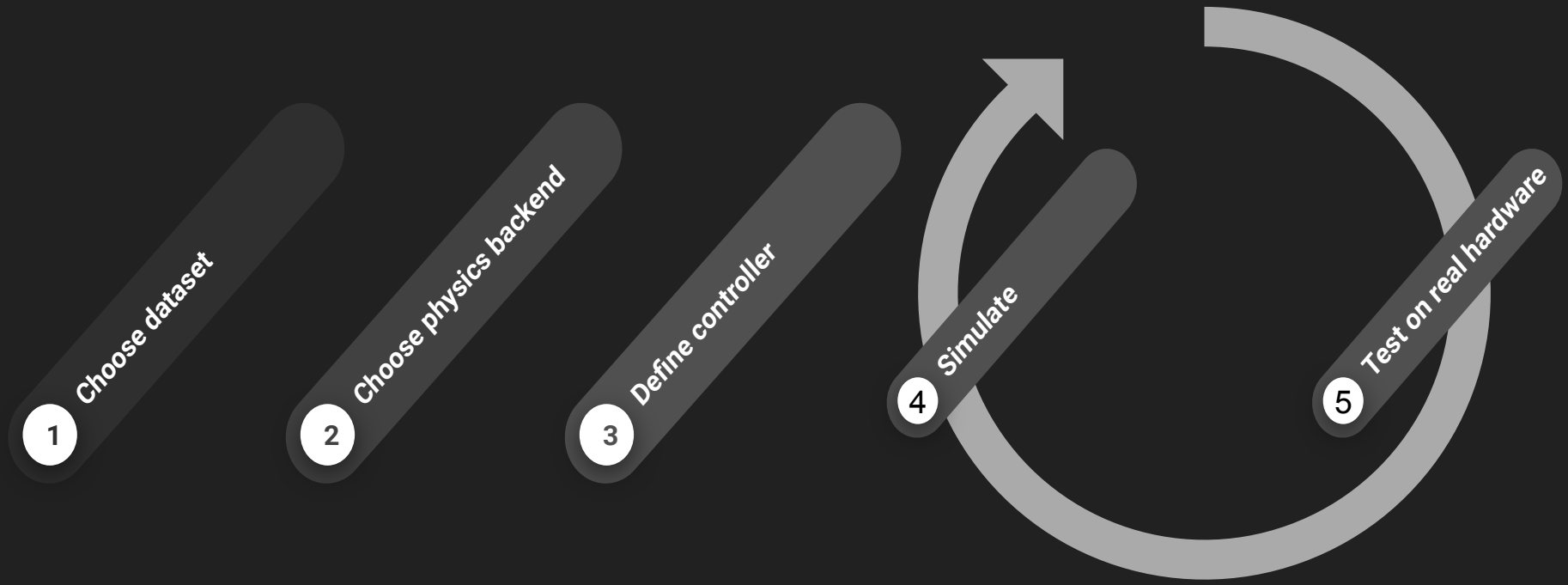
Simulation for Real World Robotics

David Watkins-Valls

Candidacy Exam

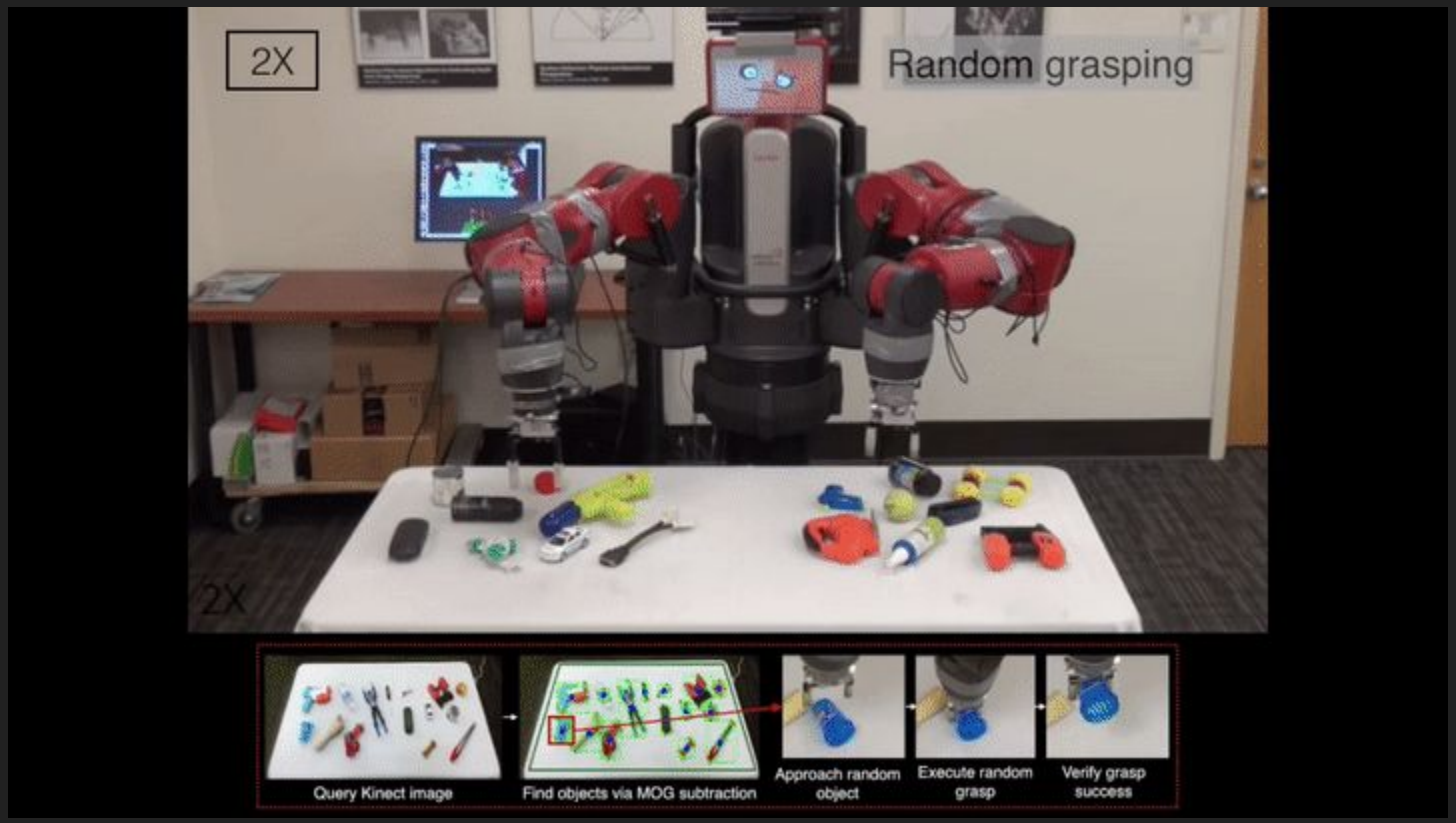
What is “Simulation”?

- Narrow scope to robotics simulation
- Specifically looking at how the field creates a useful pipeline for enabling real world robotics
- Throughout this talk we will be discussing issues related to this pipeline and how to improve it



Over 50K and 700 robot hours for 80% grasp accuracy

Supersizing self-supervision: Learning to grasp from 50K tries and 700 robot hours



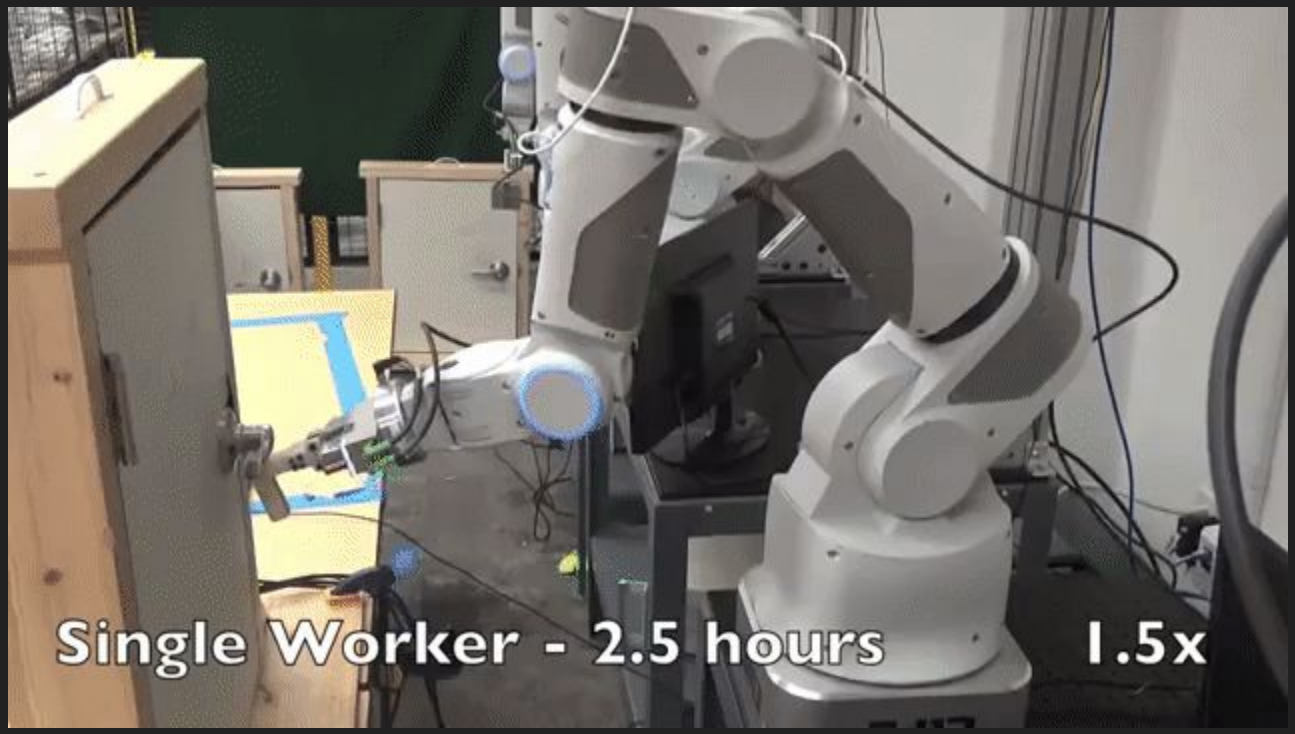
This is not feasible for every robotic task

New methods in simulation can reduce real-world experiment time

Pinto, L., & Gupta, A. (2016). Supersizing self-supervision: Learning to grasp from 50K tries and 700 robot hours. Proceedings - IEEE International Conference on Robotics and Automation, 2016–June, 3406–3413. http://doi.org/10.1109/ICRA.2016.7487517

4 workers and human assistance required to achieve simple tasks

Deep reinforcement learning for robotic manipulation with asynchronous off-policy updates



Faster than previous paper, but domain space is too small to generalize to all tasks and too expensive

Gu, S., Holly, E., Lillicrap, T., & Levine, S. (2017). Deep reinforcement learning for robotic manipulation with asynchronous off-policy updates. In Proceedings - IEEE International Conference on Robotics and Automation (pp. 3389–3396). http://doi.org/10.1109/ICRA.2017.7989385

Advantages and Shortcomings of Simulation

- Advantages
  - Simulation allows for rapid iteration
  - Can collect data substantially faster than in the real world
  - No risk to expensive hardware
  - Scalable across multiple machines

- Disadvantages
  - Tradeoff between accuracy and speed
  - Do not generalize well to real world problems
  - Missing many sensory modalities
  - High level APIs required for faster iteration in software development

- We can solve these disadvantages through domain randomization, newer physics simulators, and larger datasets that did not exist before

Choosing a Dataset

Real vs. Simulated Data

- Real data
  - Allows simulation to have higher realism
  - Sensors for capturing data have become cheaper
  - Time consuming to capture and annotate

