https://deeprob.org/w26/

ROB 498/599: Deep Learning for Robot Perception (DeepRob)

Lecture 14:

More on Robot Grasping models;

(Previous) Final Project Showcase;

Great Lakes Compute Resources

02/25/2026

M | ROBOTICS

Today

- Feedback and Recap (5min)
- More on robot grasping (25min)
- Final Project Showcase (30min)
- Great Lakes Computing Resources (15min)
- Summary and Takeaways (5min)

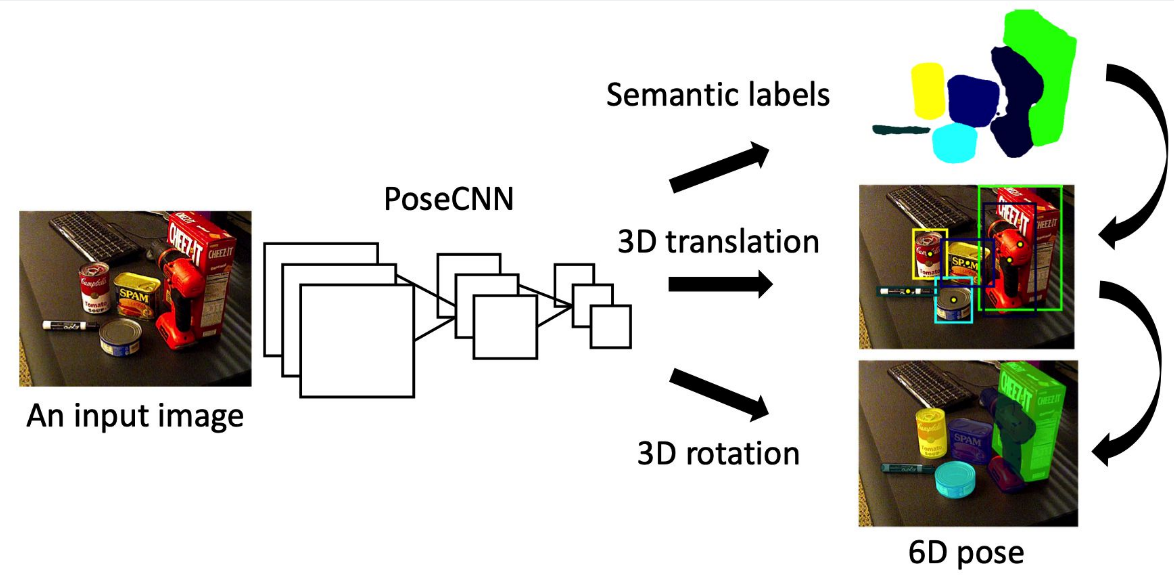
Final Project Group Sign-Up (3-4 persons per group):
https://docs.google.com/spreadsheets/d/1Jjv6vBQwAtCpUSQDAx0Gj5pDCWWqcvSQEeBFIF5prqo/edit?usp=sharing

M | ROBOTICS

More on Robotic Grasping

Recap: Pose CNN

(Monday Feb.23 lecture - will be useful in P4)



Robotic Grasping

robotic manipulation

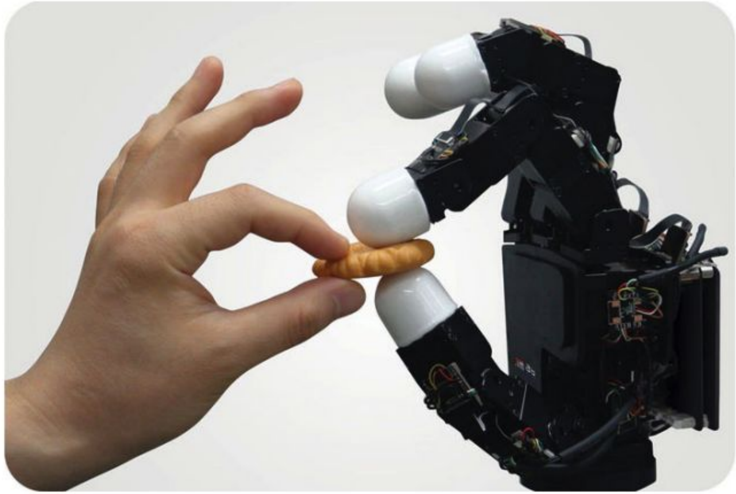


Figure 5.1: The Allegro Hand. Image retrieved from `wiki.wonikrobotics.com`.

Definition (Grasp):
A grasp is an act of restraining an object's motion through application of forces and torques at a set of contact points.

Robotic Grasping

robotic manipulation

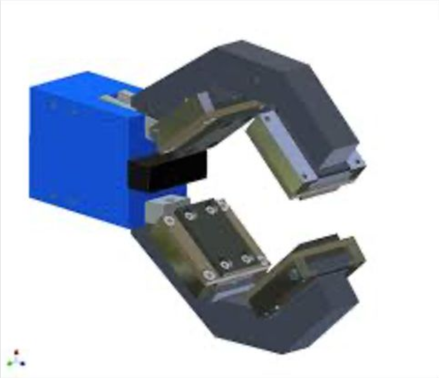
Challenges (Grasp):

- High-dimensional configuration of the gripper
- Choosing contact points can be difficult
- Ensure robot doesn't collide with environment
- Evaluate grasp quality (robust grasp, uncertainty)

Robotic Grasping - End Effectors

Parallel Gripper

https://onrobot.com/en/products/2fg7



Jaw Gripper

https://www.agi-automation.com/design-guidelines-for-pneumatic-gripper/

Dexterous Hand Gripper

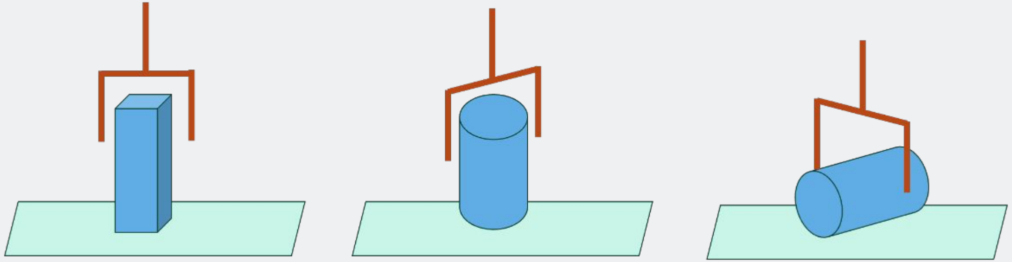
https://www.shadowrobot.com/

Suction Gripper

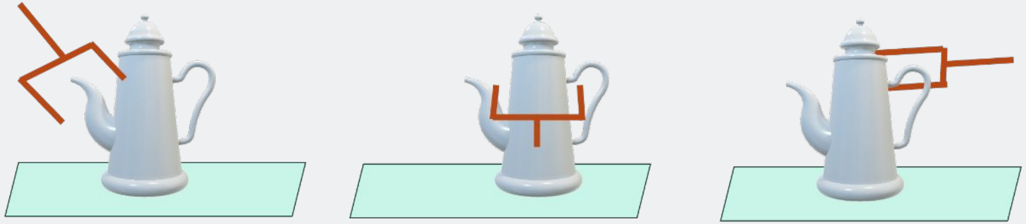
https://test.tm-robot.com/en/product/robotiq-vacuum-gripper-epick/

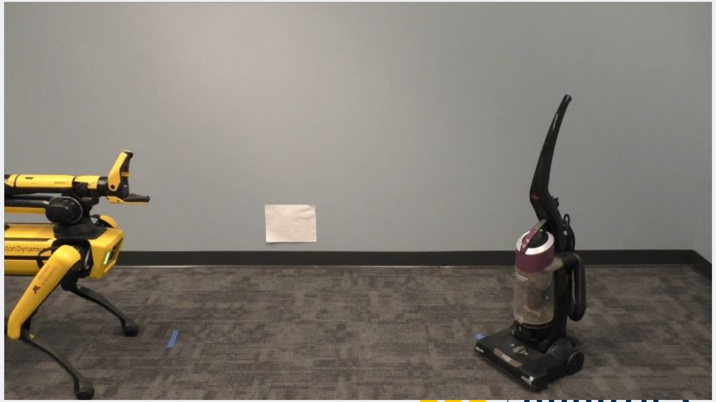
Robotic Grasping - Grasp Pose

- Grasping in SE(2) pose



- Grasping in SE(3) pose





Robotic Grasping - SE(2) Pose

Action Direction:

Top-down

Input:

RGB-D images, point cloud

Output:

(x, y, θ)

- Location
- Rotation angle

M | ROBOTICS

Robotic Grasping - SE(3) Pose

Action Direction:

any 3D direction

Input:

volumetric representations (mesh, point cloud, TSDF, etc.)

Output:

(R , t)

- Rotation
- Translation

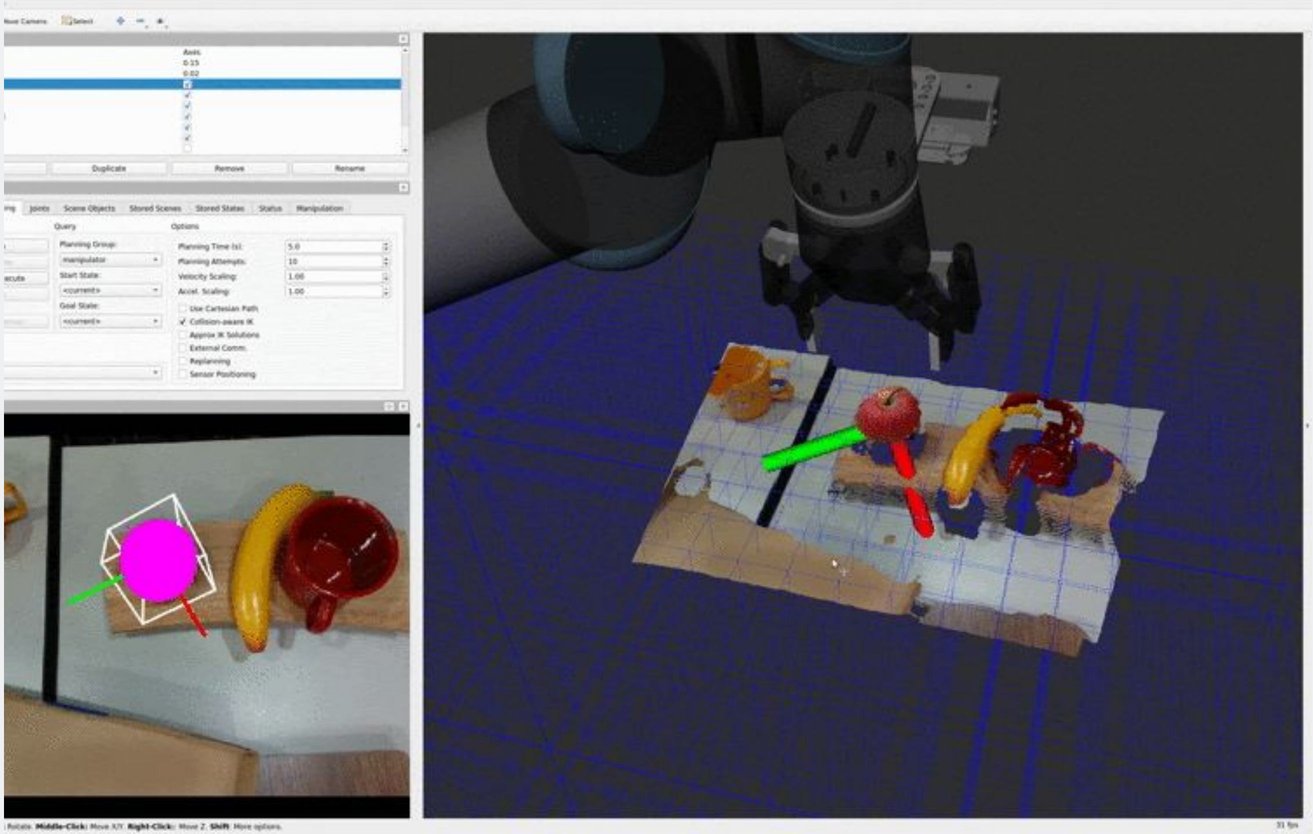
Robotic Grasping - SE(2) Pose

- Supersizing Self-supervision: Learning to Grasp from 50K Tries and 700 Robot Hours https://arxiv.org/pdf/1509.06825v1
- Dex-Net 2.0: Deep Learning to Plan Robust Grasps with Synthetic Point Clouds and Analytic Grasp Metrics https://arxiv.org/abs/1703.09312
- Sample Efficient Grasp Learning Using Equivariant Models https://arxiv.org/pdf/2202.09468
- Grasping with kirigami shells https://www.science.org/doi/10.1126/scirobotics.abd6426

