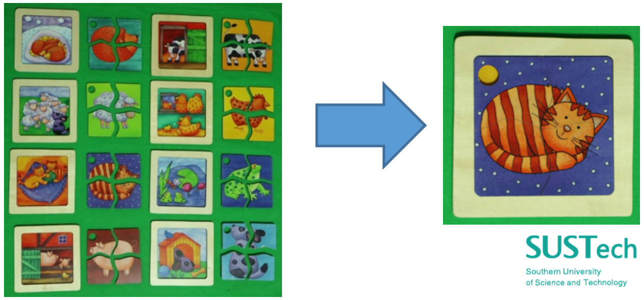
- **[General Task Description]**
  - A robot system autonomously picks up the four fragment pieces into a matching base frame of the jigsaw puzzle correctly using the shortest amount of time with the best accuracy base on textural and/or geometrical relationship.



- **[sub-task-1] a simple 4-piece jigsaw**
  - Only the 4 fragmented pieces of 1 jigsaw set are involved, and the finish the jigsaw on a flat desktop



- **[sub-task-2] a simple 5-piece jigsaw**
  - All 5 pieces of 1 jigsaw set are used, finish the jigsaw by placing all4 fragmented pieces inside the given base frame piece.



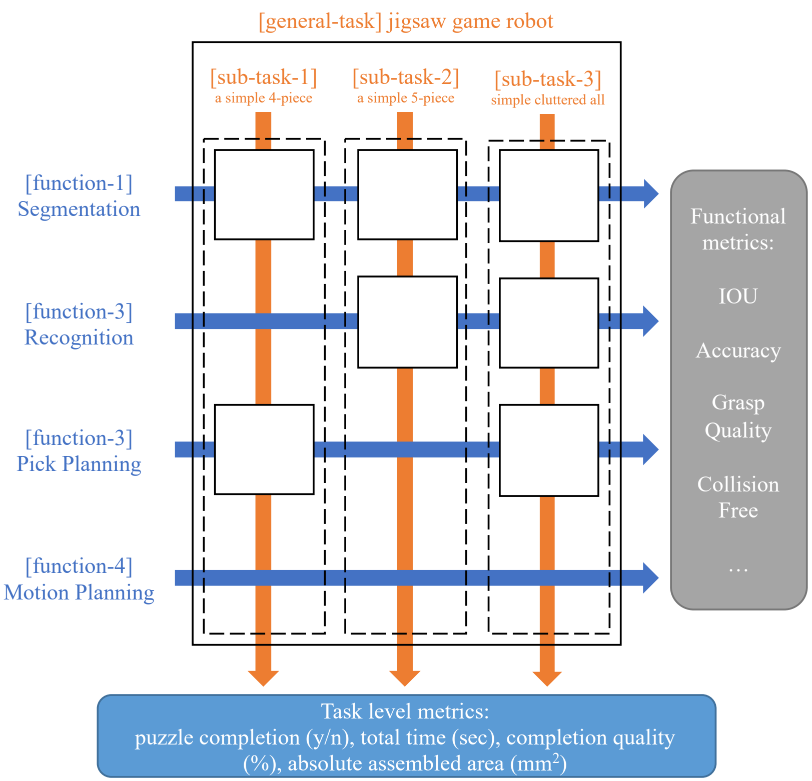
- **[sub-task-3] simple cluttered all jigsaws**
  - All pieces of 8 jigsaw sets are involved, finish 1 jigsaw based on a given base frame piece by placing all 4 matching fragmented pieces inside the base frame piece

AncoraSIR.com

SUSTech
Southern University of Science and Technology

Detailed Metric Description

Team #	1, 2, 3		**Sub-task-#**	1, 2, 3	**Name**		**Date**	
Trail Start Time	**Task-level metrics**				**Functional-level Description**			
(YYMMDD-HHMMSS)	**Game Completion (yes/no)**	**Time to Completion (sec)**	**Bounding Sqare Area (mm^2)**	**Completion Area Ratio (%)**	**Segmentation**	**Recognition**	**Pick Planning**	**Motion Planning**
(1/5 experiment trails)	*Record*	*Record*	*Record*	*Record*	*Describe if any*	*Describe if any*	*Describe if any*	*Describe if any*



In paper, each team try to complete each task for 5 times and record the results.

- Must record the whole experiments.

In class, each team try to complete each task for 1 time and record the results.

- Live demonstration.

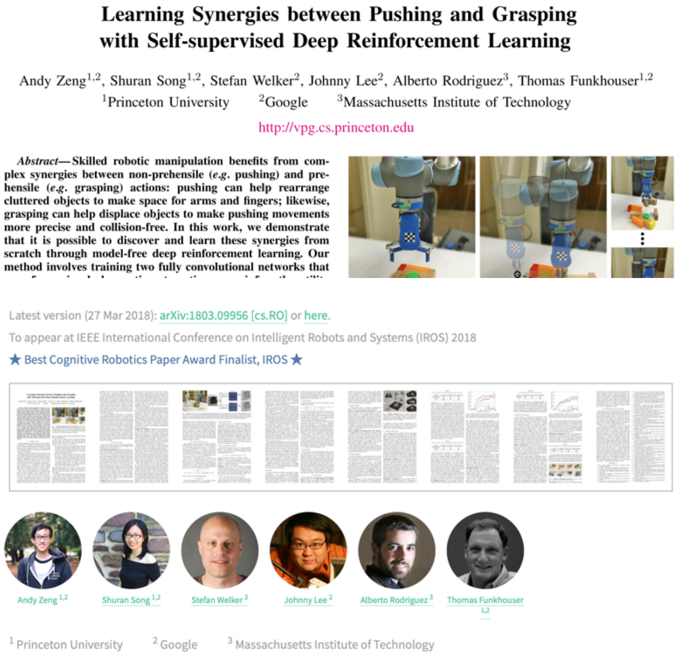
AncoraSIR.com

SUSTech
Southern University of Science and Technology

Sample Submissions

“Visual Pushing and Grasping Toolbox” by Andy Zeng

- [paper] https://arxiv.org/pdf/1803.09956.pdf
  - IEEE Template (https://www.ieee.org/conferences/publishing/templates.html)
- [video] https://www.youtube.com/watch?v=-OkyX7ZlhiU
- [poster] http://vpg.cs.princeton.edu
- [code] https://github.com/andyzeng/visual-pushing-grasping



Acknowledgement

Make sure to add the following to your final poster, paper & video

- *"The project is part of the required assignment at the class of ME336 Collaborative Robotic Learning at Southern University of Science and Technology, delivered by Prof. Song Chaoyang (songcy@sustech.edu.cn) and Dr. Wan Fang (sophie.fwan@hotmail.com) at the Bionic Design and Learning Lab in the Spring term of 2019."*
  - *If any plan for future publications, please add corresponding authorship of the instructors with the above acknowledgement to your submissions.*

SUSTech
Southern University of Science and Technology

AncoraSIR.com

Thank you!

Prof. Song Chaoyang

- Dr. Wan Fang (sophie.fwan@hotmail.com)

AncoraSIR.com