- Why simulated data?
  - Easier to randomize the domain of the dataset
  - Can be much higher resolution than is supported by real world sensors
  - Meshes can be too perfect when compared to real world data
  - Some object geometries are not able to be simulated such fluids deformable objects

Example: Choosing a dataset for grasping

- Benchmarking in Manipulation Research : YCB Object and Model Set and Benchmarking Protocols
- ShapeNet: An Information-Rich 3D Model Repository
- Jacquard: A Large Scale Dataset for Robotic Grasp Detection
- Leveraging big data for grasp planning

- Need high fidelity dataset with diverse examples
- Dataset needs to be related to the problem: kitchen vs. home vs. industrial
- For grasping, four datasets are immediately relevant:

ShapeNet

51,300 unique 3D models
55 object categories
Simulated
Textured

YCB

77 unique 3D models
5 object categories
Real
Textured

GRASP

280 unique 3D models
87 object categories
Simulated
Untextured

Jacquard

240 unique 3D models
8019 hand-labeled grasp rectangles
Real
Textured

Calli, B., Walsman, A., Member, S., Singh, A., Member, S., Srinivasa, S., … Member, S. (n.d.). Benchmarking in Manipulation Research : The YCB Object and Model Set and Benchmarking Protocols.
Chang, A. X., Funkhouser, T., Guibas, L., Hanrahan, P., Huang, Q., Li, Z., … Yu, F. (2015). ShapeNet: An Information-Rich 3D Model Repository. http://doi.org/10.1145/3005274.3005291
Depierre, A., Dellandréa, E., & Chen, L. (2018). Jacquard: A Large Scale Dataset for Robotic Grasp Detection, 2–7. Retrieved from http://arxiv.org/abs/1803.11469
Kappler, Daniel, Jeannette Bohg, and Stefan Schaal. "Leveraging big data for grasp planning." 2015 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2015.

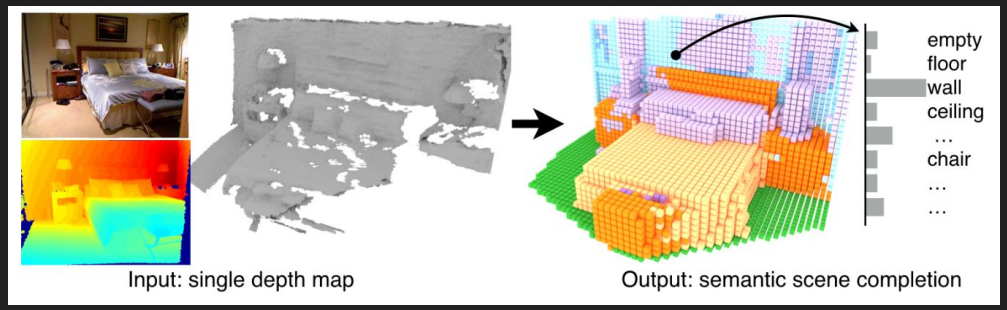
Example: Choosing a dataset for navigation

- 3D Semantic Parsing of Large-Scale Indoor Spaces (Stanford 2D-3D-S)
- Matterport3D : Learning from RGB-D Data in Indoor Environments
- Semantic Scene Completion from a Single Depth Image (SUNCG)

Matterport3D

10,800 panoramic views from
194,400 RGB-D images of 90
building-scale scenes
40 object categories
Real

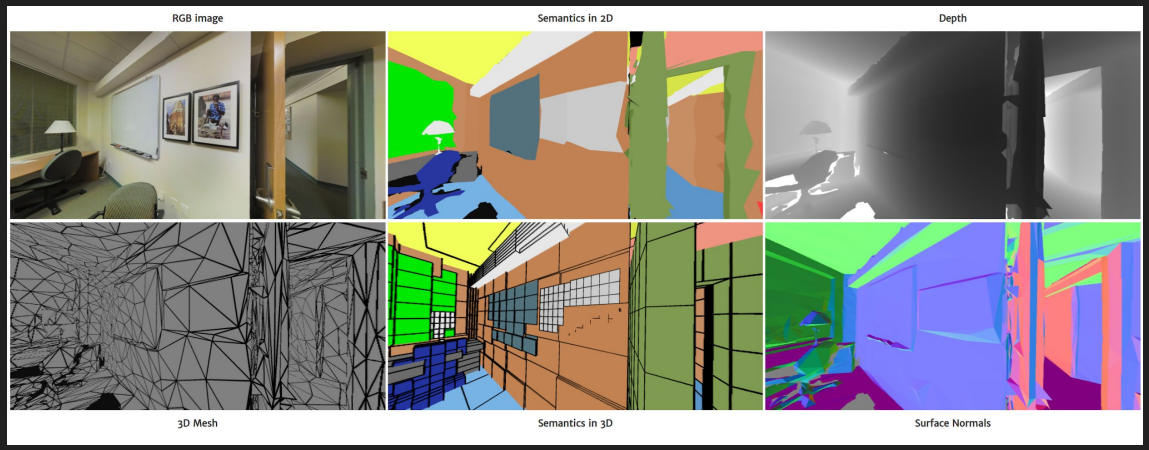
SUNCG



45,622 houses with 775,574 rooms
2644 unique object meshes covering 84 categories
Synthetic

Stanford 2D-3D-S

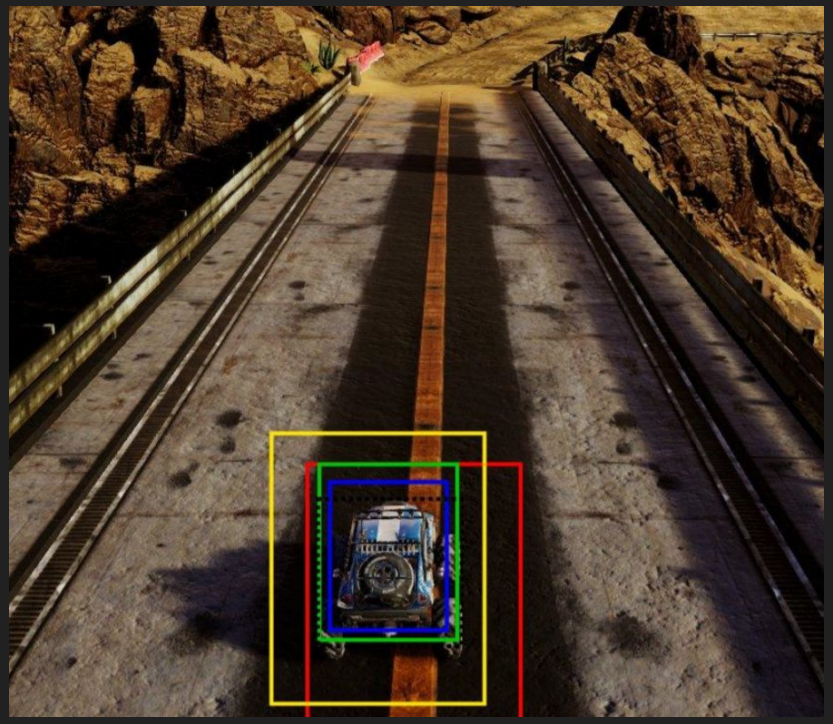
6 Areas
Over 70,000 RGB images
13 object categories
Real



Armeni, I., Sener, O., Zamir, A. R., Jiang, H., Brilakis, I., Fischer, M., & Savarese, S. (2016). 3D Semantic Parsing of Large-Scale Indoor Spaces. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1534–1543. http://doi.org/10.1109/CVPR.2016.170
Chang, A., Dai, A., Funkhouser, T., Savva, M., & Song, S. (n.d.). Matterport3D : Learning from RGB-D Data in Indoor Environments.
Song, S., Yu, F., Zeng, A., Chang, A. X., Savva, M., & Funkhouser, T. (2016). Semantic Scene Completion from a Single Depth Image, 1746–1754. http://doi.org/10.1109/CVPR.2017.28

Example: Dataset for UAV simulation

A benchmark and simulator for UAV tracking



UAV123

Provided with a benchmarking system and video dataset

123 Sequences, 110K frames

Bounding box and attribute annotation per frame

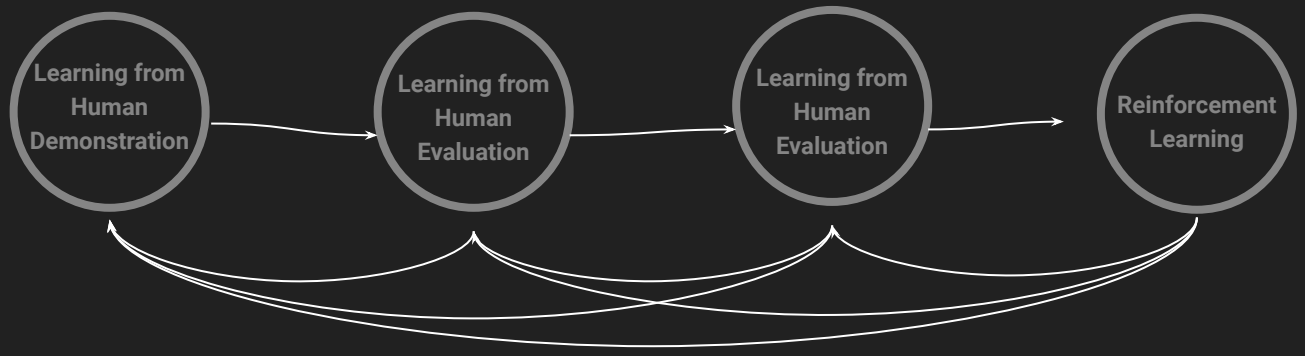
Compared four tracking algorithms as initial benchmark for accuracy of bounding box and success
STRUCK, MEEM, SAMF, SRDCF

Accurate real world drone data for training in simulation

Mueller, M., Smith, N., & Ghanem, B. (2016). A benchmark and simulator for UAV tracking. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 9905 LNCS, 445–461. http://doi.org/10.1007/978-3-319-46448-0_27

Human in the Loop

- How do we receive optimal training data?
- We can incorporate data from human operators into our training pipeline
  - Human centered reinforcement learning (HCRL)
- Collecting valid samples can be performed by:
  - Virtual reality teleoperated sessions
  - Mass consensus through mechanical turk
  - Having a human operator perform the task correctly



ROS Reality: collecting data in VR

Comparing Robot Grasping Teleoperation across Desktop and Virtual Reality with ROS Reality

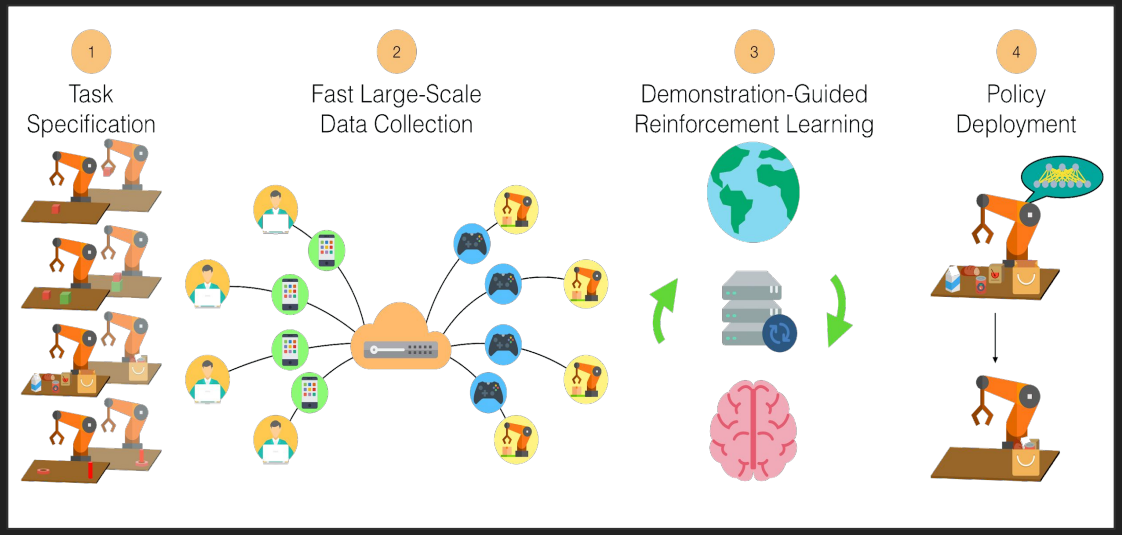
- Using remote teleoperation we can collect correct action sequences from a human operator
- Potential latency issues
- Allows for global data collection
- Currently reduced action space
- **VR allows for collecting ground truth data remotely using real/simulated robots**