Robotic Grasping - SE(3) Pose

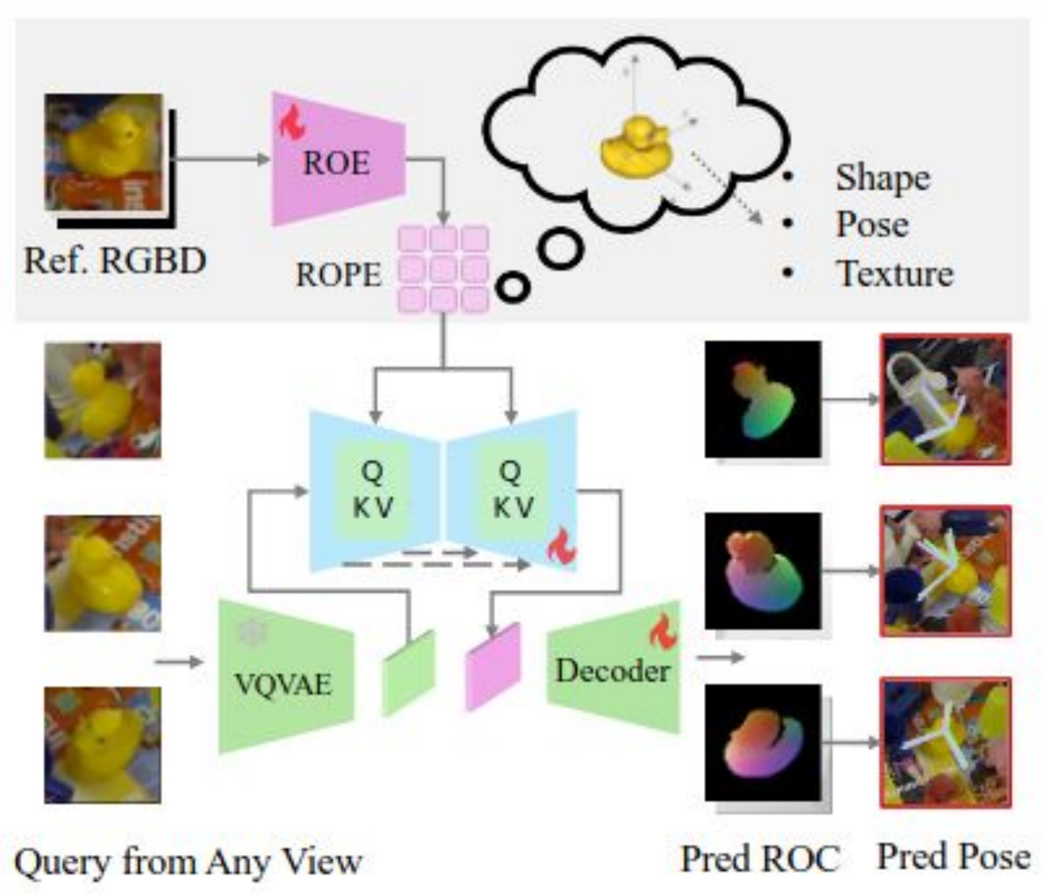
- High precision grasp pose detection in dense clutter
  https://arxiv.org/pdf/1603.01564
- GraspNet-1Billion (large-scale benchmark)
  https://openaccess.thecvf.com/content_CVPR_2020/papers/Fang_GraspNet-1Billion_A_Large-Scale_Benchmark_for_General_Object_Grasping_CVPR_2020_paper.pdf
- Contact-GraspNet (Cluttered scene)
  https://arxiv.org/pdf/2103.14127
  https://github.com/NVlabs/contact_graspnet
- GraspNeRF (Multiview, Transparent and Specular Objects)
  https://pku-epic.github.io/GraspNeRF/

6DoF pose estimation - Any6D (CVPR 2025)



https://sites.google.com/view/taeyeop-lee/any6d

6DoF pose estimation - One2Any (CVPR 2025)



https://github.com/lmy1001/One2Any

(Previous) DeepRob Final Project Showcase

(some previous examples
https://deeprob.org/w24/reports/
https://deeprob.org/w25/reports/)

Aha Slides (In-class participation)

https://ahaslides.com/Q2PWN

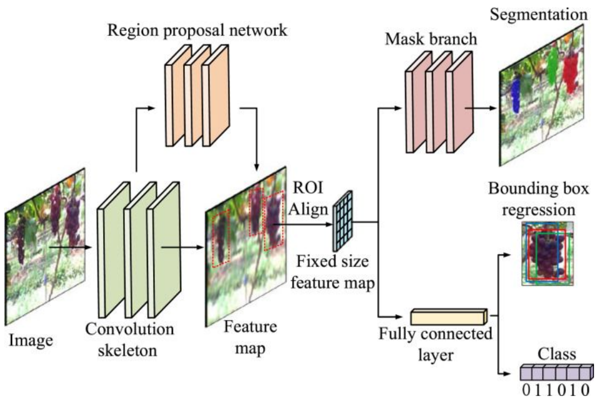
(Type in questions for our presenters - thanks!)

GrapeRob: A Grape Localization Pipeline for Automated Robotic Harvesting

Advaith Balaji
Isaac Madhavaram

Fruit Detection and Pose Estimation for Grape Cluster-Harvesting Robot Using Binocular Imagery Based on Deep Neural Networks

- Automation of Grape Harvesting
- Can be generalized for multiple fruits
- Lightweight, easily deployable model with high precision



Yin et al. 2021

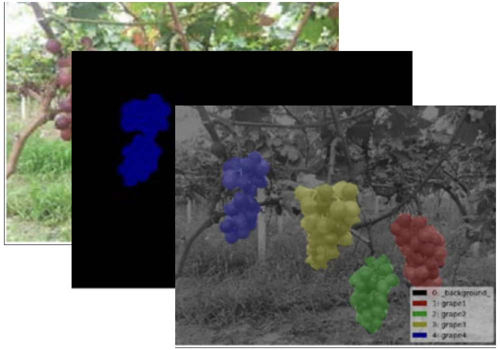
The Bigger Picture…

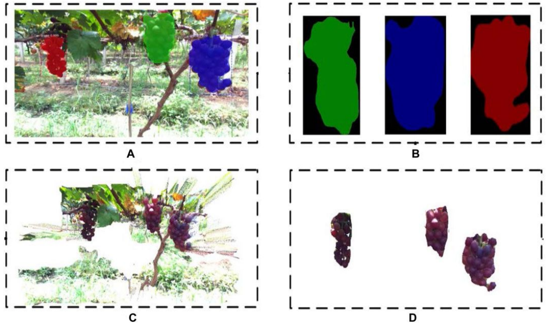
- Labor shortages and the inefficiencies of manual harvesting
- Alleviate labor shortages and ensure sustainable grape production

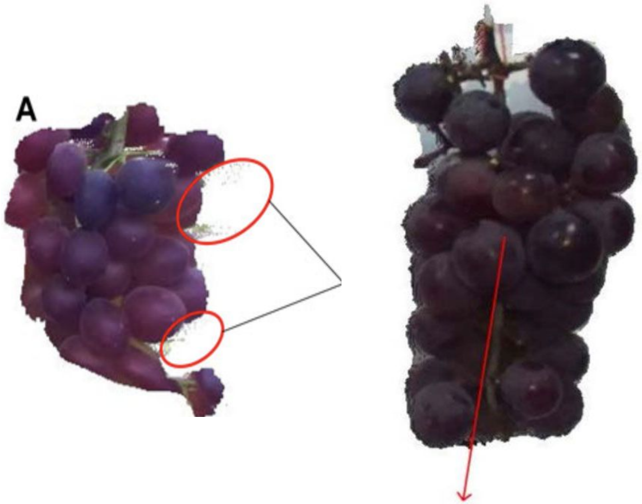
(72 million tons of grapes!!)

Approach

- Use binocular imagery to reconstruct a 3d point cloud of a segmented bunch of grapes

