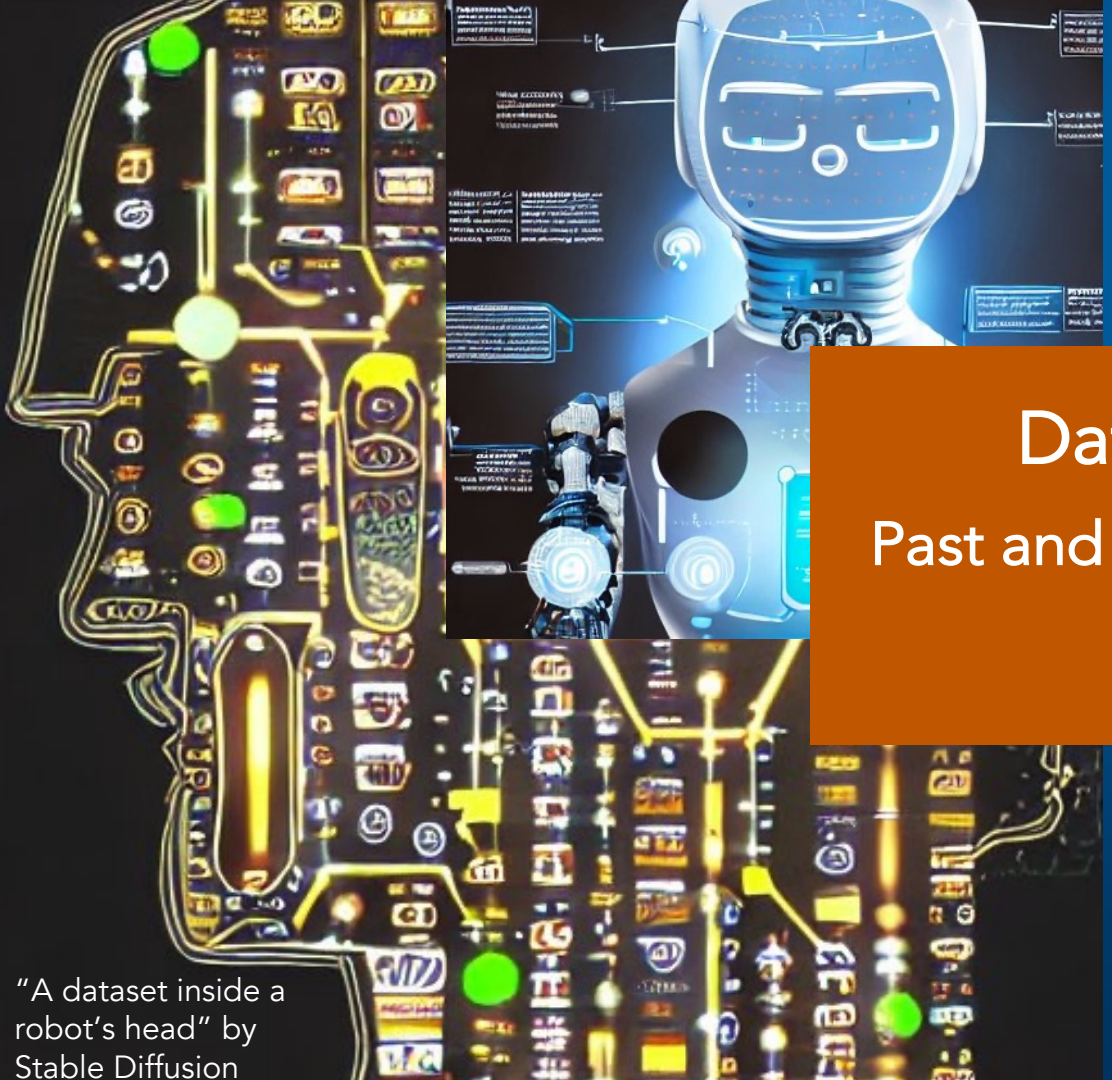


Datasets in Robotics:

Past and Future. An open discussion

Dr. Roberto Martín-Martín

RobIn
ROBOTIC INTERACTIVE INTELLIGENCE LAB

TEXAS Robotics

THE UNIVERSITY OF TEXAS AT AUSTIN

"A dataset inside a robot's head" by Stable Diffusion

Meta: Why this Presentation?

- Collaboration in the "Good Systems" team between Roboticists and experts in Information Science and Data Management

- For Information Science and Data Management:
  Understanding the use of datasets in Robotics (and other fields) will help creating better information / dataset protocol

- For Robotics:
  Explicit discussion and introspective understanding of the reasons and the goals for dataset generation

What is a Dataset?

And some other definitions

Data are observations or measurements (unprocessed or processed) represented as text, numbers, or multimedia.

A **dataset** is a structured collection of data generally associated with a unique body of work.