Whitney, D., Rosen, E., Phillips, E., Konidaris, G., & Tellex, S. (2017). Comparing Robot Grasping Teleoperation across Desktop and Virtual Reality with ROS Reality. International Symposium on Robotics Research, 1–16. http://doi.org/10.4103/1817-1737.56008

Crowdsourcing RL Data

ROBOTURK: A Crowdsourcing Platform for Robotic Skill Learning through Imitation

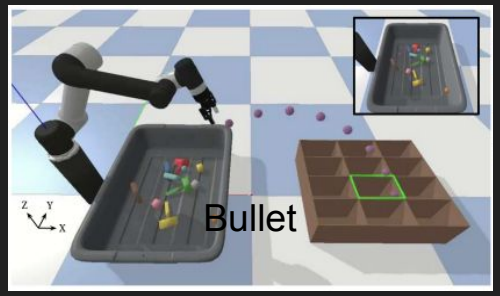
- Rapid crowdsourcing
- Allows for global data collection
- VR controller, phone, keyboard, and 3D mouse input options
- **We can collect ground truth data from people faster than ever**

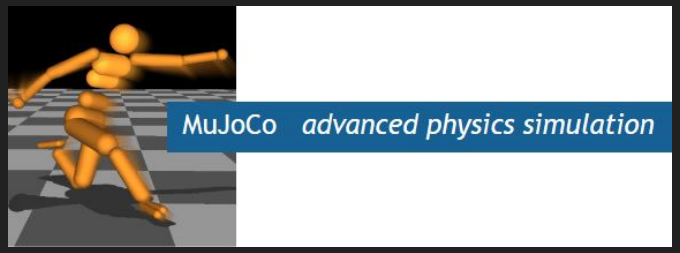


Mandlekar, A., Zhu, Y., Garg, A., Booher, J., Spero, M., Tung, A., … Fei-Fei, L. (2018). ROBOTURK: A Crowdsourcing Platform for Robotic Skill Learning through Imitation, (CoRL). Retrieved from http://vision.stanford.edu/pdf/mandlekar2018corl.pdf

Physics Simulators

Current physics backends





Comparison of different physics backends*

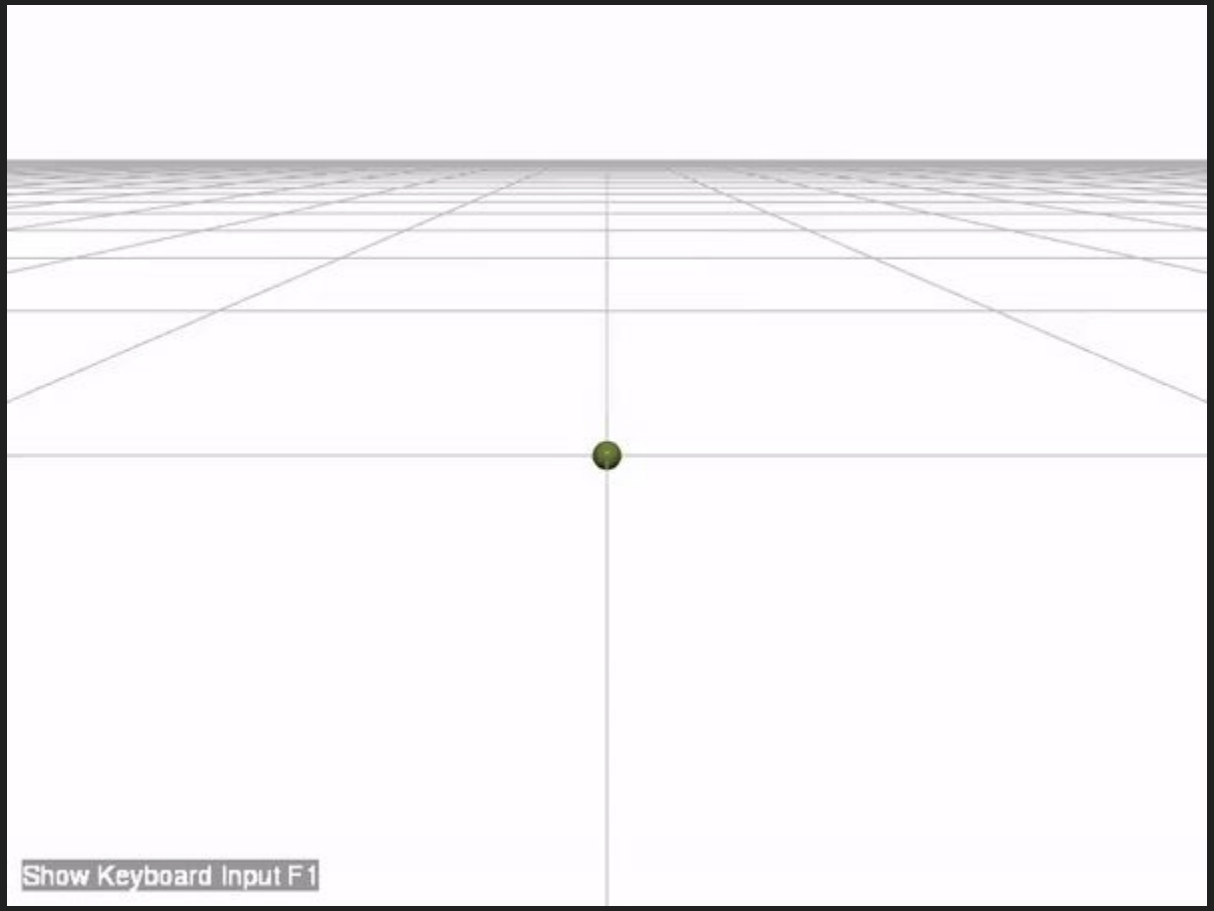
	Bullet	ODE	MuJoCo	DART	Gazebo
Initial release	2006	2001	2015	2012	2004
Author	E. Coumans	R. Smith	E. Todorov	J. Lee et al	N. Koenig
Language	C / C++	C++	C	C++	C/C++
API	C++ / Python	C	C	C++	C++/ROS
Contacts	Hard/Soft	Hard/Soft	Soft	Hard	N/A
Solver	MLCP	LCP	Newton / PGS / CG	LCP	N/A
Integrator	Semi-implicit Euler	Semi-implicit Euler	Semi-implicit Euler / RK4	Semi-implicit Euler	N/A

*https://leggedrobotics.github.io/SimBenchmark/

SimBenchmark Test

- Rolling test: friction model test
- Bouncing test: single-body elastic collision test
- 666 balls test: single-body hard contact test
- Elastic 666 balls test: single-body energy test
- ANYmal PD control test: articulated-robot-system speed test for quadrupedal robot
- ANYmal momentum test: articulated-robot-system momentum test
- ANYmal energy test: articulated-robot-system energy test

Example 666 Test



SimBenchmark Overall Results

	RaiSim	Bullet	ODE	MuJoCo	DART
Rolling	++	+++	-	+	-
Bouncing	++++	++	+++	-	+
666	+++	+	++	+	+
Elastic 666	++++	++	+++	-	+
ANYmal PD	+++++	+++	+	++++	++
ANYmal Momentum	+++	++	+++++	++++ (RK4) ++ (Euler)	+
ANYmal Energy	++++	+++	++	+++++ (RK4) +++ (Euler)	+

- more + is better
- **+**: best results
- -: cannot simulate due to inaccurate model or excepted

Physics backend benchmark

SimBenchmark

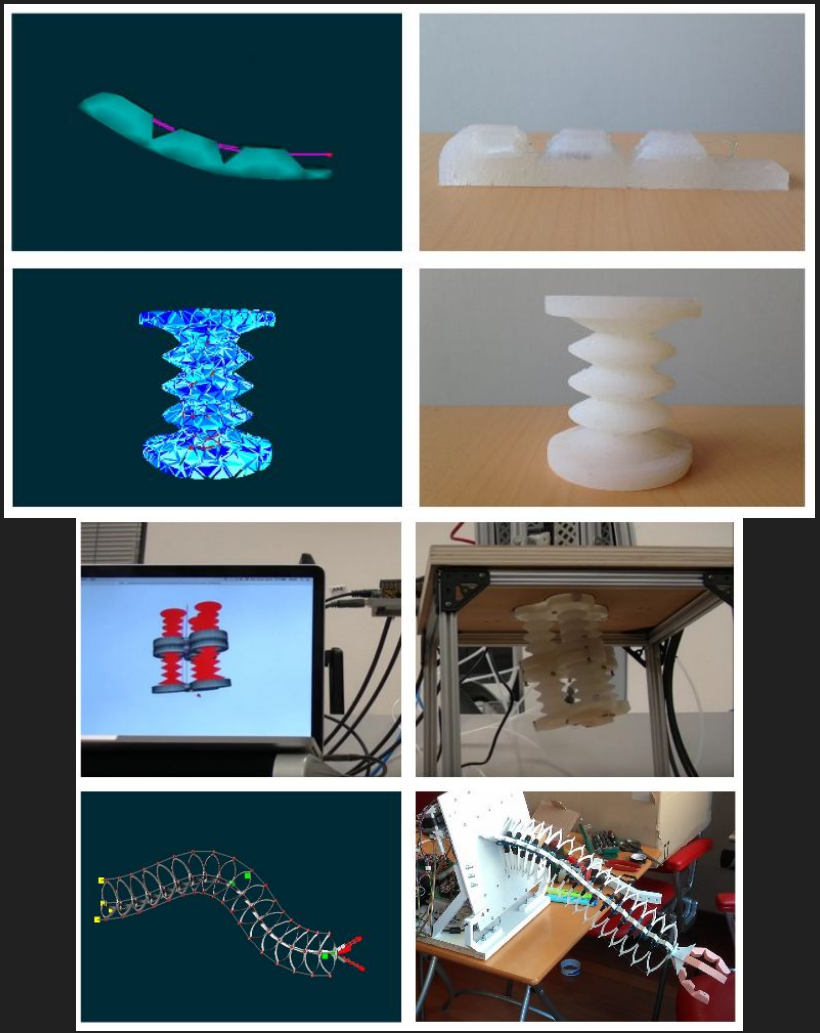
- DART's simulation pipeline is not suitable for simulation scenes with many objects - Very slow
- Bullet has severe position level drift without the post-solver error correcting method
- MuJoCo's soft contact model cannot control elasticity of contact.
- MuJoCo has consistent slip which requires additional post-process for legged robot simulation

- **Each simulator architecture has its own tradeoffs and needs to be evaluated for the problem at hand**

*https://leggedrobotics.github.io/SimBenchmark/

Soft body physics simulation: SOFA

Framework for online simulation of soft robots with optimization-based inverse mode



- Soft body actuation of materials
- Complex geometry can be slow but it is accurate
- Lack of accuracy metrics - all qualitative
- Close source plugin for open source simulation framework
- **Simulation of soft-body physics is now possible - however lacking integration**