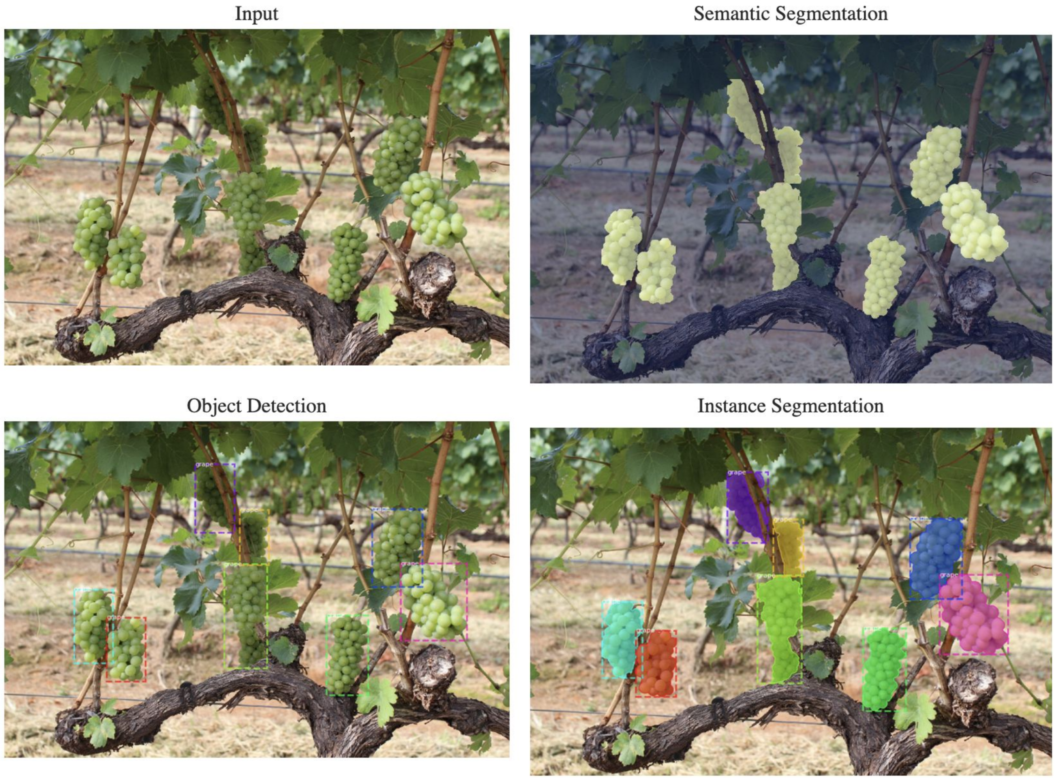






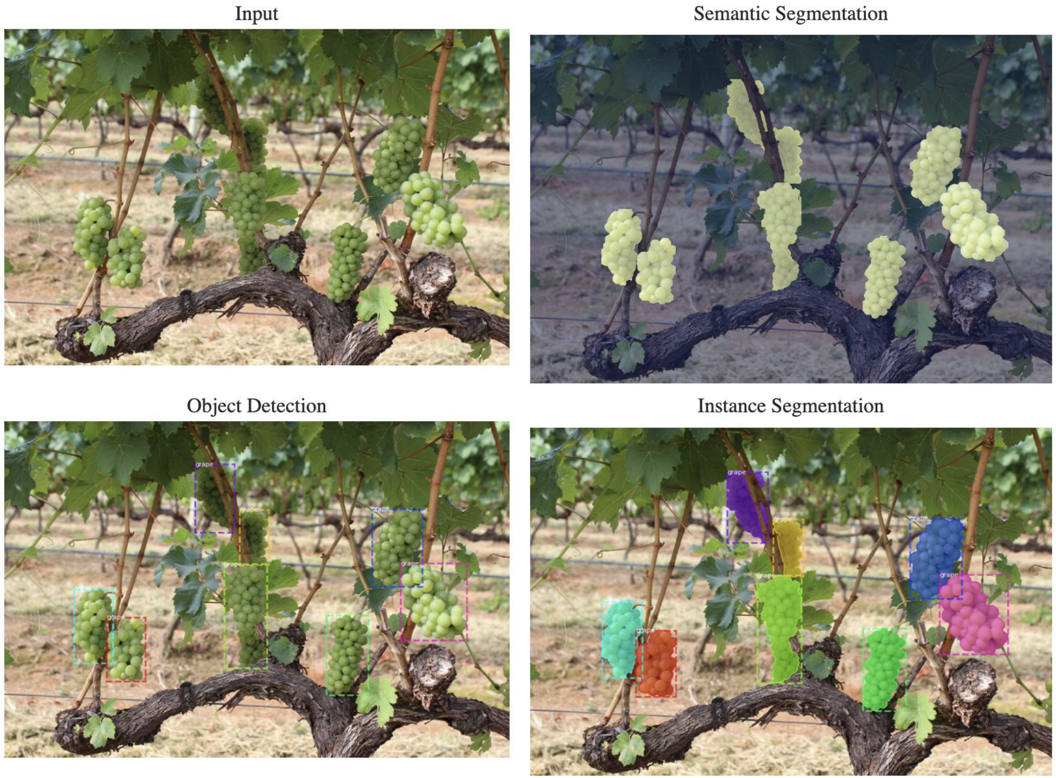
Yin et al. 2021

Data



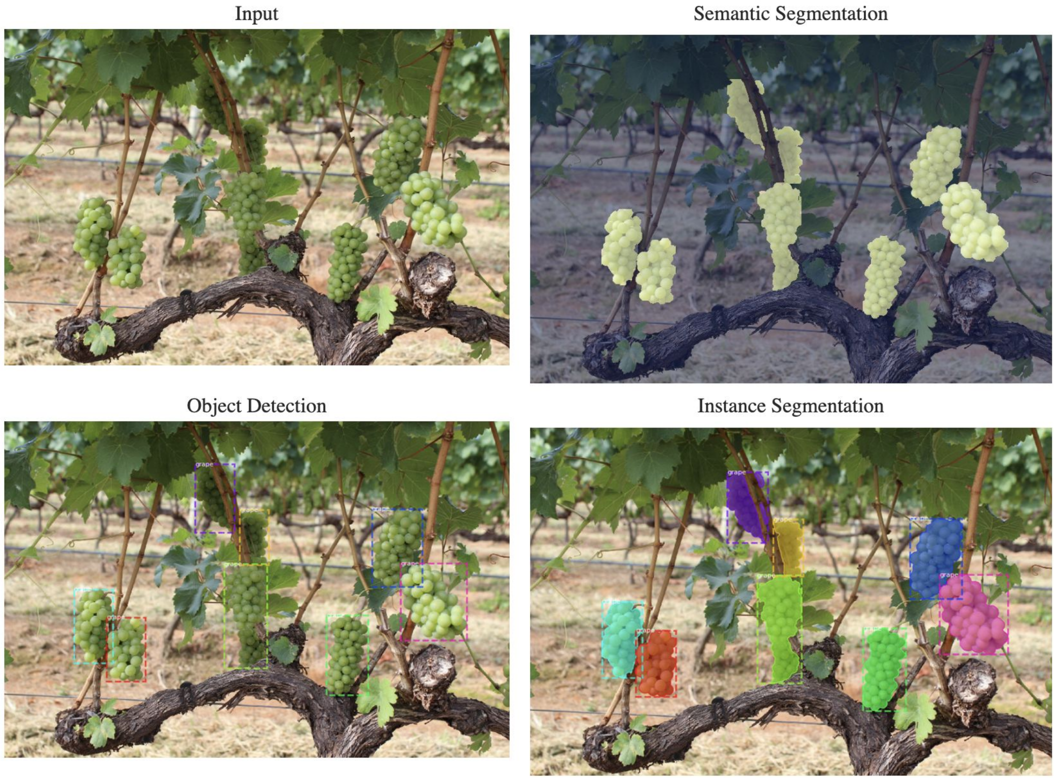
Santos et al.

What is the clear problem with this??



Santos et al.

What is the clear problem with this??



Must grasp at the stem!

Santos et al.

Results

Lighting conditions	Number of grapes	Precision/%	Recall/%	*IOU*/%
Frontlighting	72	92.31	97.30	83.23
Side-lighting	69	89.61	95.83	82.17
Back-lighting	65	86.67	92.86	80.61

Average detection time (s)	Average point cloud segmentation time (s)	Average total time (s)
1.1	0.6	1.7

Yin et al. 2021

What is the clear problem with this??

Lighting conditions	Number of grapes	Precision/%	Recall/%	*IOU*/%
Frontlighting	72	92.31	97.30	83.23
Side-lighting	69	89.61	95.83	82.17
Back-lighting	65	86.67	92.86	80.61

Average detection time (s)	Average point cloud segmentation time (s)	Average total time (s)
1.1	0.6	1.7

Yin et al. 2021

What is the clear problem with this??

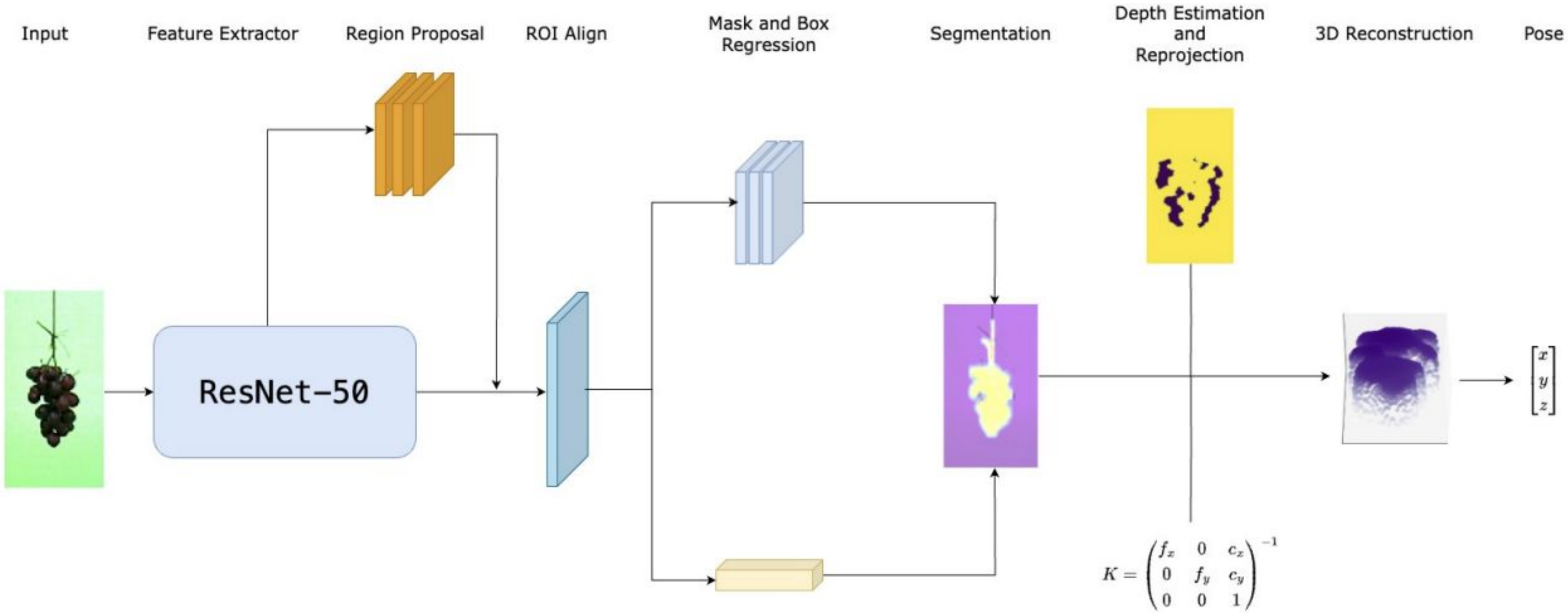
Lighting conditions	Number of grapes	Precision/%	Recall/%	*IOU*/%
Frontlighting	72	92.31	97.30	83.23
Side-lighting	69	89.61	95.83	82.17
Back-lighting	65	86.67	92.86	80.61

Average detection time (s)	Average point cloud segmentation time (s)	Average total time (s)
1.1	0.6	1.7

No robots!