A **database** is an organized collection of data stored as multiple datasets. Those datasets are generally stored and accessed electronically from a computer system that allows the data to be easily accessed, manipulated, and updated.

What is a Dataset?

"A Dataset if an opportunity for the research community to set a North Star"

Uses of Datasets in AI

Three main purposes

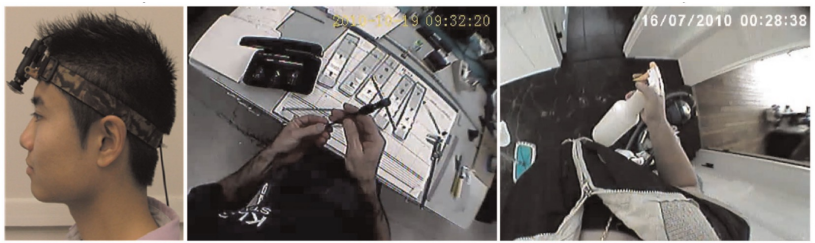
Record an event so that it can be analyzed later → extract regularities, understand events, gain knowledge

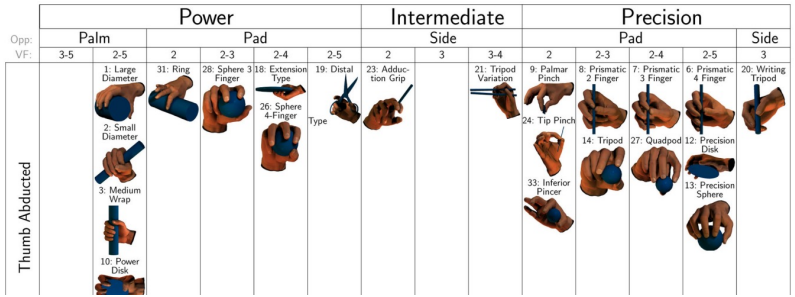
Serve as shared experimental evaluation for solutions to the same problem → Fair comparison

[RECENT!] Train solutions for a problem using data driven methods

Gaining Knowledge from a Dataset

The Yale human grasping dataset





- Dataset of humans performing common tasks
- Helps to create a comprehensive taxonomy of possible grasp types
- The taxonomy is used to analyze frequency and types of grasps in activities

Nr.	Name	Type	Opp. Type	Th. Pos	VF	Lit. Ref.	Prev. [54]	Mass [g] [57]	Size [cm] [57]	Rigidity [57]	Force [58]
1	Large Diameter	Pow.	Palm	Abd.	VF 2: 2-5 VF 3:	15	F: 1.7 % D: 1.5 %	0 500 1,000	0 5 10 15	ri sq	wt
2	Small Diameter	Pow.	Palm	Abd.	VF 2: 2-5 VF 3:	8	F: 0.7 % D: 0.3 %	0 500 1,000	0 5 10 15	ri sq	wt in
3	Medium Wrap	Pow.	Palm	Abd.	VF 2: 2-5 VF 3:	11	F: 12.7 % D: 23.8 %	0 500 1,000	0 5 10 15	ri sq	wt in
4	Adducted Thumb	Pow.	Palm	Add.	VF 2: 2-5 VF 3: 1	2	F: 1.0 % D: 0.9 %	0 500 1,000	0 5 10 15	ri	wt in
5	Light Tool	Pow.	Palm	Add.	VF 2: 2-5 VF 3: (1)	2	F: 4.8 % D: 3.6 %	0 500 1,000	0 5 10 15	ri	wt in
6	Prismatic 4 Finger	Pre.	Pad	Abd.	VF 2: 2-5 VF 3:	6	F: 4.0 % D: 2.0 %	0 500 1,000	0 5 10 15	ri sq	wt
7	Prismatic 3 Finger	Pre.	Pad	Abd.	VF 2: 2-4 VF 3:	4	F: 4.2 % D: 3.5 %	0 500 1,000	0 5 10 15	ri	wt
8	Prismatic 2 Finger	Pre.	Pad	Abd.	VF 2: 2-3 VF 3:	3	F: 6.4 % D: 4.3 %	0 500 1,000	0 5 10 15	ri	wt
9	Palmar Pinch	Pre.	Pad	Abd.	VF 2: 2 VF 3:	18	F: 3.9 % D: 2.3 %	0 500 1,000	0 5 10 15	ri fl	wt

[The Yale human grasping dataset: Grasp, object, and task data in household and machine shop environments, IJRR 2015, Ian M. Bullock, Thomas Feix and Aaron M. Dollar]

Evaluating in Fair Conditions using a Dataset

Setting a common goal for the community

RobIn
ROBOTIC INTERACTIVE INTELLIGENCE LAB

- Dataset of images annotated with category labels
- Organized in a hierarchical taxonomy
- Used to benchmark image recognition solutions