Duriez, C., Coevoet, E., Largilliere, F., Morales-Bieze, T., Zhang, Z., Sanz-Lopez, M., … Dequidt, J. (2017). Framework for online simulation of soft robots with optimization-based inverse model. 2016 IEEE International Conference on Simulation, Modeling, and Programming for Autonomous Robots, SIMPAR 2016, 111–118. https://doi.org/10.1109/SIMPAR.2016.7862384

Domain Specific Simulators

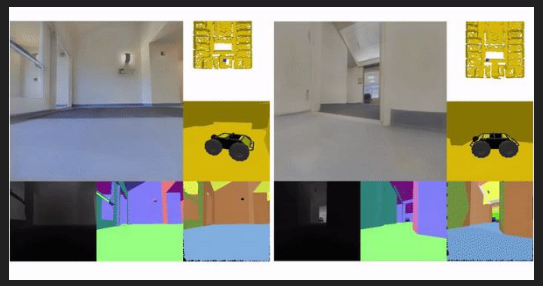
OpenRAVE : A Planning Architecture for Autonomous Robotics
Graspit: A versatile simulator for robotic grasping

- For grasping and trajectory planning
- GraspIt! Is a flexible grasping framework
- OpenRave is a flexible trajectory planning framework
- **Simulating trajectories and grasps makes real world grasp planning feasible**

Diankov, R., & Kuffner, J. (2008). OpenRAVE : A Planning Architecture for Autonomous Robotics. Tech. Rep. CMU-RI-TR-08-34, Robotics Institute, (July). http://doi.org/CMU-RI-TR-08-34
Miller, A. T., & Allen, P. K. (2004). Graspit: A versatile simulator for robotic grasping. IEEE Robotics and Automation Magazine, 11(4), 110–122. http://doi.org/10.1109/MRA.2004.1371616

Simulator and high level APIs

Gibson Env : Real-World Perception for Embodied Agents
MINOS : Multimodal Indoor Simulator
Habitat: A Platform for Embodied AI Research



Gibson

Produced in 2018
Support for
Matterport3D/Stanford 2D3DS
512x512 resolution

Produced in 2017
Support for
Matterport3D/SUNCG
Arbitrary resolution

Habitat

Produced by Facebook in 2019
Support for Matterport3D/SUNCG
10K fps
512x512 resolution

High level APIs are new and allow for simulating RL and data collection tasks easier than before

Xia, F., & Sax, A. (n.d.). Gibson Env : Real-World Perception for Embodied Agents.
Savva, M., Chang, A. X., Dosovitskiy, A., & Funkhouser, T. (n.d.). MINOS : Multimodal Indoor Simulator, 1–14.
Savva, M., Kadian, A., Maksymets, O., Zhao, Y., Wijmans, E., Jain, B., … Batra, D. (2019). Habitat: A Platform for Embodied AI Research. Retrieved from http://arxiv.org/abs/1904.01201

Transferring agents to real-world: Gibson

Gibson Env : Real-World Perception for Embodied Agents

- Imperfections in rendering can make it difficult to get photo-realistic images
- Training a network structure to map simulated image onto real world image
- Provided with open source implementation
- **We can transfer the style of simulated views onto realistic images**

Xia, F., & Sax, A. (n.d.). Gibson Env : Real-World Perception for Embodied Agents.

Vehicle/Drone Simulation: Airsim

AirSim: High-Fidelity Visual and Physical Simulation for Autonomous Vehicles

- Drone simulation is done using Airsim
- Allows for real-time simulation of drones and cars
- Synthetic data from games in the unreal engine
- Plans to use PhysX in the future but uses Unreal Engine for now

Shah, S., Dey, D., Lovett, C., & Kapoor, A. (2017). AirSim: High-Fidelity Visual and Physical Simulation for Autonomous Vehicles, 1–14. http://doi.org/10.1007/978-3-319-67361-5_40

Deep learning speeds up simulation

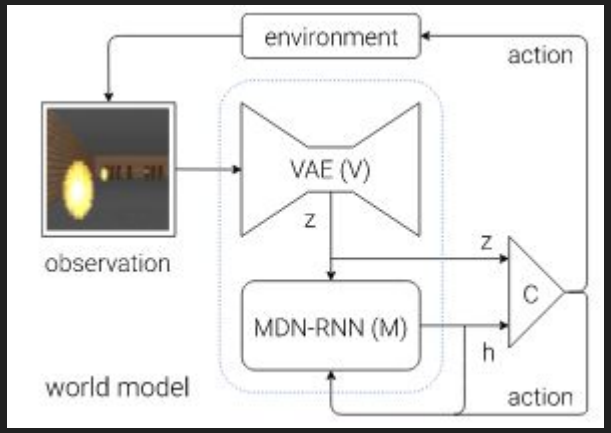
Learned Simulators

- Modern advancements in deep learning have allowed for the creation of learned representations of world models
- Learned engines allow for faster simulation of world mechanics and often can offer more accuracy in the simulated result if the training data is derived from the ultimate medium the agent will be acting in

- **We can speedup our simulation pipelines**

Learned simulator representations can allow for faster learning: World Models

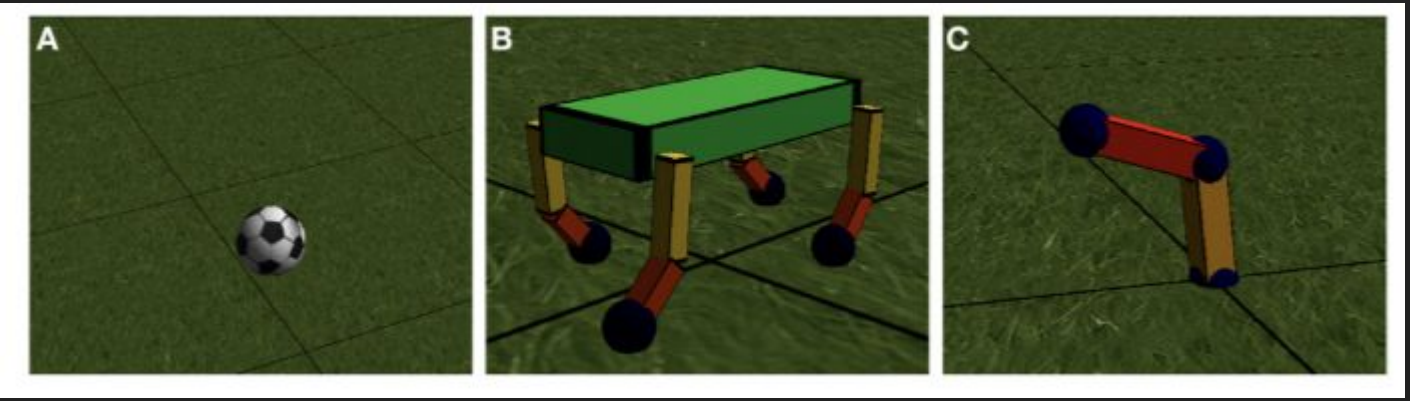
- Map the action and state of simulator onto an RNN architecture
- An agent controls an avatar in a video game capturing the state and action of each frame
- Allows for rapid learning in approximated physics environment
- Tested specifically with Doom
- **Learned simulation model is faster than running the simulation itself**



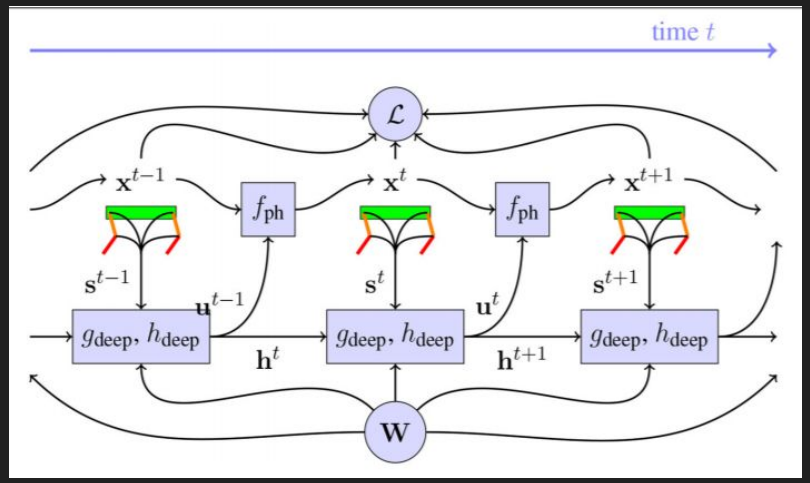
Ha, D., & Schmidhuber, J. (2018). World Models. http://doi.org/10.5281/zenodo.1207631

Differentiable physics engines allow for faster learning

A Differentiable Physics Engine for Deep Learning in Robotics,

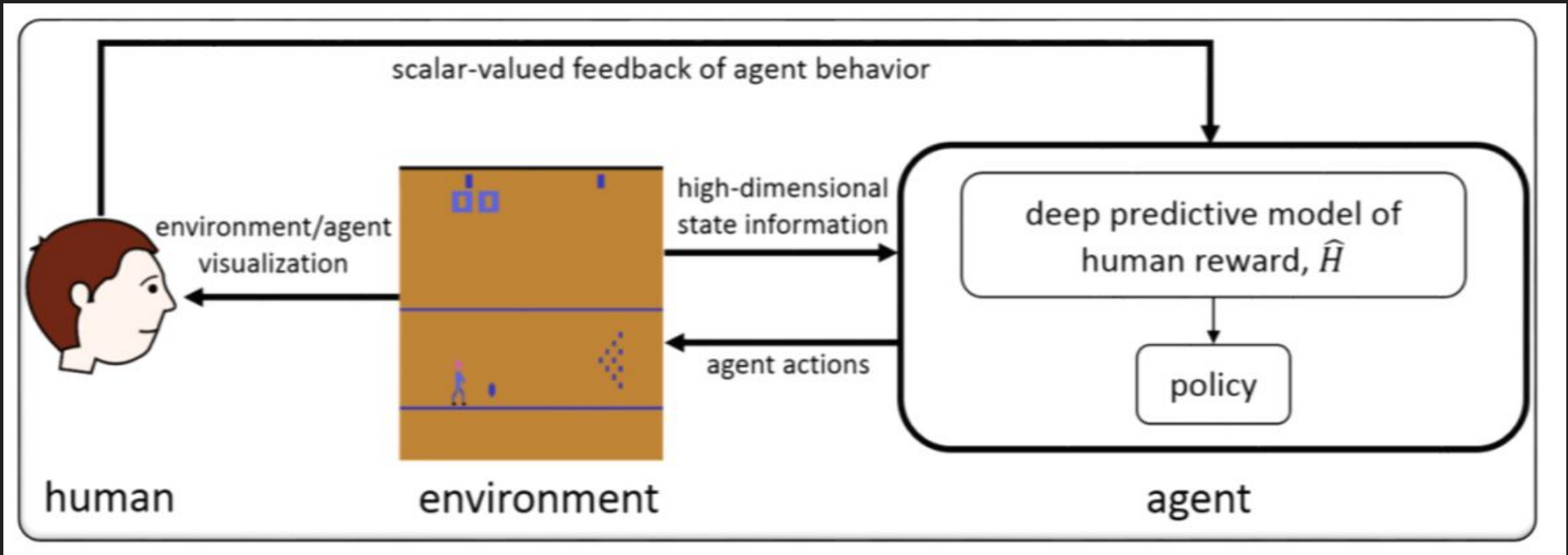


- Textures, bodies, and state are all differentiable
- Integrated in either forward or backwards propagation
- **Textures, physics, state can be integrated into the reward function of simulations**



Degrave, J., Hermans, M., Dambre, J., & wyffels, F. (2016). A Differentiable Physics Engine for Deep Learning in Robotics, 13(March), 1–9. http://doi.org/10.3389/fnbot.2019.00006