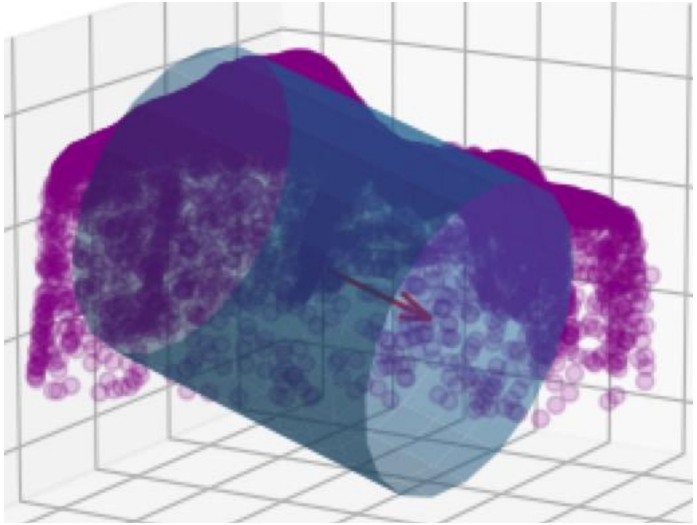
Yin et al. 2021

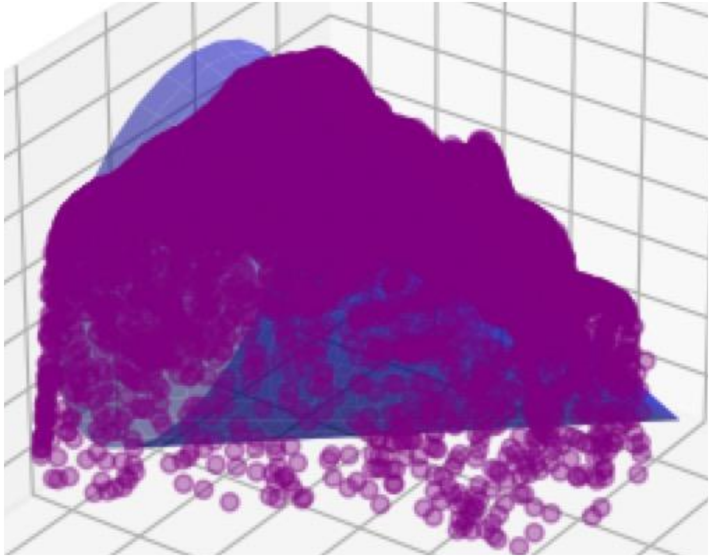
Our Pipeline



Our Contributions

Our Contributions

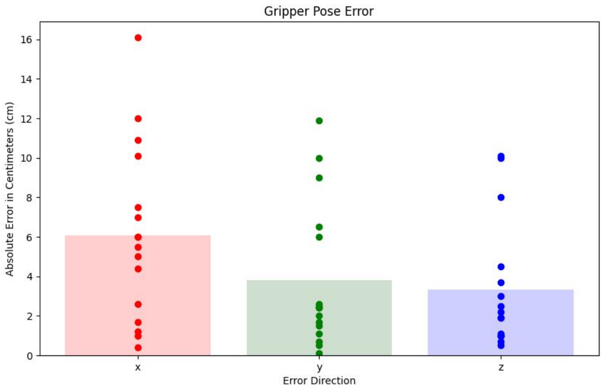




Our Contributions

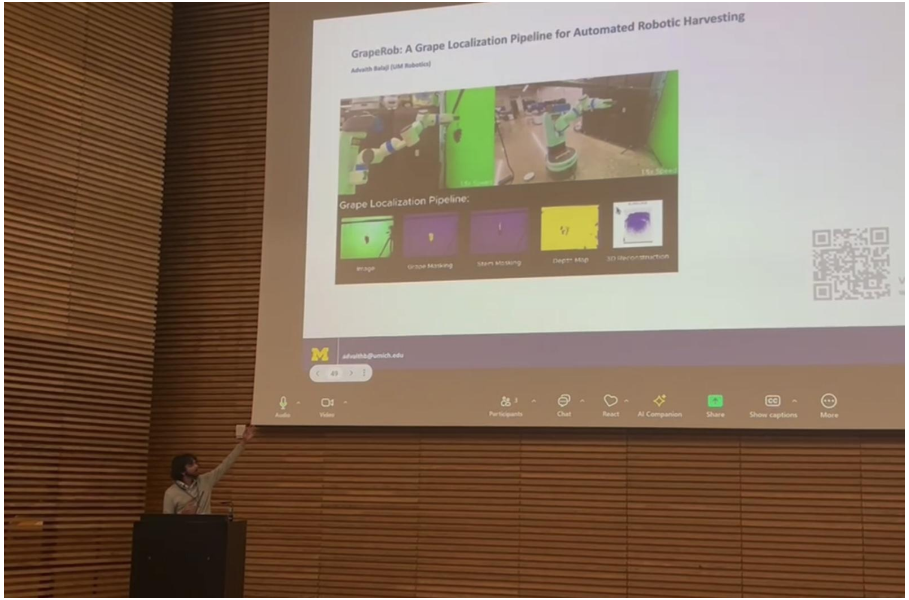
Robot Experiments

- 86% Grasp Success Rate
- Average Pose Error:
  - 6.08 cm in *x*
  - 3.80 cm in *y*
  - 3.11 cm in *z*



Robot Experiments

- 86% Grasp Success Rate
- Average Pose Error:
  - 6.08 cm in x
  - 3.80 cm in y
  - 3.11 cm in z



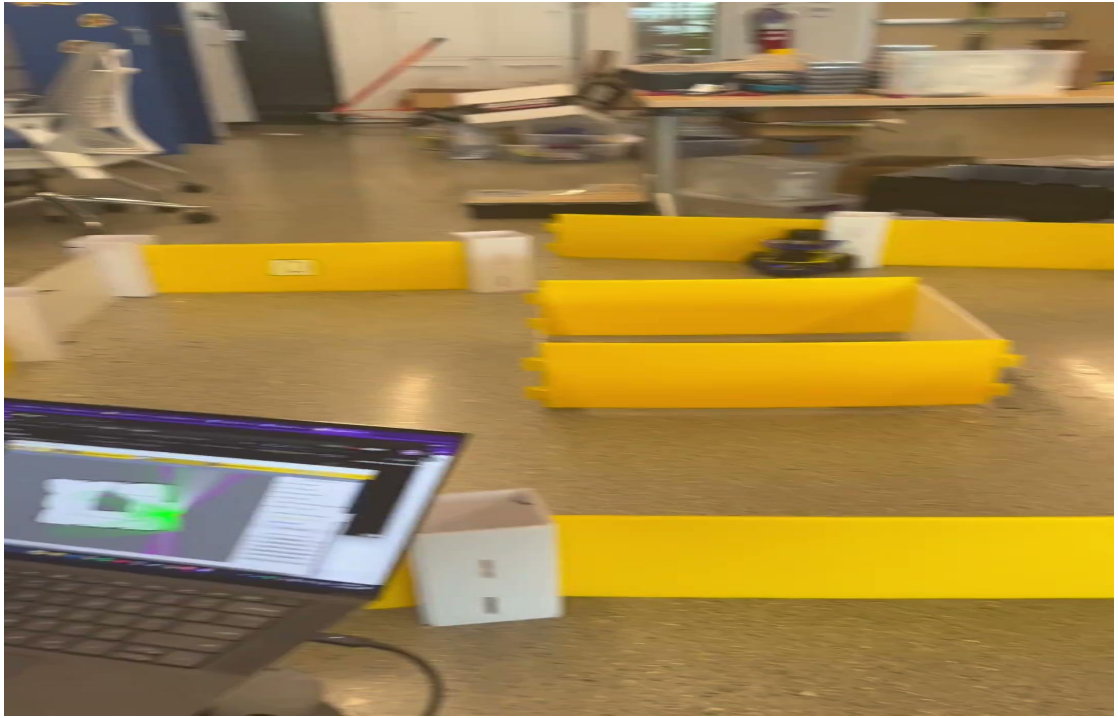
Presented at Michigan AI Symposium!

Winter 2025

Fast DPF for Lidar Localization

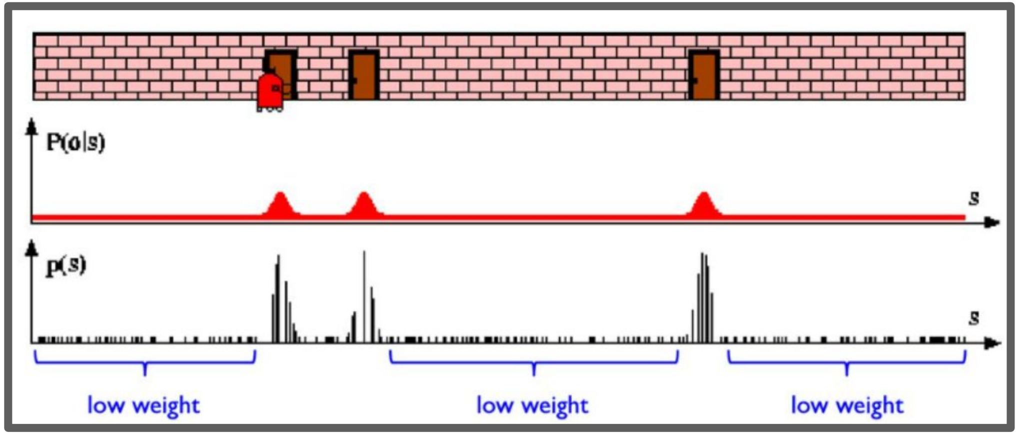
Raphael Allusson

Rob 330 - Lessons Learned

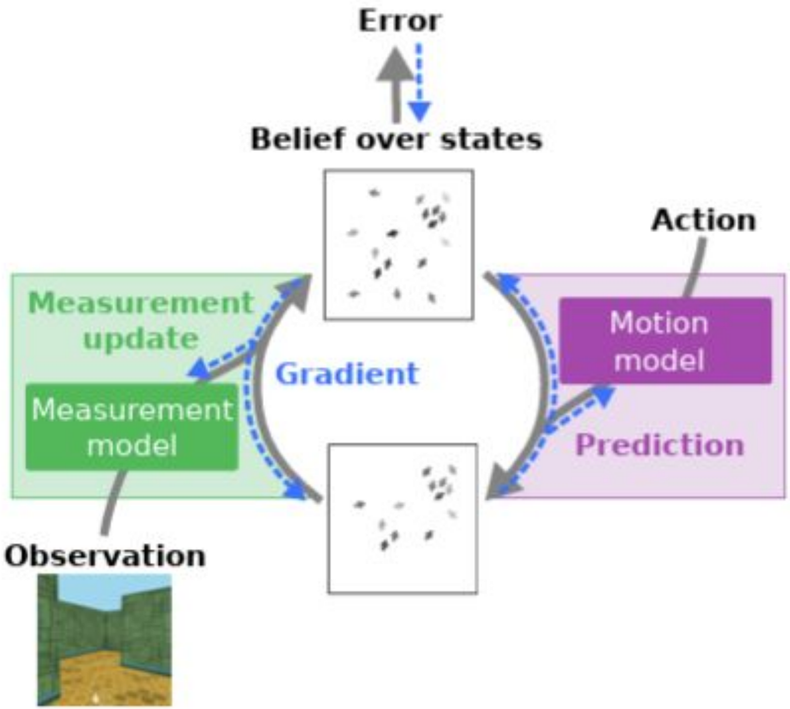


Background – Particle Filters

Particle Filters are a method of representing belief state as a sample of distribution, where the latent distribution has no intrinsic structure, allowing it to have a multi-modal representation of belief.



Differentiable Particle Filters

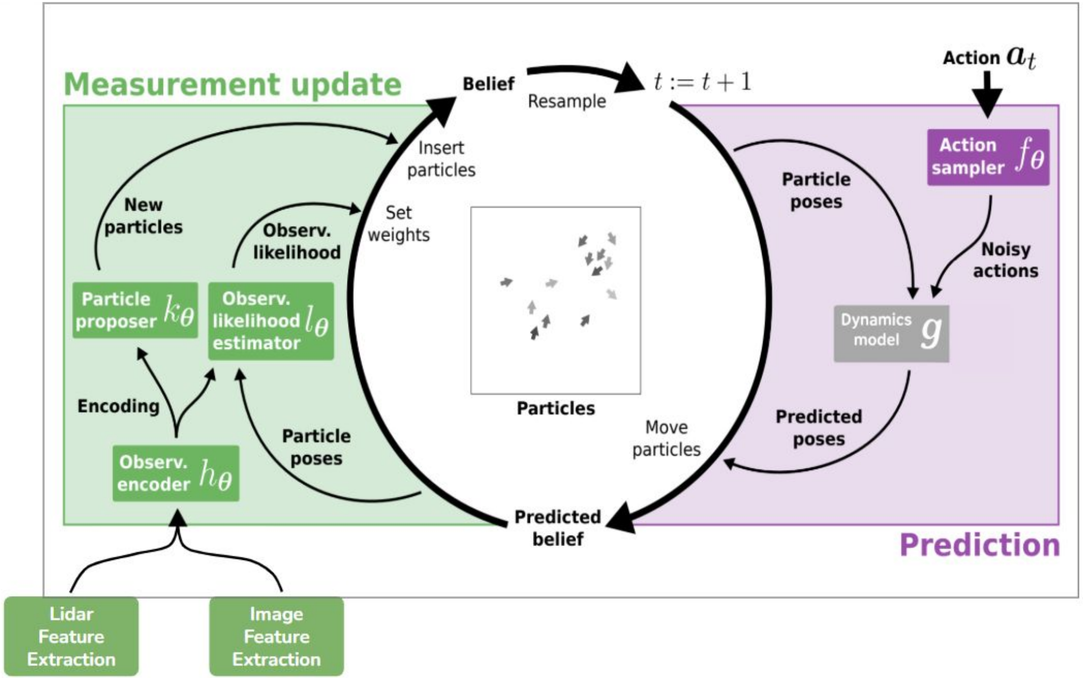


Value Proposition

How can we run DPF inference on Low Spec Hardware?