Participation in 2011

96 registrations

15 submissions

Top Entries

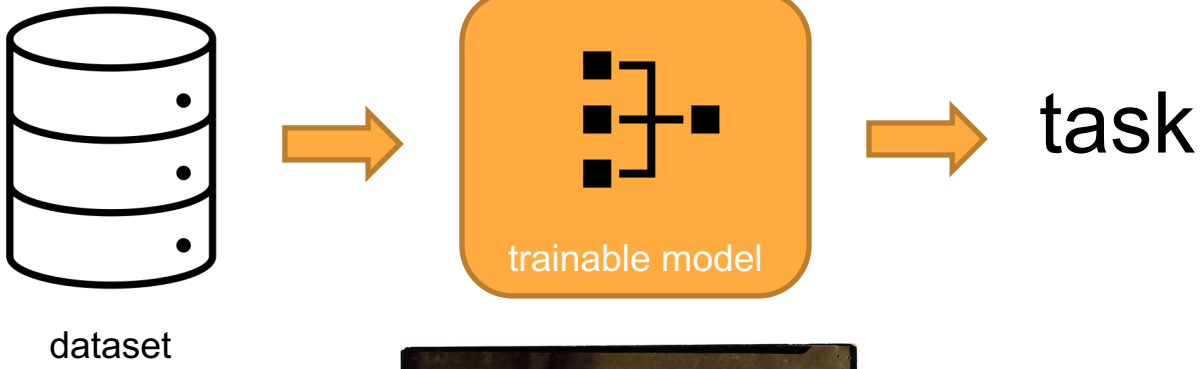
Xerox Research Centre Europe
Univ. Amsterdam & Univ. Trento
ISI Lab Univ. Tokyo
NII Japan

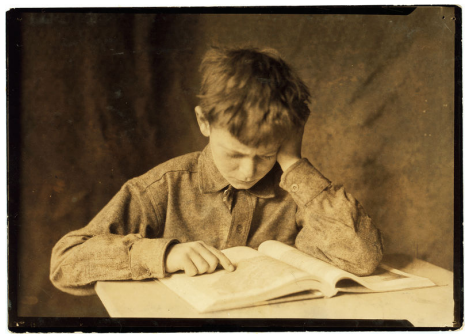
[Deng, Jia, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. "Imagenet: A large-scale hierarchical image database." In 2009 IEEE conference on computer vision and pattern recognition, pp. 248-255. Ieee, 2009.]