Deep Tamer: Using limited human trials reduces training time



- Using few human actions to decrease training time
- Training time in 15 minutes of data vs similar agents
- Exceeds human operator in 7 minutes of training
- **We can reduce time spent simulating by using human-in-the-loop ground truth data for deep learned agents**

Warnell, G., Waytowich, N., Lawhern, V., & Stone, P. (2017). Deep TAMER: Interactive Agent Shaping in High-Dimensional State Spaces. Retrieved from http://arxiv.org/abs/1709.10163

Sensory Modalities

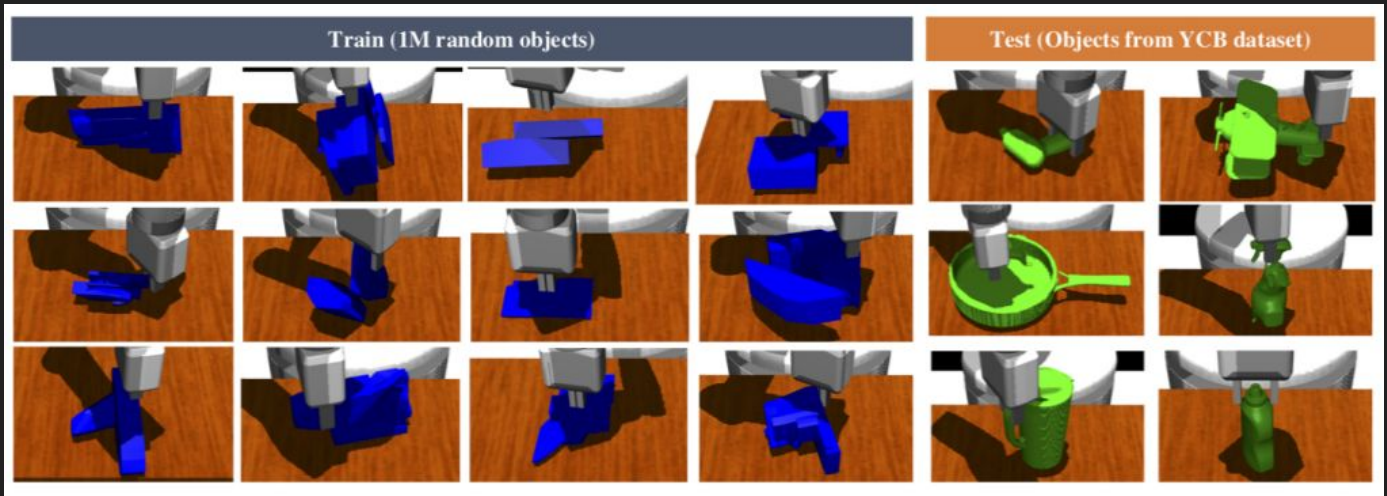
- Physics engines have certain sensory modalities pre-programmed in
- All have some form of RGB rendering scheme
- Less common
  - Depth rendering
  - Semantic labeling
  - Tactile
  - Accurate force models
  - Sound
  - Human in the loop

- Lacking sensory modalities does not mean a simulator is unusable but is a tradeoff that must be considered

Improving sim-to-real fidelity

GANs allow for better grasp generalization

Domain Randomization and Generative Models for Robotic Grasping



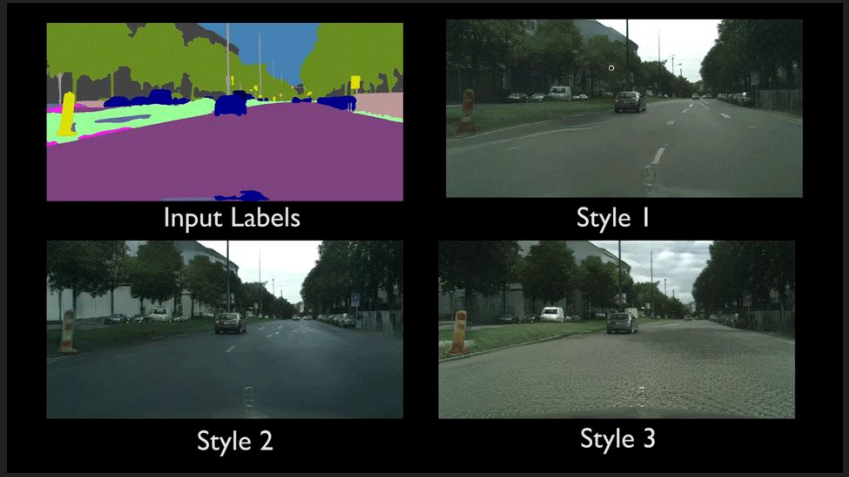
- Train model to generate models based on YCB object dataset
- Trained grasp planner in simulation on generated synthetic objects
- **Generating objects based on a real-world dataset gives us more data to train another network or test an existing pipeline with**

Tobin, J., Biewald, L., Duan, R., Andrychowicz, M., Handa, A., Kumar, V., … Abbeel, P. (2017). Domain Randomization and Generative Models for Robotic Grasping. http://doi.org/10.1017/CBO9781107415324.004

Photorealistic video synthesis improves sim-to-real

Video-to-Video Synthesis

- 2048x1024 photorealistic video-to-video translation using GANs
- It can be used for turning semantic label maps into photo-realistic videos, synthesizing people talking from edge maps, or generating human motions from poses
- Using simulated datasets we can perform style transfer for relevant task datasets
- **Transferring style based solely on semantic labels lets us map simulated data to real data**

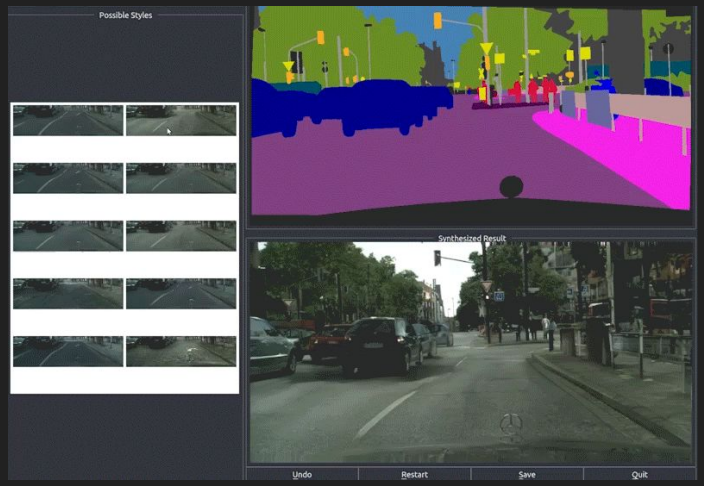


Wang, T.-C., Liu, M.-Y., Zhu, J.-Y., Liu, G., Tao, A., Kautz, J., & Catanzaro, B. (2018). Video-to-Video Synthesis, 1–14. http://doi.org/arXiv:1808.06601v1

More synthesized data means better models

High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs

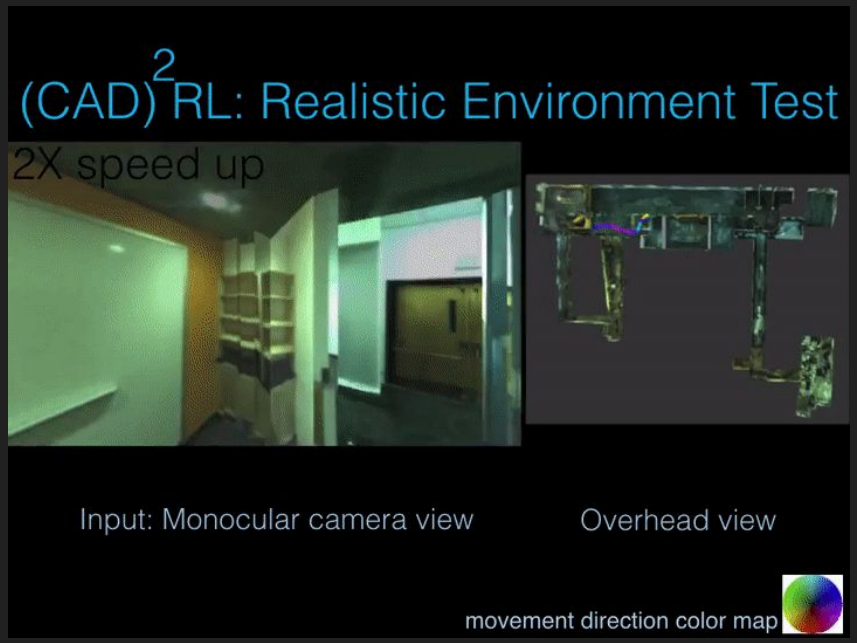
- 2048x1024 photorealistic video-to-video translation using GANs
- Similar to video-to-video synthesis but also allows for editing of style
- Using simulated datasets we can perform style transfer for relevant task datasets
- **Transferring style based solely on semantic labels lets us map simulated data to real data**
- **New User Interfaces make domain mapping easier**



Corporation, T.-C. W. M.-Y. L. J.-Y. Z. A. T. J. K. B. C. NVIDIA, & Berkeley, UC. (2017). High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs. ACM Transactions on Speech and Language Processing, 8(2), 1–18. http://doi.org/10.1145/2050100.2050101

Examples of Robotic Simulation Pipelines

CAD2RL: Utilizing drone simulation for automated flight navigation



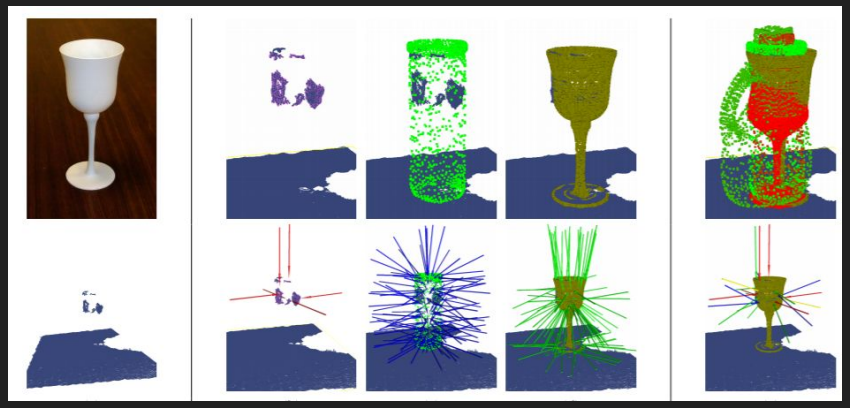
- Real Single-Image Flight without a Single Real Image - synthetic images
- Drone simulation detects collision with objects
- Environment has randomized textures for scenes
- Translated onto a real drone flying through an environment
- **We can train a navigation task for a drone in simulation**

Sadeghi, F., & Levine, S. (2016). CAD2RL: Real Single-Image Flight without a Single Real Image. http://doi.org/10.15607/RSS.2017.XIII.034

Using pose detection enhances grasp planning via ground truth mesh models

Collaborative grasp planning with multiple object representations

- **Great example of real-to-sim-to-real pipeline for grasping task**
- 25 household objects were used in generation of grasps - 45 extra 3D models
- Uses pose detection model and then simulated annealing in GraspIt! for grasp planning