- Optimize Compute
- Increase Accuracy

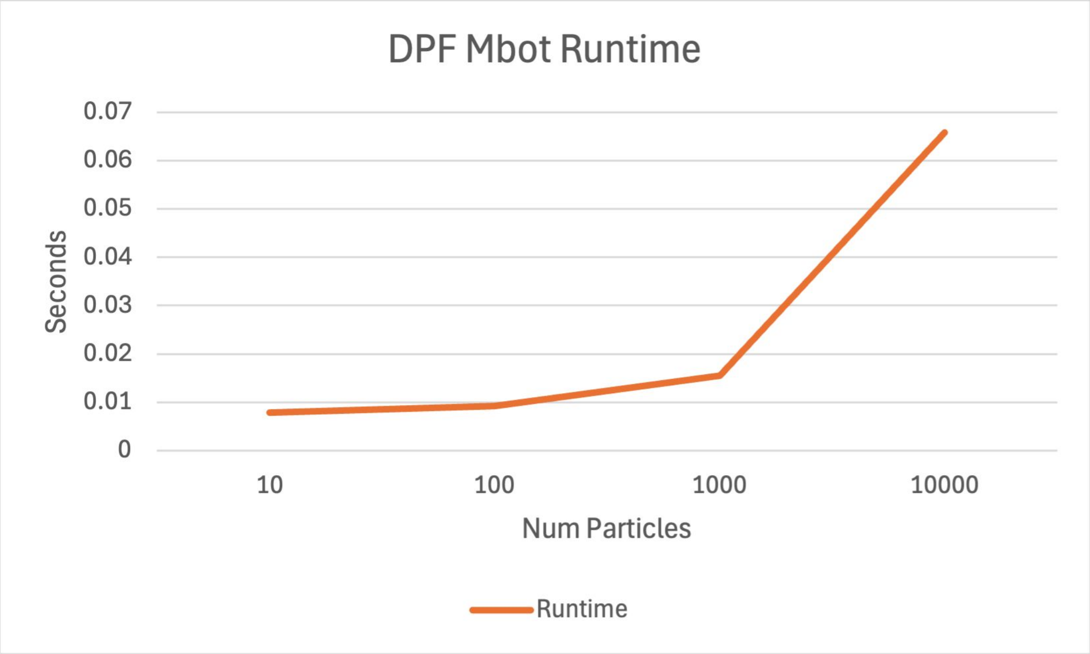
Extension - LiDAR Encoding



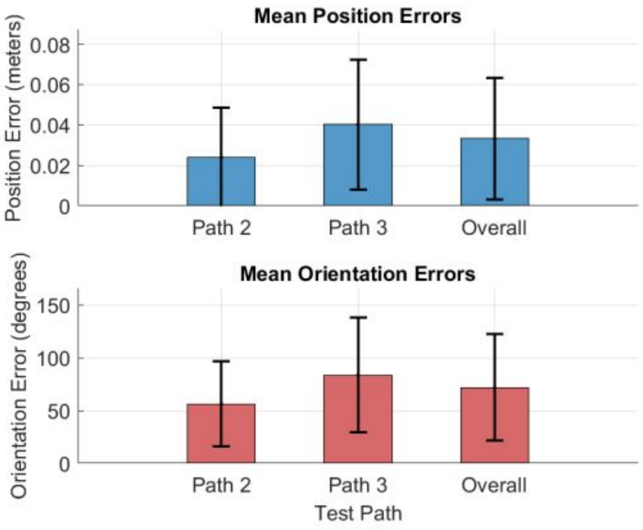
Extension - Optimization

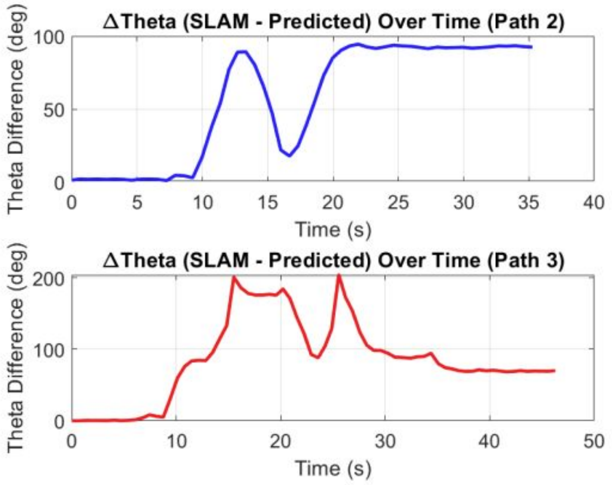
- TensorFlow 1 → Pytorch
- Vectorization
- Temporal Context

Results - Runtime



Results - Accuracy





Results - Accuracy

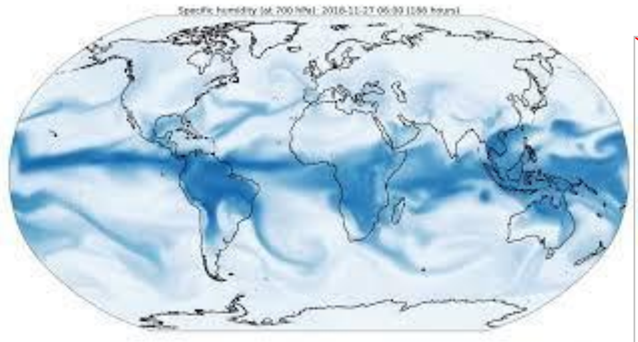
Aha Slides
(In-class participation)

https://ahaslides.com/Q2PWN

(questions for our presenters?)

Other projects (examples)

GraphCast



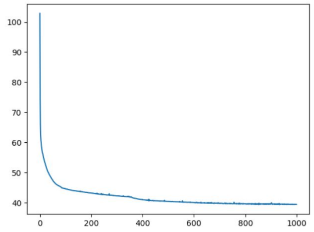


Fig. 5. Final convergence of the trained model over 1000 epochs, showing predicted solar radiation values. Y-axis is watts per square meter mean square training loss.

http://solarcast-ml.com/

SORNet: Spatial Object-Centric Representations for Sequential Manipulation

Jace Aldrich, Ariana Verges Alicea, Hannah Ho

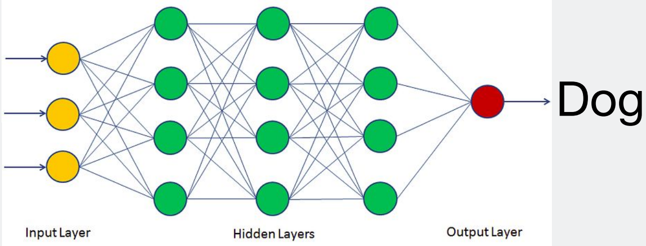
Original Paper Authors: Wentao Yuan, Chris Paxton, Karthik Desingh, Dieter Fox

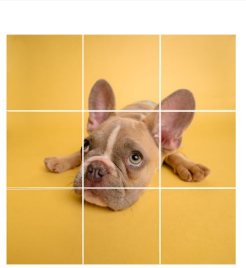
ROBOTICS

Why SORNet?

- How can we get information of objects and their spatial relationship with each other for object manipulation?
- With it, a robot can perform sequential tasks with objects (e.g. stacking blocks).

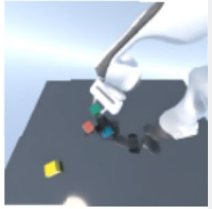
Vision Transformer Models





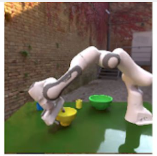
Overall Architecture

RGB Image



Canonical Object Views

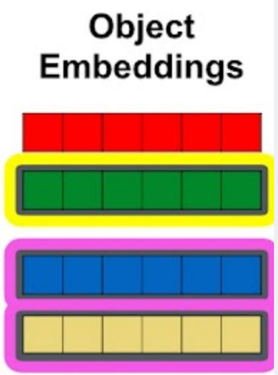
Object Embedding Network


RGB observation view 1


RGB observation view 2 *(optional)*


Depth observation view x *(optional)*

Readout Networks



Leonardo Dataset

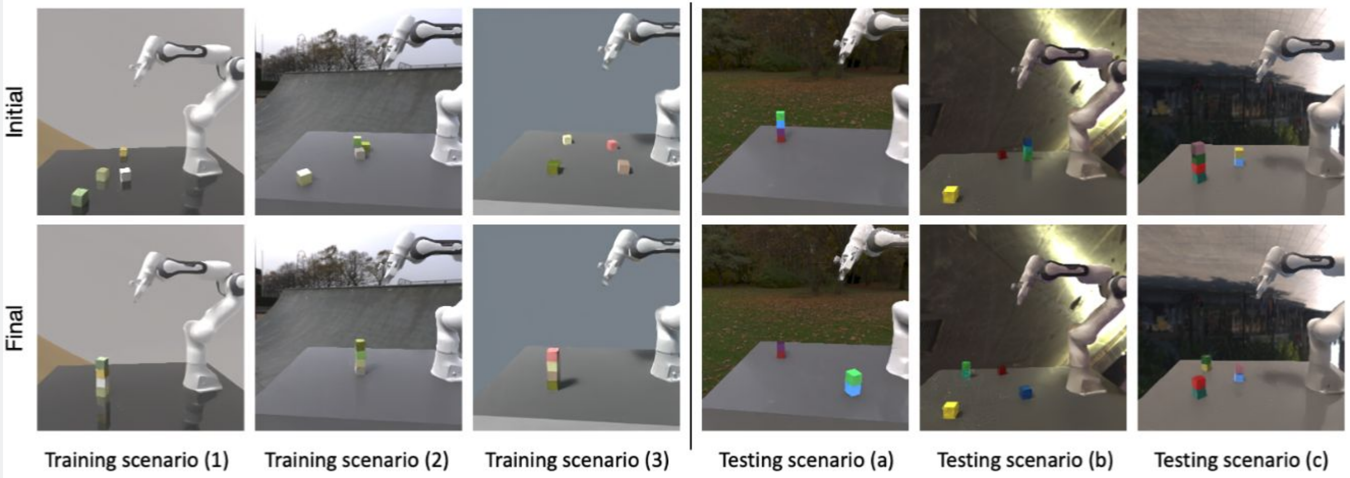
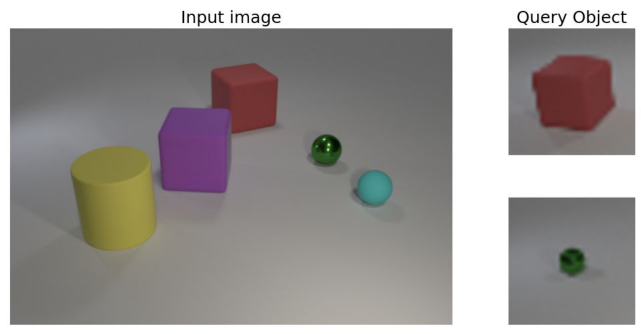


Figure 5: Sample scenes from training and testing scenarios in the Leonardo dataset. Top row shows the initial configuration of a sequence and the bottom row shows the goal configuration. The training scenarios contain 4 blocks with a single goal condition. The testing scenarios contain 4-7 blocks with heldout colors and various goal conditions involving multi-tower stacking.