Training Machine Learning Based Solutions

A recent(-ish) trend

RobIn
ROBOTIC INTERACTIVE INTELLIGENCE LAB



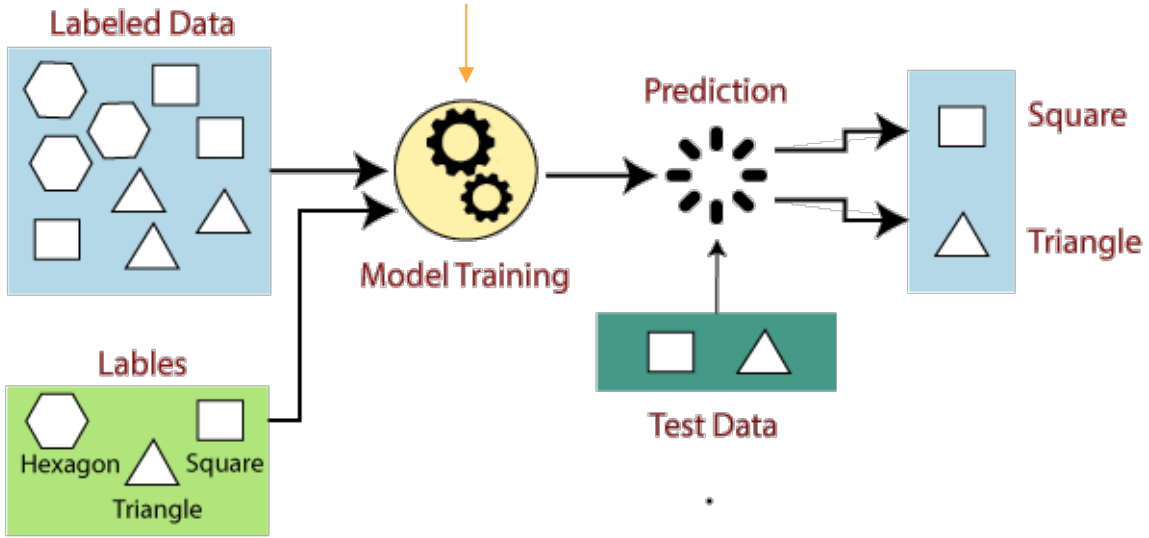


Background: Supervised Learning

How to consume datasets with machine learning

$$f(x) = label$$

$$(x, label)$$

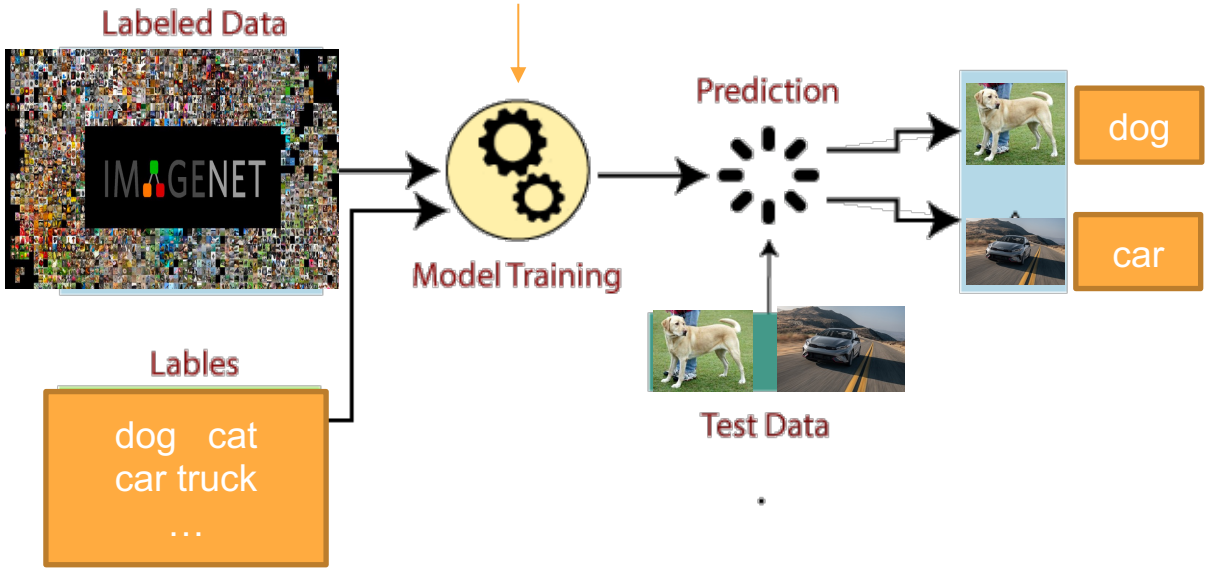


Supervised Learning + Datasets

Encode the information in the dataset into a model

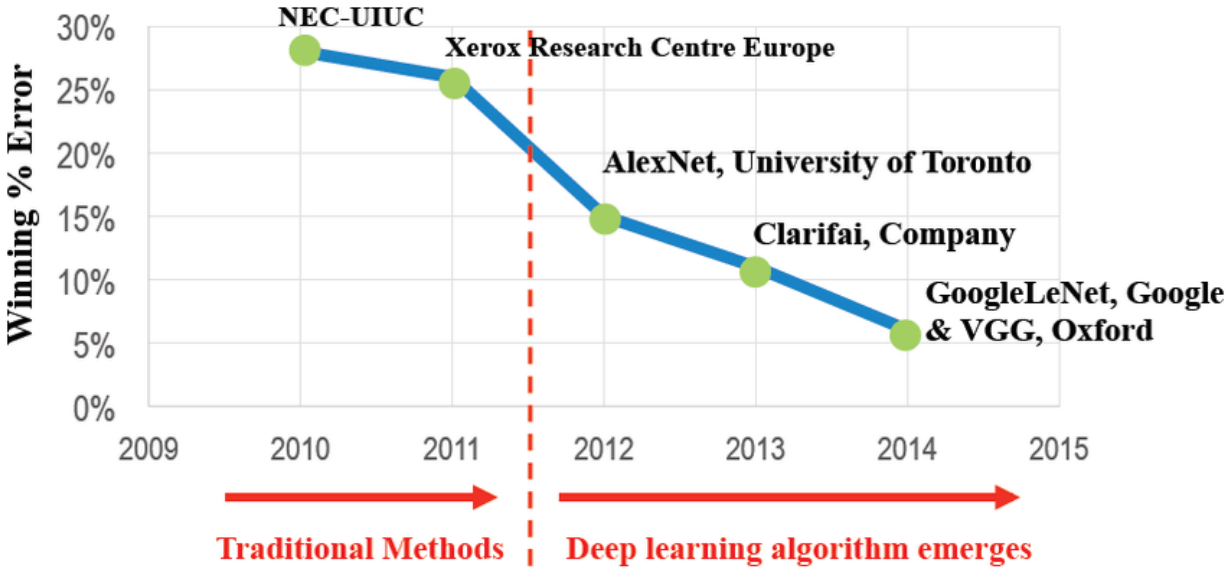
$$f(x) = label$$

$$(x, label)$$



The Dataset Revolution in Computer Vision

Data-driven methods win over classical (hardcoded) solutions



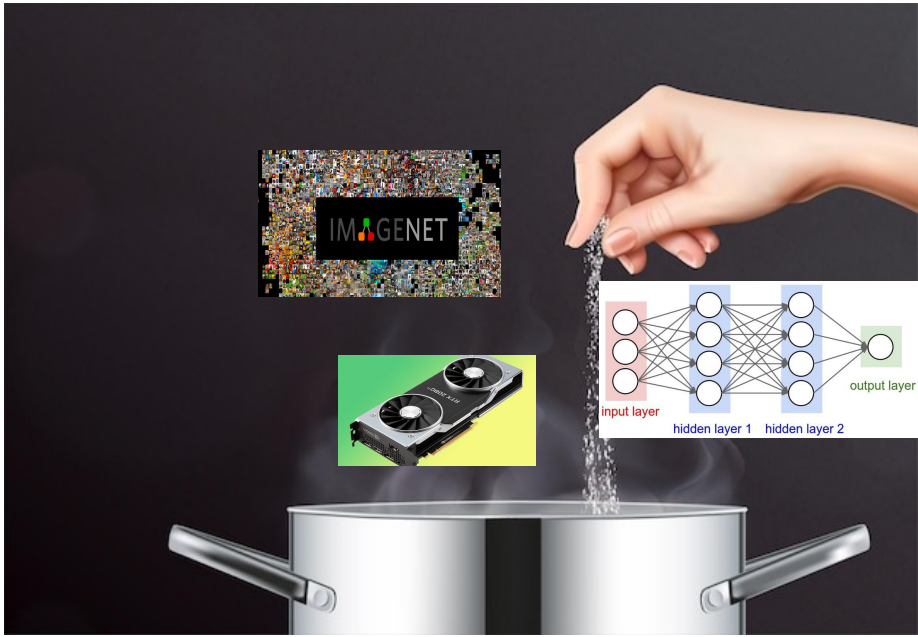
Results on ImageNet

"Image classification is solved"

The Ingredients of the Revolution

It is not (just) the algorithm; it is the data(set)!

- Large Dataset
- Model with enough capacity → Deep Neural Network
- Hardware to train the large model using the large dataset



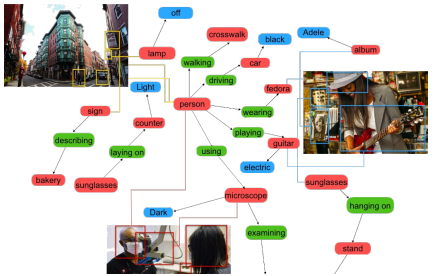
[Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Communications of the ACM 60, no. 6 (2017): 84-90.]

Image Datasets

The fuel for the AI revolution

- Image Classification
- Image Segmentation
- Scene Understanding
- Static
- Passive

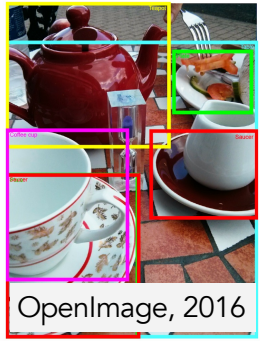
ImageNet, 2009 [Deng et al.]



VisualGenome, 2016 [Krishna et al.]

Pascal VOC, 2010 [Everingham et al.]

MS COCO, 2014 [Lin et al.]



OpenImage, 2016

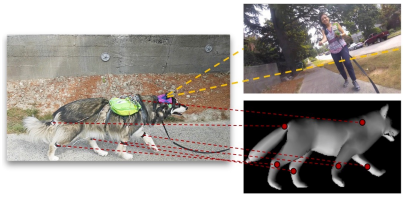
Extending the First Successes to other CV Tasks

Bringing in the temporal dimension

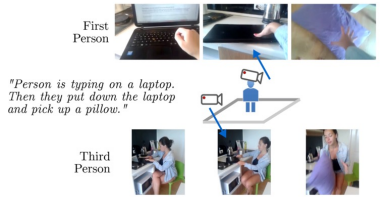
- Activity Recognition
- **Motion Understanding**
- **Human-Object Interaction**
- **Dynamic**
- **Passive**

Something-Something, 2017 [Goyal et al.]

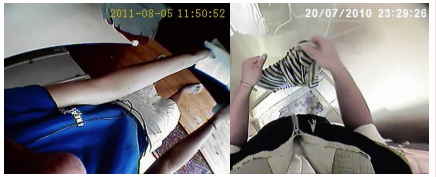
Epic Kitchens, 2018 [Damen et al.]



Decade-Dog, 2018 [Ehsani et al.]



Charades-Ego, 2018 [Sigurdsson et al.]