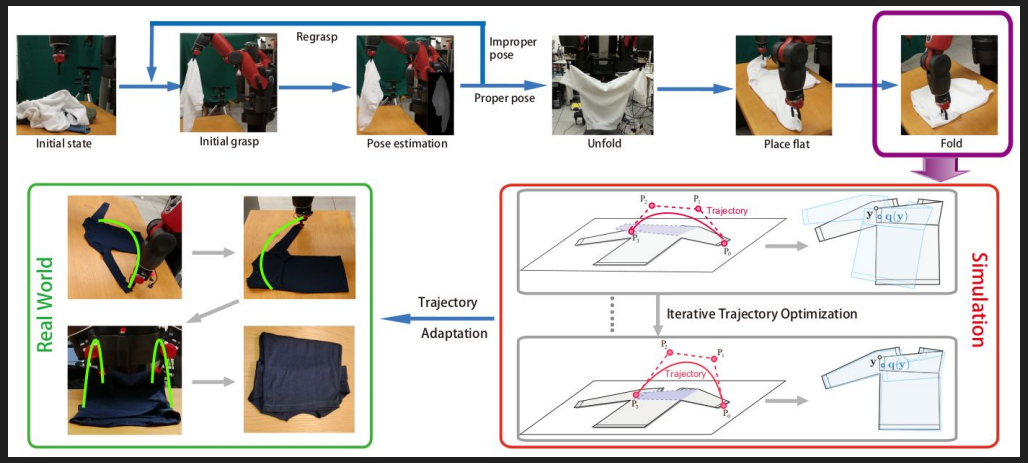


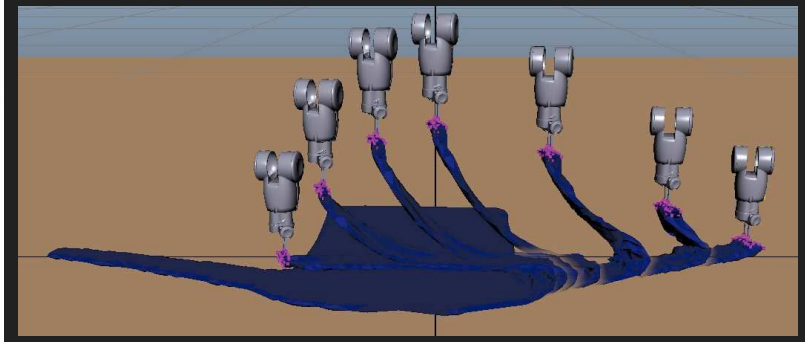
Brook, P., Ciocarlie, M., & Hsiao, K. (2011). Collaborative grasp planning with multiple object representations. Proceedings - IEEE International Conference on Robotics and Automation, 2851–2858. http://doi.org/10.1109/ICRA.2011.5980490

Simulation of deformable objects aids in planning folding tasks

Folding Deformable Objects using Predictive Simulation and Trajectory Optimization

- **Great example of real-to-sim-to-real pipeline for folding task**
- Approximated geometry of clothes using 2D image
- Used maya for mesh simulation
- Folding actions were successful for optimal clothing placement





Li, Y., Yue, Y., Xu, D., Grinspun, E., & Allen, P. K. (2015). Folding Deformable Objects using Predictive Simulation and Trajectory Optimization, 6000–6006.

Using simulated object renderings transfers better into the real world

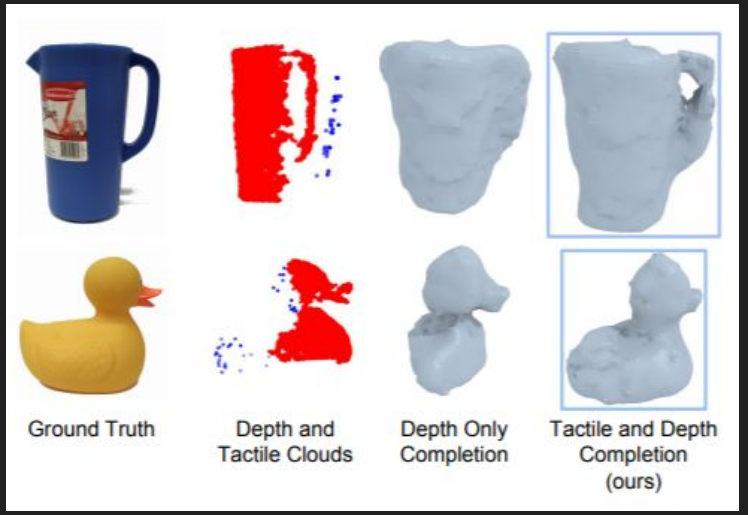
Multi-Modal Geometric Learning for Grasping and Manipulation

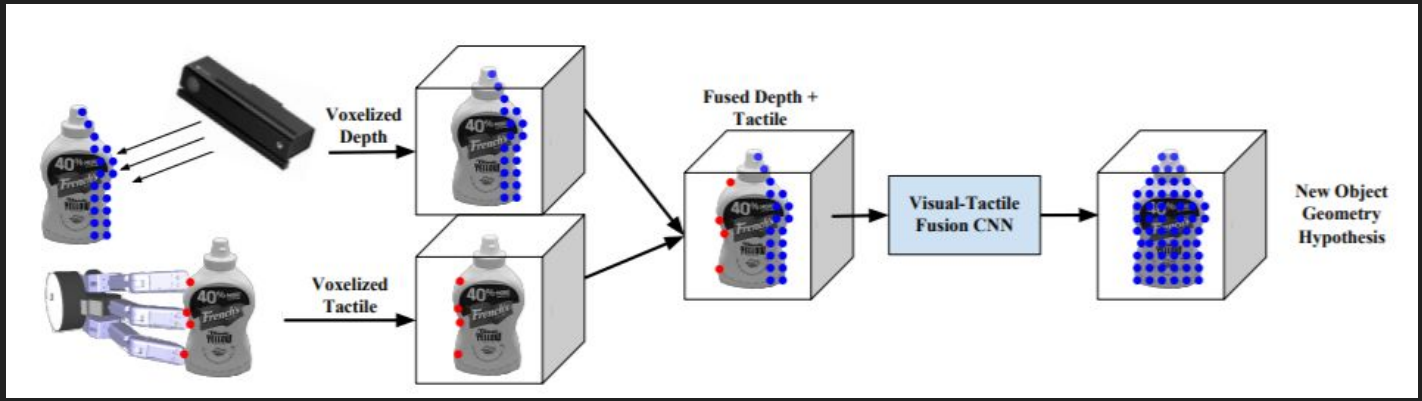
Grasping was enabled through simulated point cloud acquisition

Synthetic tactile points were collected

Use of the YCB and GRASP datasets

Great example of real-to-sim-to-real pipeline with multiple sensory modalities

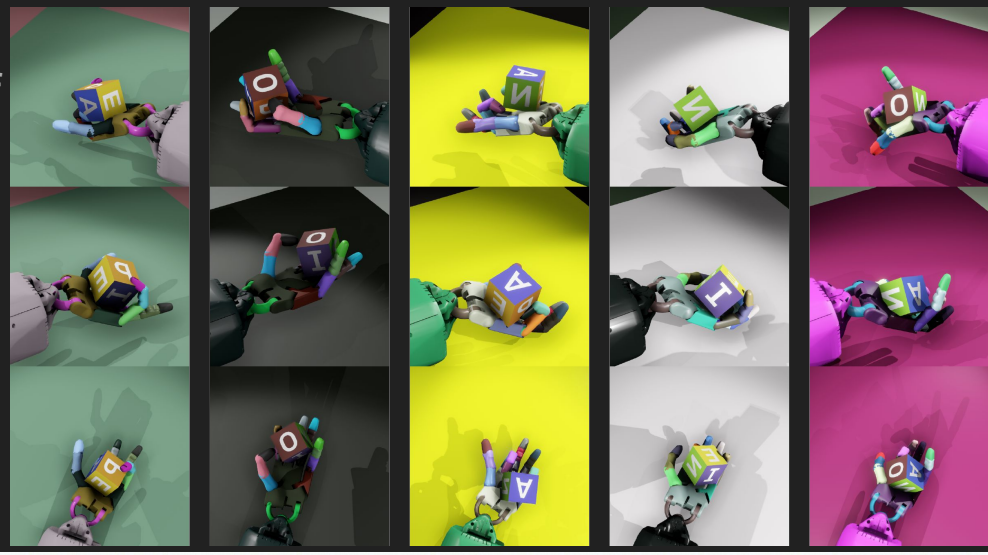


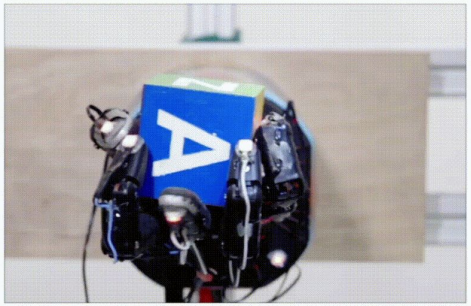


Varley, J., Watkins-Valls, D., & Allen, P. (2018). Multi-Modal Geometric Learning for Grasping and Manipulation. http://doi.org/arXiv:1803.07671v2

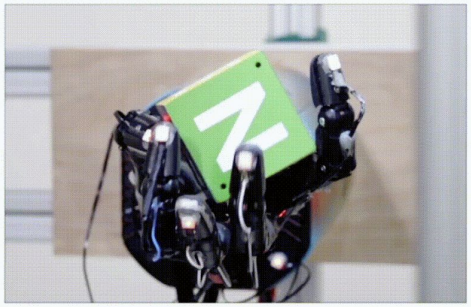
Learning Dexterous In-Hand Manipulation: Reducing the domain of the task in simulation allows for better transfer

- Reducing the domain complexity of the task allowed for easy transfer of the action onto the real world
- Shadow hand model in simulation was accurate
- **By reducing the domain you can decrease training time and have a higher likelihood of mapping to real-world**

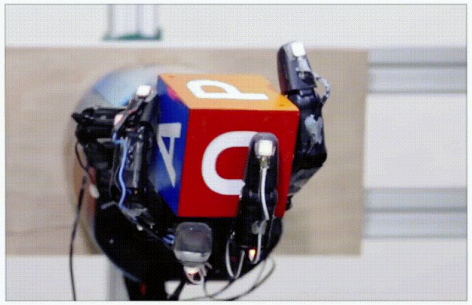




FINGER PIVOTING



SLIDING



FINGER GAITING

OpenAI, :, Andrychowicz, M., Baker, B., Chociej, M., Jozefowicz, R., … Zaremba, W. (2018). Learning Dexterous In-Hand Manipulation, 1–27. http://doi.org/arXiv:1808.00177v2

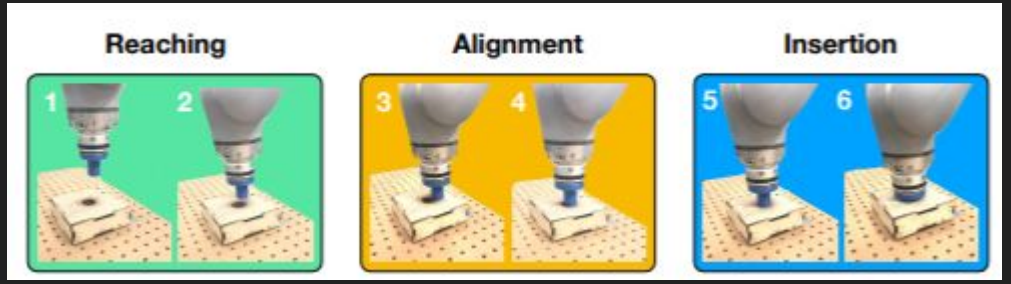
Using action primitives allows for better transfer onto the real world