Kitchen Dataset

Figure 6: Sample scenes from the kitchen dataset. The top and bottom rows show two different views. SORNet can leverage additional views to improve performance, but does not require multiple views.

Model Output



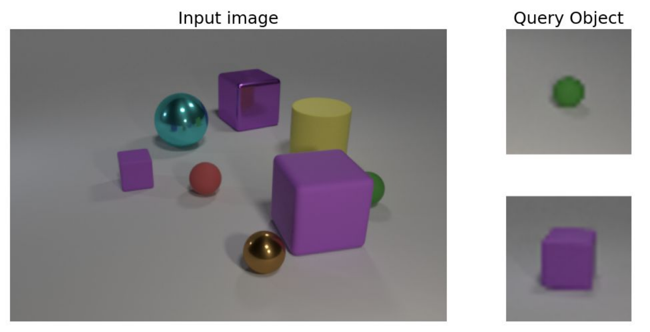
Question

Is the large red rubber cube to the left of the small green metal sphere?

Answer

SORNet: Yes

Ground truth: Yes



Question

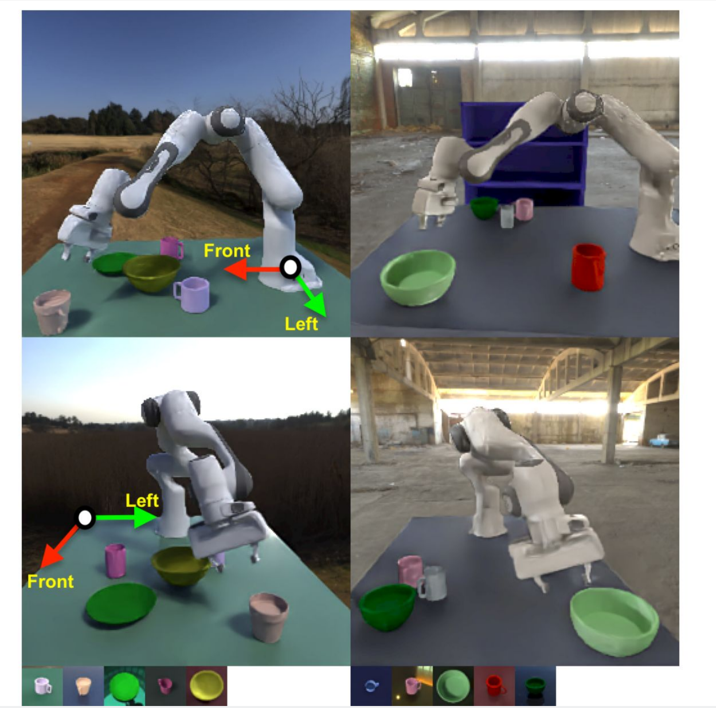
Is the small green rubber sphere to the right of the large purple rubber cube?

Answer

SORNet: Yes

Ground truth: Yes

Kitchen – Simulated



on_surface(00_pastel_purple_Mug, tabletop)
on_surface(01_pinkish_tan_Mug, tabletop)
on_surface(02_bottle_green_Bowl, tabletop)
on_surface(03_barbie_pink_Mug, tabletop)
on_surface(04_brownish_green_Bowl, tabletop)
left_of(01_pinkish_tan_Mug, 02_bottle_green_Bowl)
left_of(00_pastel_purple_Mug, 03_barbie_pink_Mug)
left_of(04_brownish_green_Bowl, 03_barbie_pink_Mug)
right_of(03_barbie_pink_Mug, 04_brownish_green_Bowl)
right_of(03_barbie_pink_Mug, 00_pastel_purple_Mug)
right_of(02_bottle_green_Bowl, 01_pinkish_tan_Mug)
in_front_of(02_bottle_green_Bowl, 03_barbie_pink_Mug)
in_front_of(01_pinkish_tan_Mug, 04_brownish_green_Bowl)
in_front_of(01_pinkish_tan_Mug, 00_pastel_purple_Mug)
behind(00_pastel_purple_Mug, 01_pinkish_tan_Mug)
behind(04_brownish_green_Bowl, 01_pinkish_tan_Mug)
behind(03_barbie_pink_Mug, 02_bottle_green_Bowl)
touching(02_bottle_green_Bowl, 04_brownish_green_Bowl)
touching(04_brownish_green_Bowl, 00_pastel_purple_Mug)
touching(00_pastel_purple_Mug, 04_brownish_green_Bowl)
touching(04_brownish_green_Bowl, 02_bottle_green_Bowl)

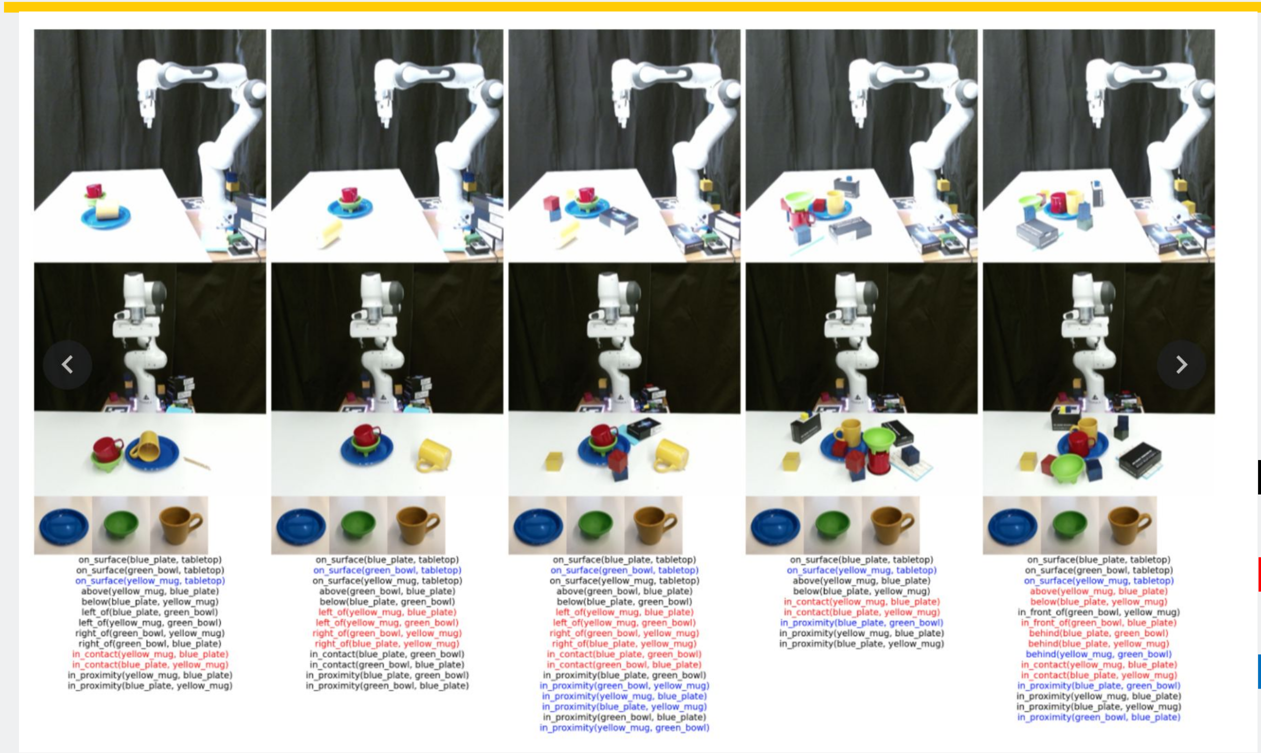
on_surface(00_baby_blue_Mug, tabletop)
on_surface(01_bubblegum_pink_Mug, tabletop)
on_surface(02_lightish_green_Bowl, tabletop)
on_surface(03_deep_red_Mug, tabletop)
on_surface(04_dark_forest_green_Bowl, tabletop)
left_of(02_lightish_green_Bowl, 04_dark_forest_green_Bowl)
left_of(02_lightish_green_Bowl, 00_baby_blue_Mug)
right_of(00_baby_blue_Mug, 02_lightish_green_Bowl)
right_of(04_dark_forest_green_Bowl, 02_lightish_green_Bowl)
in_front_of(02_lightish_green_Bowl, 03_deep_red_Mug)
in_front_of(04_dark_forest_green_Bowl, 01_bubblegum_pink_Mug)
behind(01_bubblegum_pink_Mug, 04_dark_forest_green_Bowl)
behind(03_deep_red_Mug, 02_lightish_green_Bowl)
touching(00_baby_blue_Mug, 01_bubblegum_pink_Mug)
touching(01_bubblegum_pink_Mug, 04_dark_forest_green_Bowl)
touching(04_dark_forest_green_Bowl, 00_baby_blue_Mug)
touching(00_baby_blue_Mug, 04_dark_forest_green_Bowl)
touching(04_dark_forest_green_Bowl, 01_bubblegum_pink_Mug)
touching(01_bubblegum_pink_Mug, 00_baby_blue_Mug)

Black – Correct

Red – False Negative

Blue – False Positive

Kitchen – Real World



Black – Correct

Red – False Negative

Blue – False Positive

PROPS Relation Dataset (Ours)

PROPS Relation Dataset (Ours)

Input image

Query Object

Question

Is the potted meat can

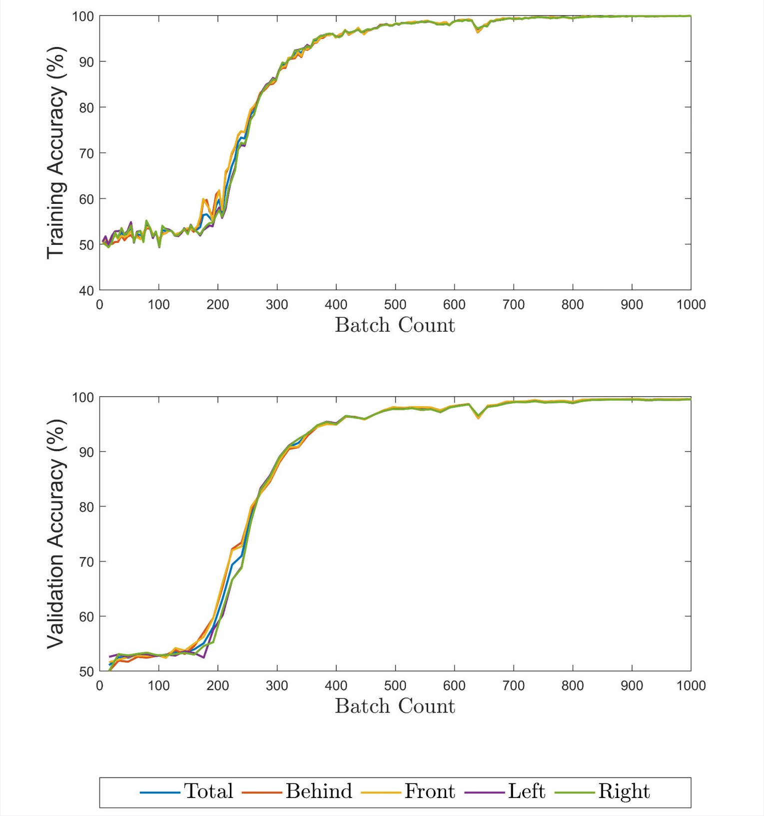
behind the master chef can?