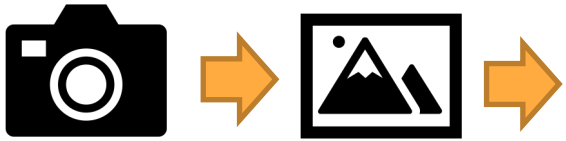


Grasping Dataset, 2014 [Bullock et al.]

Meanwhile, in Robotics…

Difficulties to find common ground

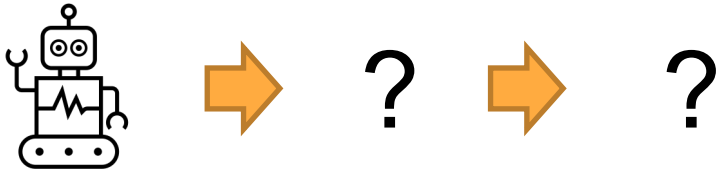
In Computer Vision:



- Image Classification
- Object Detection
- Segmentation
- Action Recognition
- Optical Flow Estimation

...

In Robotics:



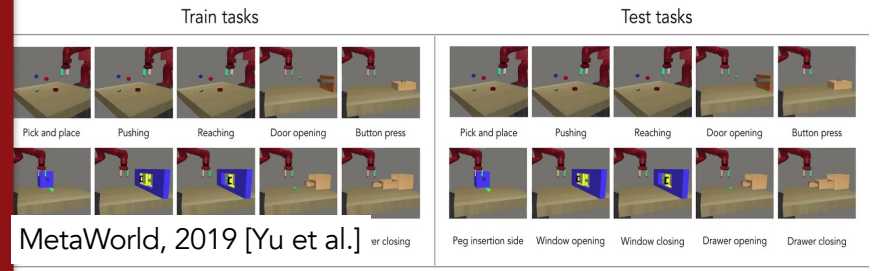
Robotics: Active Datasets?

From datasets to simulation environments?

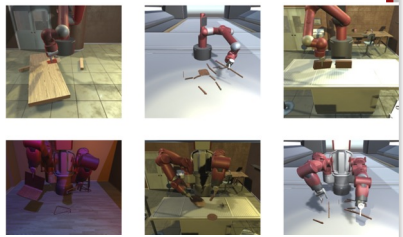
- Robot Task Learning
- **Benchmark for Solutions**
- **Task-Specific**
- **Dynamic**
- **Active**

RLBench, 2020 [James et al.]

DoorGym, 2019 [Urakami et al.]



MetaWorld, 2019 [Yu et al.]

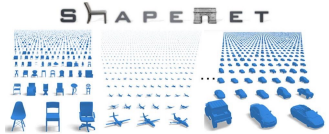


Ikea Assembly, 2019 [Lee et al.]

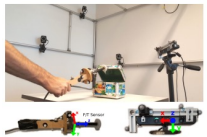
Some Examples of Robotics Datasets

Two types of Datasets

Datasets that are useful for Robots



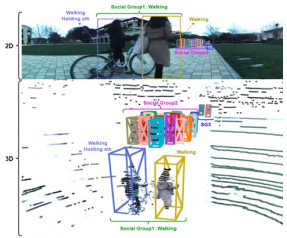
20Bn-smth-smth



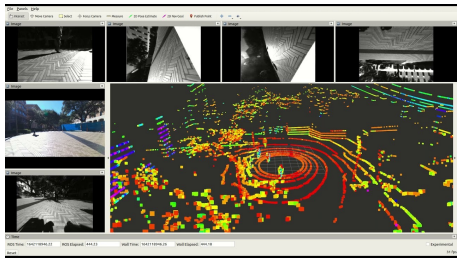
RBO Dataset

YCB Dataset

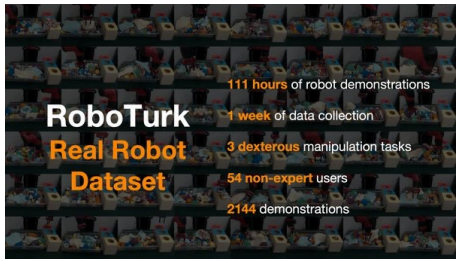
Datasets from Robots



JRDB



ScanD

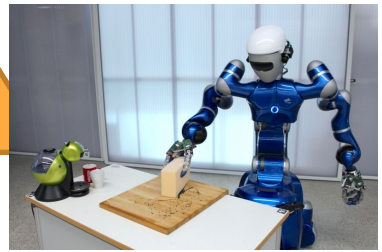


What does a Good Dataset in Robotics need?