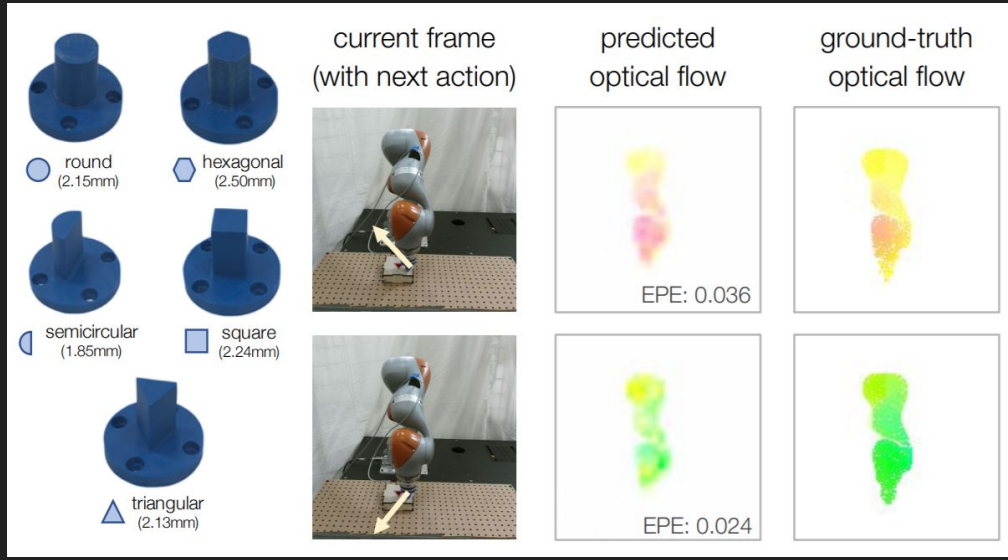
Making Sense of Vision and Touch: Self-Supervised Learning of Multimodal Representations for Contact-Rich Tasks

Action primitives are reaching, alignment, and insertion

Stated in paper that they would like to integrate more complex tasks

Using action primitives makes it easier to generalize to the real world





Lee, M. A., Zhu, Y., Srinivasan, K., Shah, P., Savarese, S., Fei-Fei, L., … Bohg, J. (2018). Making Sense of Vision and Touch: Self-Supervised Learning of Multimodal Representations for Contact-Rich Tasks. http://doi.org/arXiv:1810.10191v1

Using modular object detection and grasping priors allows for better real world transfer

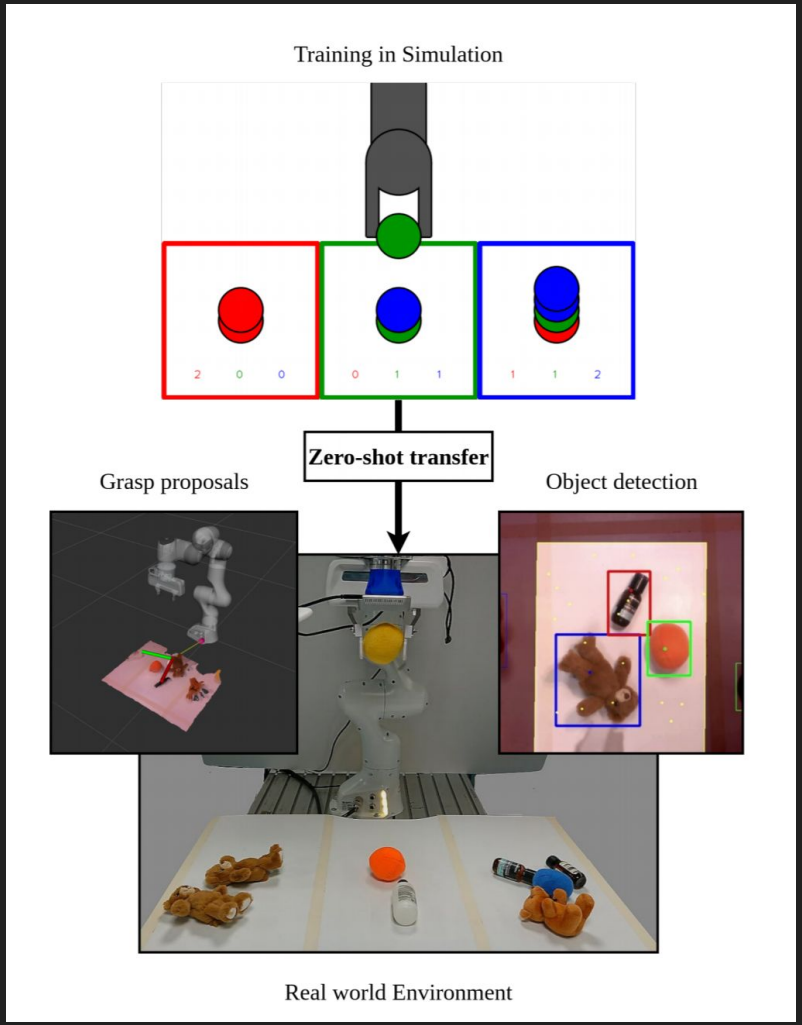
Zero-shot Sim-to-Real Transfer with Modular Priors

Simple manipulation task and sorting tasks

Pre-trained object detector

Deep sets encoder that enables reinforcement learning - effectively reducing domain space

Modular systems make real-to-sim-to-real more effective



Lee, R., Mou, S., Dasagi, V., Bruce, J., Leitner, J., & Sünderhauf, N. (2018). Zero-shot Sim-to-Real Transfer with Modular Priors. Retrieved from http://arxiv.org/abs/1809.07480

Looking Ahead

- We need better ways of simulating soft body robots
  - SOFA is a closed source solution but open source likely soon
- Need larger datasets of real world scenes applicable to robotics
  - Matterport3D is good for home based navigation tasks, but there are many more complex environments that are difficult to simulate
- High level APIs need to be more accessible
  - Habitat-Sim is a step in the right direction which is endorsed by Facebook

Looking Ahead

- VR integration into frameworks is in production now
  - PyBullet is developing a plugin for VR
  - Unity has support for VR
- ROS 2 and MoveIt! 1.0
  - Solving simulator integration into deep learning tasks and generalizing API better than before
- Better domain randomization
- Better and easier to use APIs for simulation e.g. OpenAI
- Industry support is increasing for simulators
  - Facebook and Habitat Sim
  - Microsoft and ROS/Gazebo

Looking Ahead

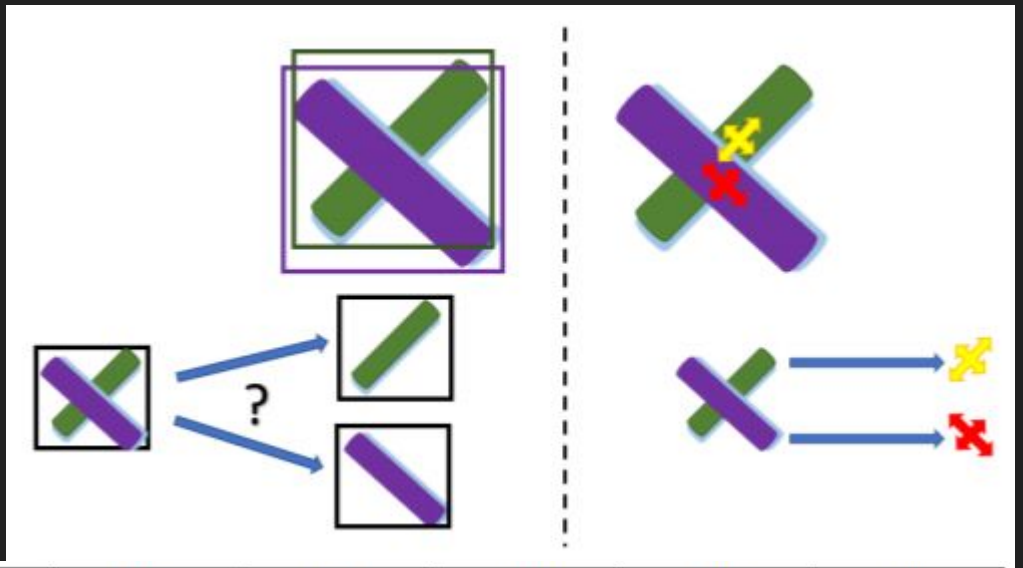
- Creating custom sensory modalities is easier than it was before
- GANs and domain randomization produce larger datasets
- Using deep learning for speeding up networks is very new and should be researched further

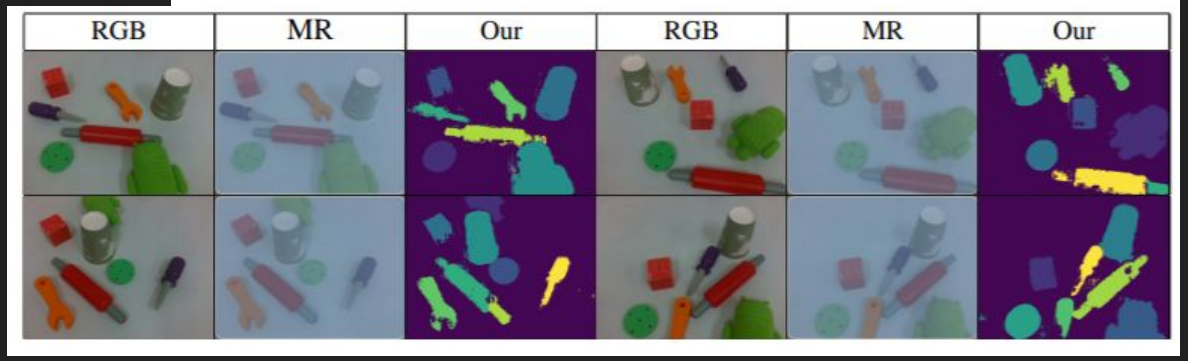
Synthetic domain randomized object segmentation translates onto real world grasping tasks

Grasping was enabled through synthetic cloud generation

Segmentation was free given models in simulation

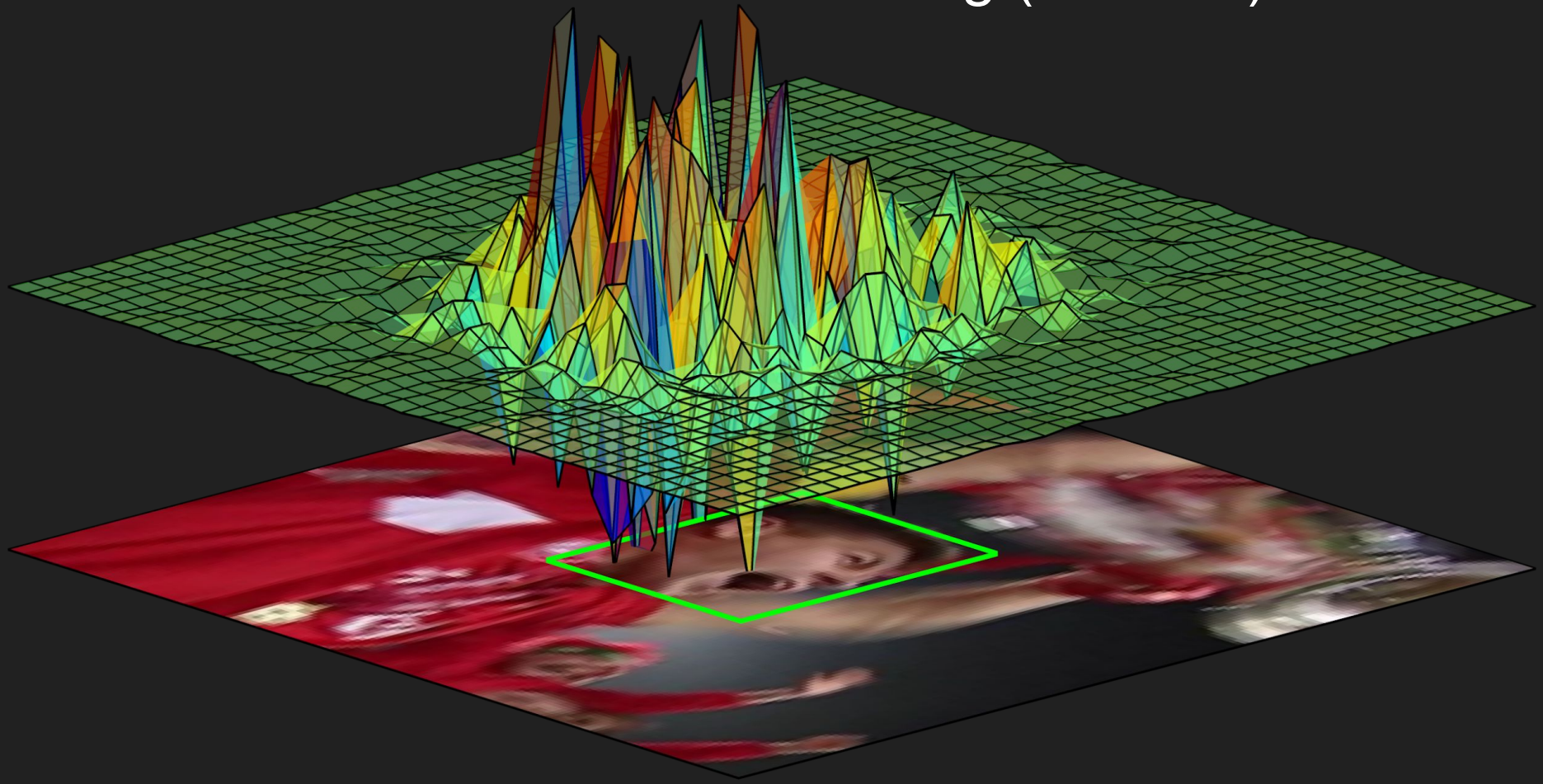
Able to cluster objects using their COM



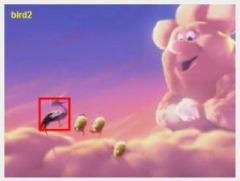


Shao, L., Tian, Y., & Bohg, J. (2018). ClusterNet: 3D Instance Segmentation in RGB-D Images, (1). http://doi.org/arXiv:1807.08894v2

UAV Tracking Algorithms: Spatially regularized correlation filters for visual tracking (SRDCF)



Visual Tracker Benchmark (OTB50)

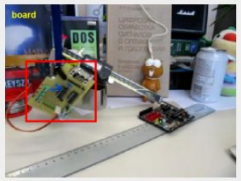


Bird2
OCC, DEF, FM, IPR, OPR

BlurCar1
MB, FM

BlurCar3
MB, FM

BlurCar4
MB, FM



Board
SV, MB, FM, OPR, OV, BC

Bolt2
DEF, BC

Boy
SV, MB, FM, IPR, OPR

Car2
IV, SV, MB, FM, BC

Car24
IV, SV, BC

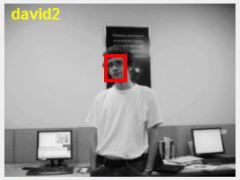
Coke
IV, OCC, FM, IPR, OPR, BC

Coupon
OCC, BC

Crossing
SV, DEF, FM, OPR, BC

Dancer
SV, DEF, IPR, OPR

Dancer2
DEF



David2
IPR, OPR

David3
OCC, DEF, OPR, BC

Dog
SV, DEF, OPR

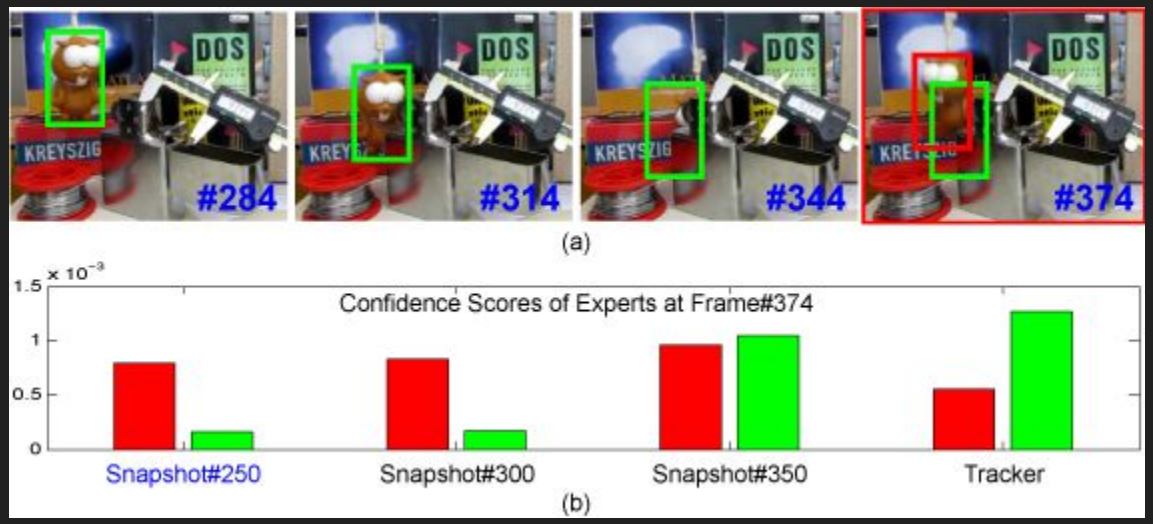
Dog1
SV, IPR, OPR

Doll
IV, SV, OCC, IPR, OPR

FaceOcc1
OCC

FaceOcc2
IV, OCC, IPR, OPR

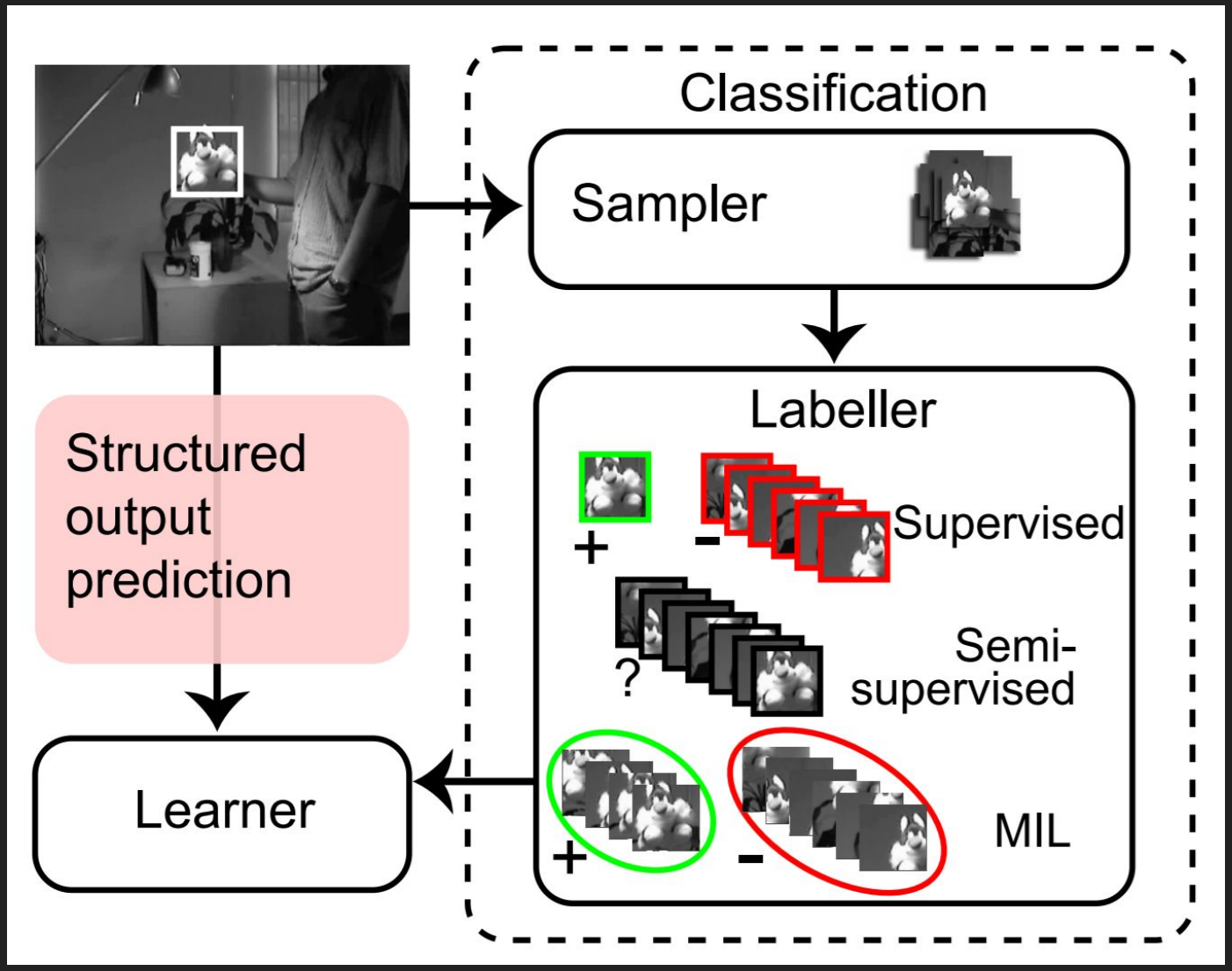
MEEM: Robust Tracking via Multiple Experts using Entropy Minimization



A Scale Adaptive Kernel Correlation Filter Tracker with Feature Integration (SAMF)

- Tracker based on the correlation filter framework
- Adaptive scale of filtering to track objects Moreover, the powerful features including HoG and color-naming are integrated together to further boost the overall tracking performance.
- Our method successfully tracked the targets in about 72% videos and outperformed the state-of-the-art trackers on the benchmark dataset with 51 sequences

Struck: Structured Output Tracking with Kernels



LCP: Linear Complementarity Problem

Given a real matrix *M* and vector *q*, the linear complementarity problem LCP(*M*, *q*) seeks vectors *z* and *w* which satisfy the following constraints:

- $w, z \geqslant 0,$ (that is, each component of these two vectors is non-negative)
- $z^T w = 0$ or equivalently $\sum_i w_i z_i = 0.$ This is the complementarity condition, since it implies that, for all i, at most one of w_i and z_i can be positive.
- $w = Mz + q$

MLCP: Mixed Linear Complementarity Problem

- Same principle as MLCP but allows for both of the variables to be free
- Global LCP solvers can work, but not global linear solvers, if LCPs are mixed into the system. A mixed-linear complementary problem (MLCP) is a system of both linear and complementary problems.

$$\begin{aligned} \mathbf{y} &= \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z} + \mathbf{b} \\ \mathbf{w} &= \mathbf{C}\mathbf{x} + \mathbf{M}\mathbf{z} + \mathbf{q} \end{aligned}$$

$$\mathbf{w}, \mathbf{z} \geq 0,\ \mathbf{w}^T\mathbf{z} = 0,\ \mathbf{x}\ \text{free},\ \mathbf{y} = 0$$

Projected Gauss–Seidel method (PGS)

The Gauss–Seidel method is an iterative technique for solving a square system of *n* linear equations with unknown $\mathbf{x}$:

$$A\mathbf{x} = \mathbf{b}.$$

It is defined by the iteration

$$L_* \mathbf{x}^{(k+1)} = \mathbf{b} - U\mathbf{x}^{(k)},$$

where $\mathbf{x}^{(k)}$ is the *k*th approximation or iteration of $\mathbf{x}$, $\mathbf{x}^{(k+1)}$ is the next or *k* + 1 iteration of $\mathbf{x}$, and the matrix *A* is decomposed into a lower triangular component L_*, and a strictly upper triangular component *U*: $A = L_* + U$.[2]

Semi-implicit Euler

The semi-implicit Euler method can be applied to a pair of differential equations of the form

$$\frac{dx}{dt} = f(t, v)$$

$$\frac{dv}{dt} = g(t, x),$$

where *f* and *g* are given functions. Here, *x* and *v* may be either scalars or vectors. The equations of motion in Hamiltonian mechanics take this form if the Hamiltonian is of the form

$$H = T(t, v) + V(t, x).$$

The differential equations are to be solved with the initial condition

$$x(t_0) = x_0, \qquad v(t_0) = v_0.$$