Meta information

Robot Embodiment:
- Actuation
- Sensing
- Morphology

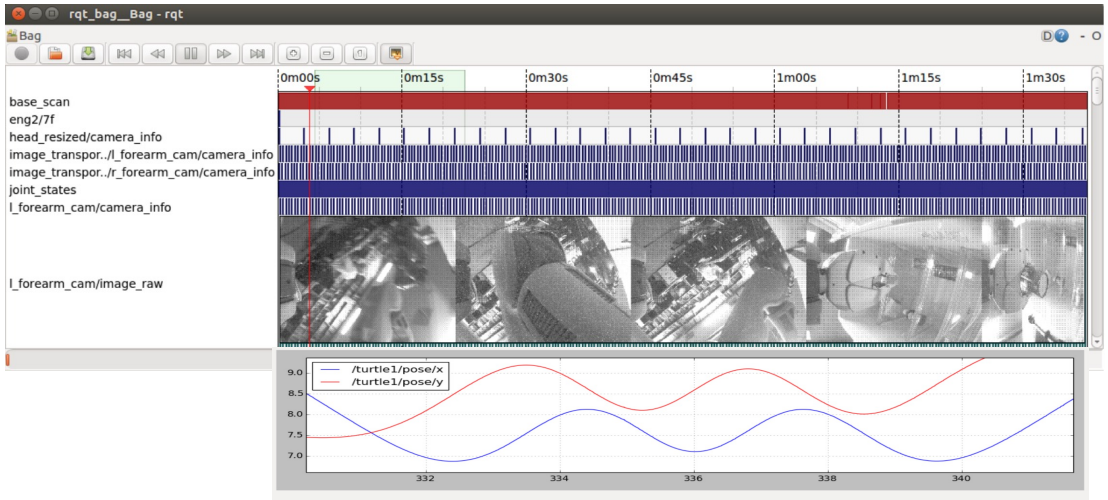
Environment:
- Layout
- Type
- Other agents
- Context



Event:
- Task/Goal
- Objectives
- Conditions

What does a Good Dataset in Robotics need?

Temporal information



- Timestamps of each signal
  - ideally synced
- Values in a protocoled format
  - Careful with compression!
  - Better raw
- Annotations

Resources

Where to find datasets that are relevant for Robotics

- IJRR Data Papers: https://journals.sagepub.com/topic/collections-ijr/ijr-3-datapapers/ijr
  - ”Dry” description of the datasets with focus on reproducibility

- NeurIPS Dataset track: https://neurips.cc/Conferences/2022/CallForDatasetsBenchmarks
  - New track focused on Datasets and Benchmarks for Robotics and AI in general

- Awesome Datasets for Robotics: https://github.com/mint-lab/awesome-robotics-datasets
  - Collection of links collected by a lab in Korea

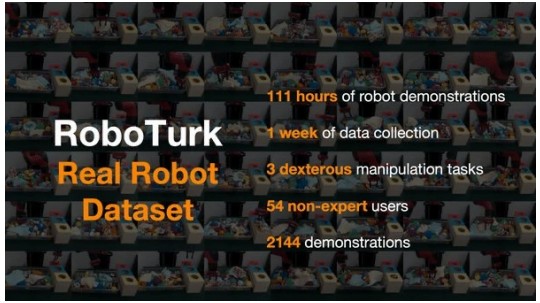
The Future in Other Fields

Removing the need for (costly) annotations

"The University of Texas at [ ] (UT Austin, UT, or Texas) is a public research [ ] in Austin, Texas. It was founded in 1883 and is the oldest [ ] in the University of [ ] System.

The Future in Robotics

Learning to act from large datasets?



$(x, label)$

actions

$f(x) = label$

policy

YouTube

Summary

- Datasets have three usages in AI/Robotics
  - Analyze/understand events
  - Benchmark solutions
  - Train solutions with ML
- This last usage is becoming more and more important in robotics
- Robotics datasets are harder than in CV due to the unclear/variable interface
- Datasets will still play an important role in AI/robotics but researchers want to avoid manual annotations

Open Questions:

- What type of datasets did/do you use for robotics?
  - What data
  - What annotations
  - What size

- What do you think are the main current and future uses of datasets in robotics?

- What information do our non-robotics colleagues need from us when helping us creating datasets?

References

- Grasping dataset: The Yale human grasping dataset: Grasp, object, and task data in household and machine shop environments, IJRR 2015, Ian M. Bullock, Thomas Feix and Aaron M. Dollar
- ImageNet: Deng, Jia, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. "Imagenet: A large-scale hierarchical image database." In 2009 IEEE conference on computer vision and pattern recognition, pp. 248-255. Ieee, 2009
- alexNet: Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Communications of the ACM 60, no. 6 (2017): 84-90
- VisualGenome: Krishna, Ranjay, Yuke Zhu, Oliver Groth, Justin Johnson, Kenji Hata, Joshua Kravitz, Stephanie Chen et al. "Visual genome: Connecting language and vision using crowdsourced dense image annotations." International journal of computer vision 123 (2017): 32-73.
- PascalVOC: Everingham, Mark, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. "The pascal visual object classes (voc) challenge." International journal of computer vision 88 (2009): 303-308.
- MSCOCO: Lin, Tsung-Yi, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. "Microsoft coco: Common objects in context." In Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13, pp. 740-755. Springer International Publishing, 2014.
- OpenImage: Kuznetsova, Alina, Hassan Rom, Neil Alldrin, Jasper Uijlings, Ivan Krasin, Jordi Pont-Tuset, Shahab Kamali et al. "The open images dataset v4: Unified image classification, object detection, and visual relationship detection at scale." International Journal of Computer Vision 128, no. 7 (2020): 1956-1981.
- Something-Something: Goyal, Raghav, Samira Ebrahimi Kahou, Vincent Michalski, Joanna Materzynska, Susanne Westphal, Heuna Kim, Valentin Haenel et al. "The" something something" video database for learning and evaluating visual common sense." In Proceedings of the IEEE international conference on computer vision, pp. 5842-5850. 2017.
- Epic Kitchens: Damen, Dima, Hazel Doughty, Giovanni Maria Farinella, Sanja Fidler, Antonino Furnari, Evangelos Kazakos, Davide Moltisanti et al. "Scaling egocentric vision: The epic-kitchens dataset." In Proceedings of the European Conference on Computer Vision (ECCV), pp. 720-736. 2018.
- Decade Dog: Ehsani, Kiana, Hessam Bagherinezhad, Joseph Redmon, Roozbeh Mottaghi, and Ali Farhadi. "Who let the dogs out? modeling dog behavior from visual data." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4051-4060. 2018.
- Charades Ego: Sigurdsson, Gunnar A., Abhinav Gupta, Cordelia Schmid, Ali Farhadi, and Karteek Alahari. "Charades-ego: A large-scale dataset of paired third and first person videos." arXiv preprint arXiv:1804.09626 (2018).

References

- ShapeNet: Chang, Angel X., Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese et al. "Shapenet: An information-rich 3d model repository." arXiv preprint arXiv:1512.03012 (2015).
- PartNet: Mo, Kaichun, Shilin Zhu, Angel X. Chang, Li Yi, Subarna Tripathi, Leonidas J. Guibas, and Hao Su. "Partnet: A large-scale benchmark for fine-grained and hierarchical part-level 3d object understanding." In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 909-918. 2019.
- Ego4D: Grauman, Kristen, Andrew Westbury, Eugene Byrne, Zachary Chavis, Antonino Furnari, Rohit Girdhar, Jackson Hamburger et al. "Ego4d: Around the world in 3,000 hours of egocentric video." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 18995-19012. 2022.
- RBO Dataset: Martín-Martín, Roberto, Clemens Eppner, and Oliver Brock. "The RBO dataset of articulated objects and interactions." The International Journal of Robotics Research 38, no. 9 (2019): 1013-1019.
- YCB Dataset: Calli, Berk, Arjun Singh, James Bruce, Aaron Walsman, Kurt Konolige, Siddhartha Srinivasa, Pieter Abbeel, and Aaron M. Dollar. "Yale-CMU-Berkeley dataset for robotic manipulation research." The International Journal of Robotics Research 36, no. 3 (2017): 261-268.
- JRDB Dataset: Martin-Martin, Roberto, Mihir Patel, Hamid Rezatofighi, Abhijeet Shenoi, JunYoung Gwak, Eric Frankel, Amir Sadeghian, and Silvio Savarese. "Jrdb: A dataset and benchmark of egocentric robot visual perception of humans in built environments." IEEE transactions on pattern analysis and machine intelligence (2021).
- ScanD: Karnan, Haresh, Anirudh Nair, Xuesu Xiao, Garrett Warnell, Sören Pirk, Alexander Toshev, Justin Hart, Joydeep Biswas, and Peter Stone. "Socially compliant navigation dataset (scand): A large-scale dataset of demonstrations for social navigation." IEEE Robotics and Automation Letters 7, no. 4 (2022): 11807-11814.
- RoboNet: Dasari, Sudeep, Frederik Ebert, Stephen Tian, Suraj Nair, Bernadette Bucher, Karl Schmeckpeper, Siddharth Singh, Sergey Levine, and Chelsea Finn. "Robonet: Large-scale multi-robot learning." arXiv preprint arXiv:1910.11215 (2019).
- Roboturk: Mandlekar, Ajay, Yuke Zhu, Animesh Garg, Jonathan Booher, Max Spero, Albert Tung, Julian Gao et al. "Roboturk: A crowdsourcing platform for robotic skill learning through imitation." In Conference on Robot Learning, pp. 879-893. PMLR, 2018.
- ChatGPT: Ouyang, Long, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll L. Wainwright, Pamela Mishkin, Chong Zhang et al. "Training language models to follow instructions with human feedback." arXiv preprint arXiv:2203.02155 (2022).