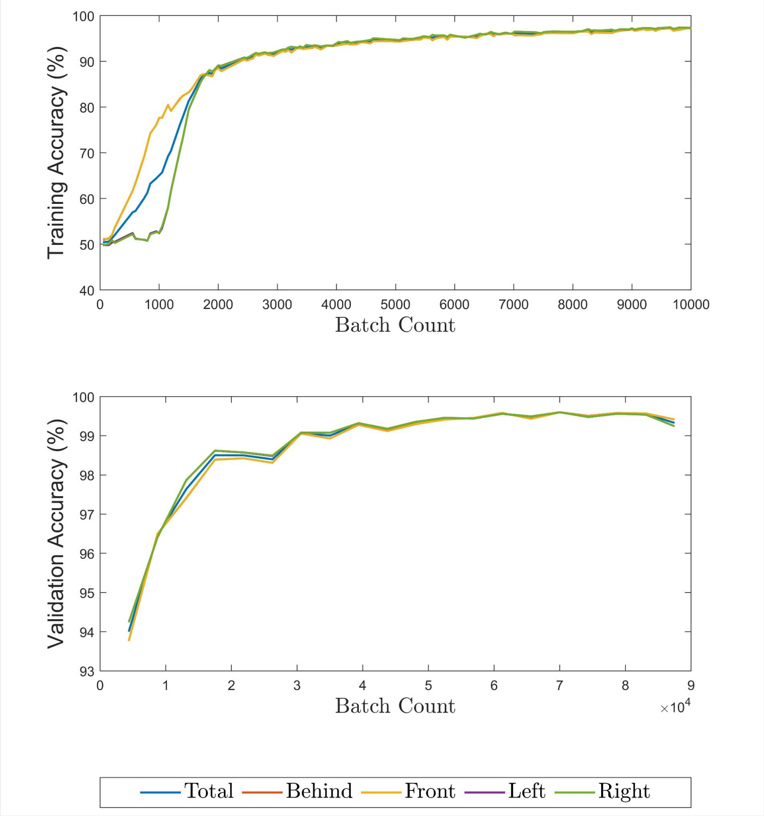
Answer

SORNet: Yes

Ground truth: Yes

CLEVR (left) vs PROPS (right)





PROPS Object Accuracy

	Master Chef Can	Cracker Box	Sugar Box	Tomato Soup Can	Mustard Bottle	Tuna Fish Can	Gelatin Box	Potted Meat Can	Mug	Large Marker	Average
Master Chef Can	-	99.30	99.77	98.80	98.90	98.77	98.65	99.20	99.15	98.85	99.04
Cracker Box	99.10	-	99.37	99.80	99.20	99.39	98.54	98.70	99.55	98.3000	99.11
Sugar Box	99.20	99.14	-	99.09	99.37	98.89	99.75	99.32	99.54	99.20	99.28
Tomato Soup Can	98.40	99.65	99.26	-	99.40	98.87	98.86	99.60	99.00	99.15	99.13
Mustard Bottle	99.30	98.90	99.26	99.90	-	98.87	99.68	98.95	98.55	98.95	99.15
Tuna Fish Can	98.98	99.28	99.41	98.98	97.95	-	99.11	99.13	98.98	99.28	99.01
Gelatin Box	99.19	99.40	99.88	99.51	99.89	99.33	-	98.81	99.78	99.03	99.43
Potted Meat Can	99.20	98.70	99.03	99.75	98.30	99.38	98.81	-	98.90	98.20	98.92
Mug	98.80	99.45	99.49	98.80	98.70	98.92	99.51	99.65	-	99.45	99.20
Large Marker	98.30	98.10	99.43	99.20	98.95	99.23	99.24	99.20	99.55	-	99.03
Average	98.94	99.10	99.43	99.31	98.96	99.08	99.13	99.17	99.22	98.93	99.13
Complete Average	98.99	99.10	99.36	99.22	99.06	99.04	99.28	99.05	99.21	98.98	

TABLE I: Full Size PROPS Data Validation Accuracy Percentages for all Relationships. The row is object 1 in the relationship, the column is object 2 in the relationship. The complete average is the average over the object's row and column, as SORNet treats the first and second patches differently.

Site Page Link

Automatic Data Generation for SORNet: PROPS Relation Dataset | DeepRob: Deep Learning for Robot Perception

Self-Supervised Learning for 6D Object Pose Estimation

Sydney Belt, Conghao Jin, Gurnoor Kaur, Joshua Symonds

Original Paper (Wild6D) Authors:

K. Zhang, Y. Fu, S. Borse, H. Cai, F. Porikli, X. Wang

M | ROBOTICS

Background

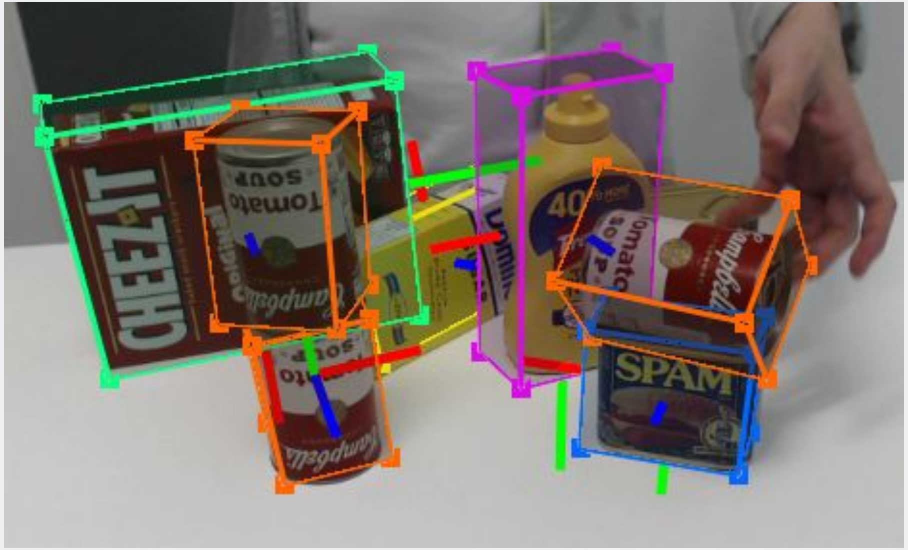
6D Object Pose Estimation

2D Images / Video Frames

3D Position

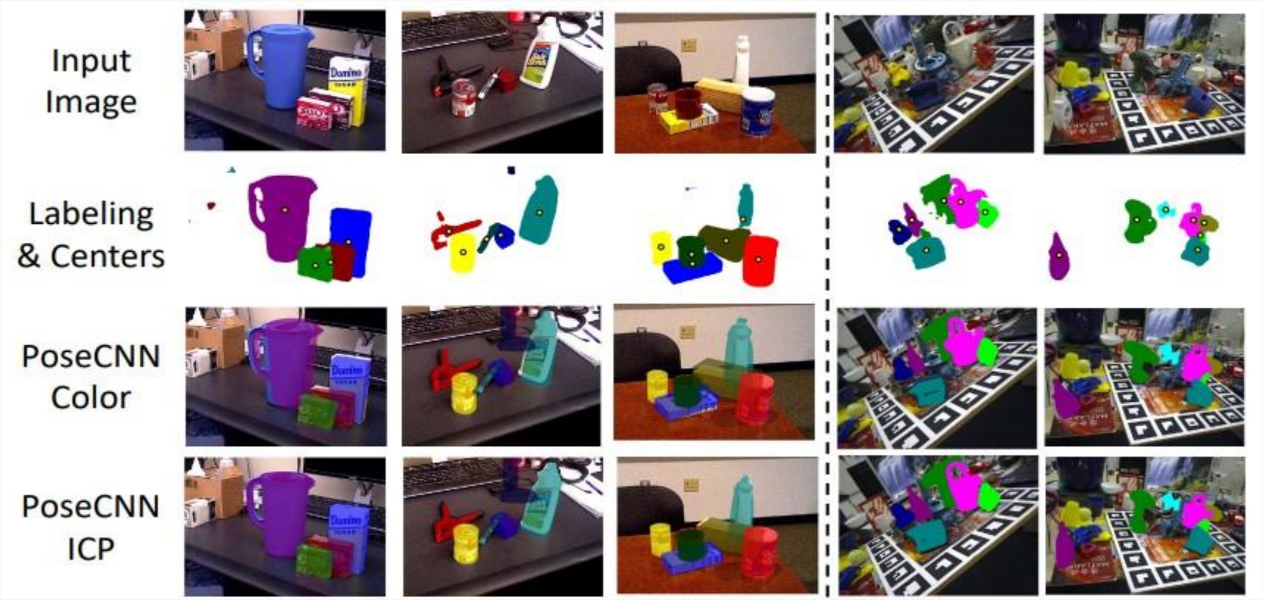
3D Orientation

6D Pose



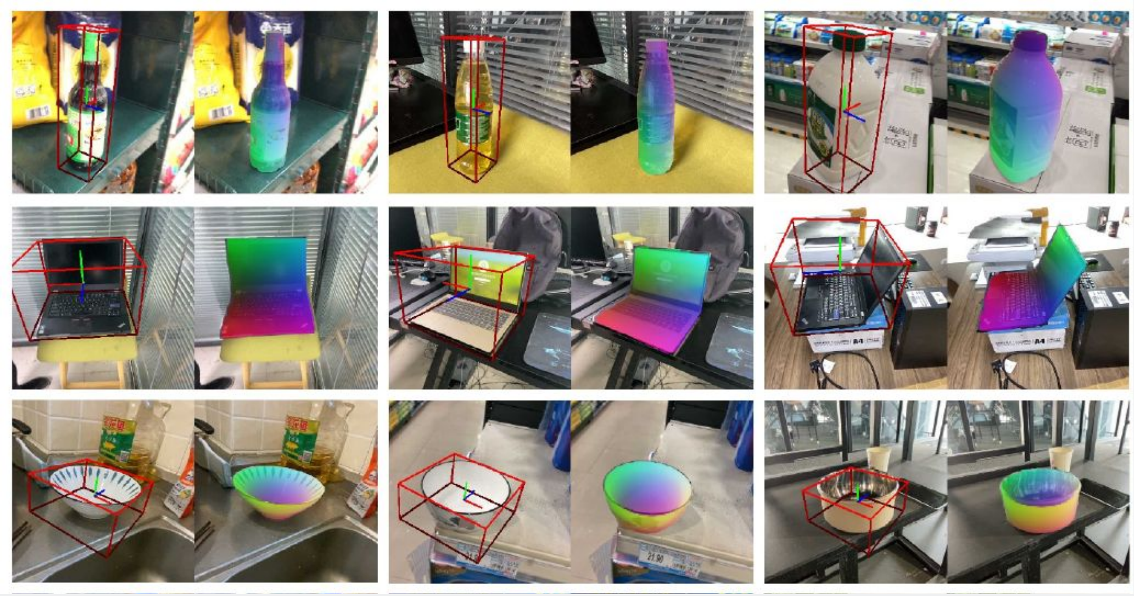
Background

1. Lack of Generalization



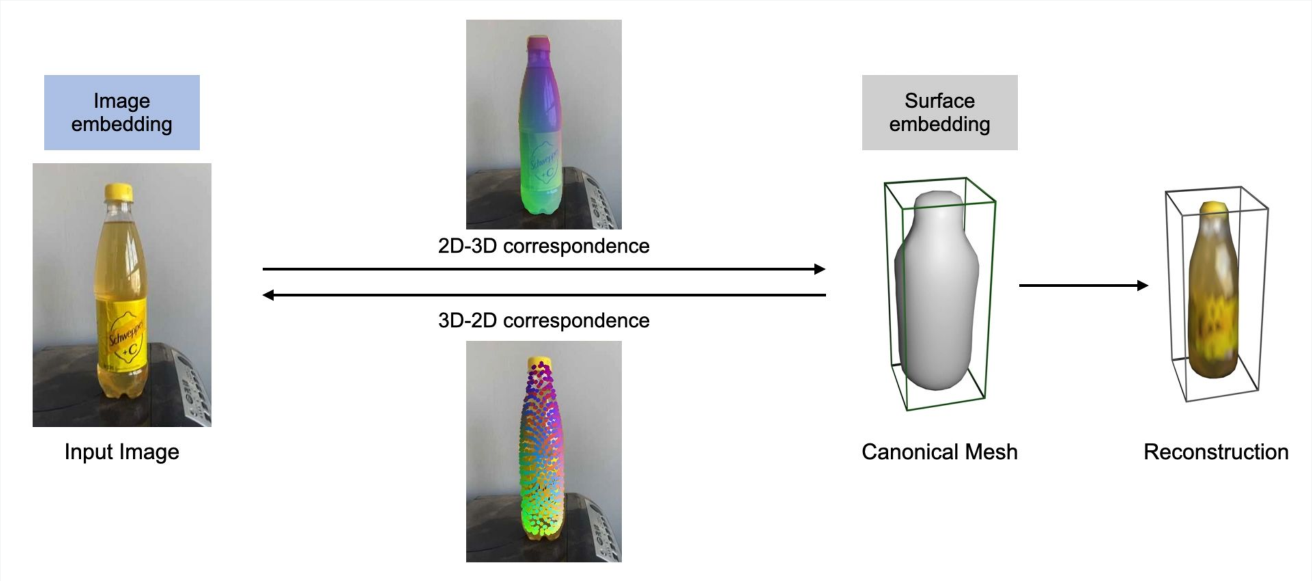
Background

2. Self-Supervised 6D Pose Estimation in the Wild



Background

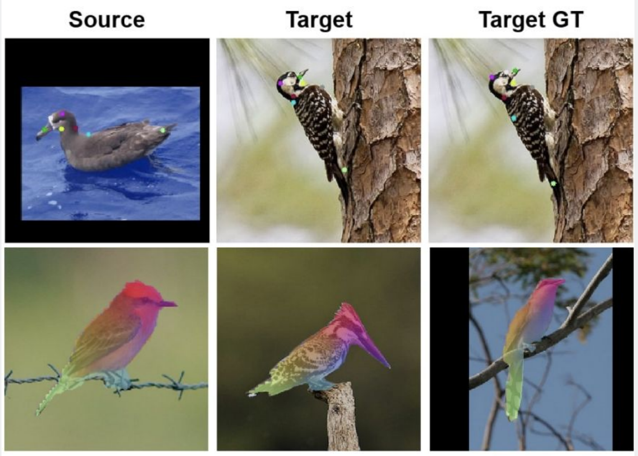
3. Categorical Surface Embedding



Related Works

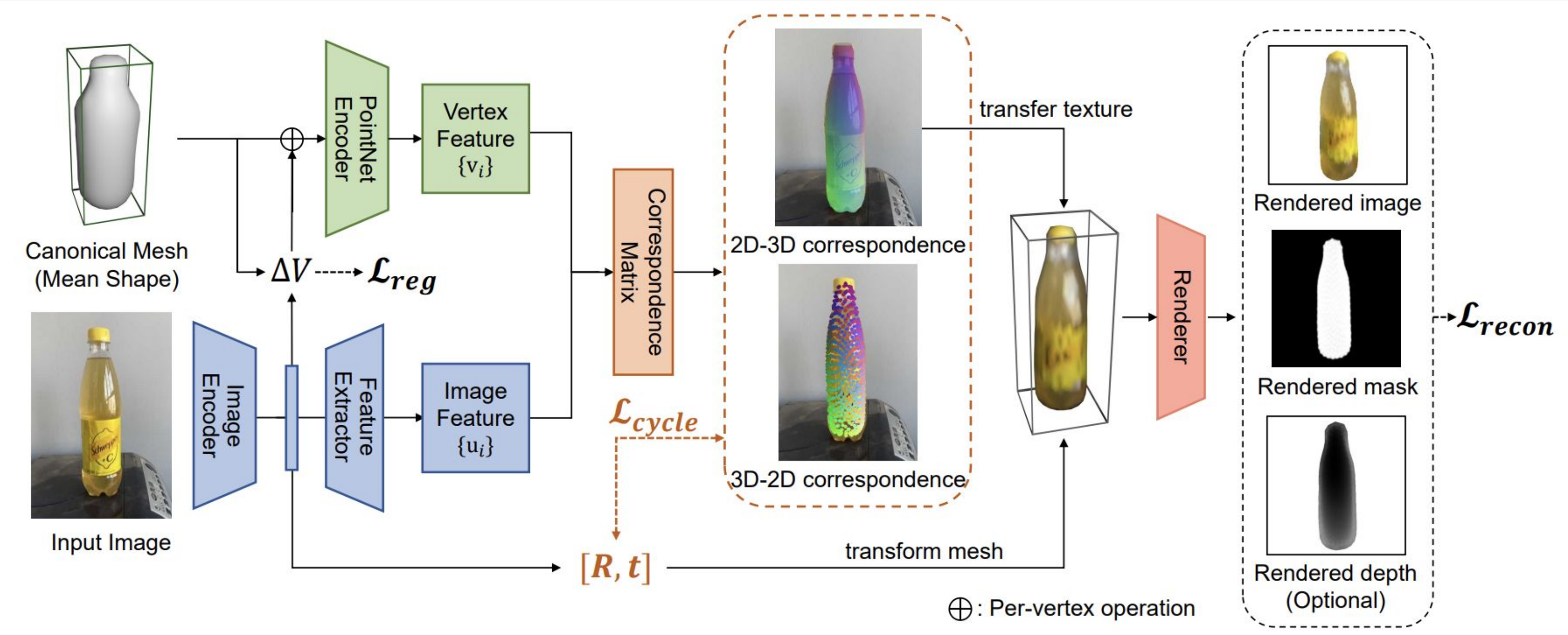
Semantic Encoding of Pixels from a DINO Trained Model

KeyPoint Transfer Task and Performance



Reproduction

Model Architecture



Pretrained Models

Model Accuracy					
Metric	Laptop	Camera	Bottle	Bowl	Mug
3D IOU: 25%	0.997	0.867	0.931	0.982	0.889
3D IOU: 50%	0.956	0.145	0.835	0.890	0.652
5°, 2cm	0.151	0.000	0.739	0.778	0.011
5°, 5cm	0.176	0.000	0.813	0.829	0.011
10°, 2cm	0.242	0.000	0.808	0.862	0.045
10°, 5cm	0.457	0.000	0.911	0.945	0.056

Retraining

Model Accuracy		
Metric	Pretrained Model	Our Training
3D IOU: 25%	0.997	0.999
3D IOU: 50%	0.956	0.864
5°, 2cm	0.151	0.024
5°, 5cm	0.176	0.028
10°, 2cm	0.242	0.083
10°, 5cm	0.457	0.167

Extension

Motivation

1. Limited hyperparameter tuning

Hyperparameters	Wild6D	REAL275	CUB
# of iterations	20,000	10,000	5,000
$(\beta_{\text{texture}}, \beta_{\text{mask}}, \beta_{\text{depth}})$ (Eq. 3)	$(0.05, 0.15, 0.1)$	$(0.05, 0.15, 0.1)$	$(0.05, 0.15, 0)$
$(\beta_{\text{2D-3D}}, \beta_{\text{2D-3D}}, \beta_{\text{inst}}, \beta_{\text{inst}})$ (Sec. 3.3)	$(0.02, 0.02, 0.05, 0.05)$	$(0.02, 0.02, 0.05, 0.05)$	$(0.01, 0.01, 0.1, 0.1)$
τ (Eq. 1)	0.1	0.1	0.1
k (Eq. 6)	200	200	200
$(\lambda_{\text{recon}}, \lambda_{\text{cycle}}, \lambda_{\text{reg}})$ (Sec. 3.3)	$(1, 1, 1)$	$(1, 1, 1)$	$(1, 1, 1)$

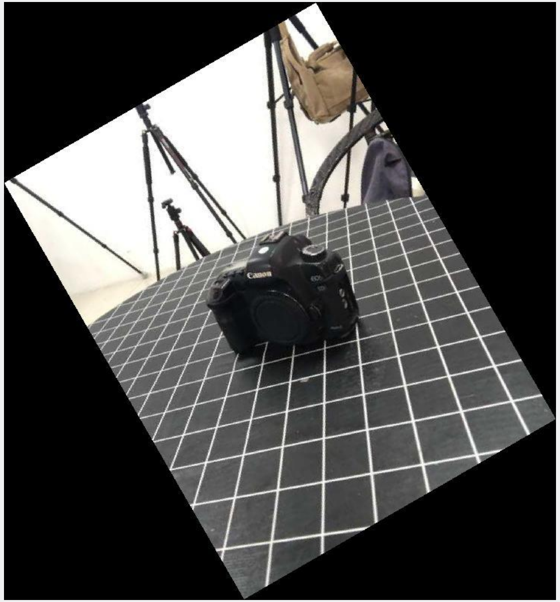
Hyperparameter Tuning

Model Accuracy			
Metric	weight-0.02	weight-0.05	weight-0.08
3D IOU: 25%	0.854	0.844	0.855
3D IOU: 50%	0.109	0.118	0.098

Motivation

2. Data Augmentation

rotate

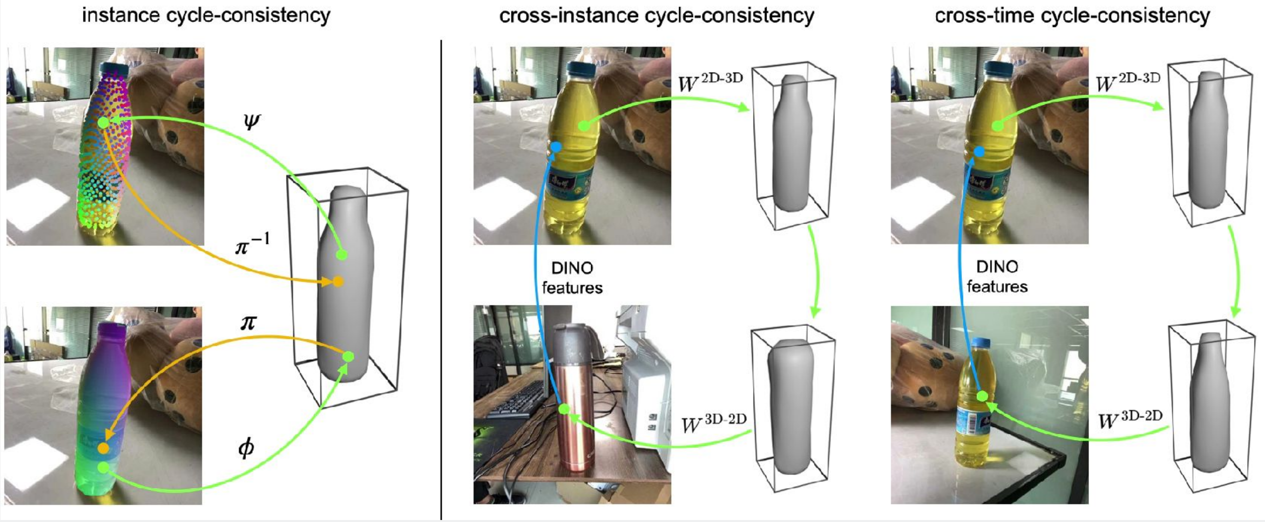


Data Augmentation

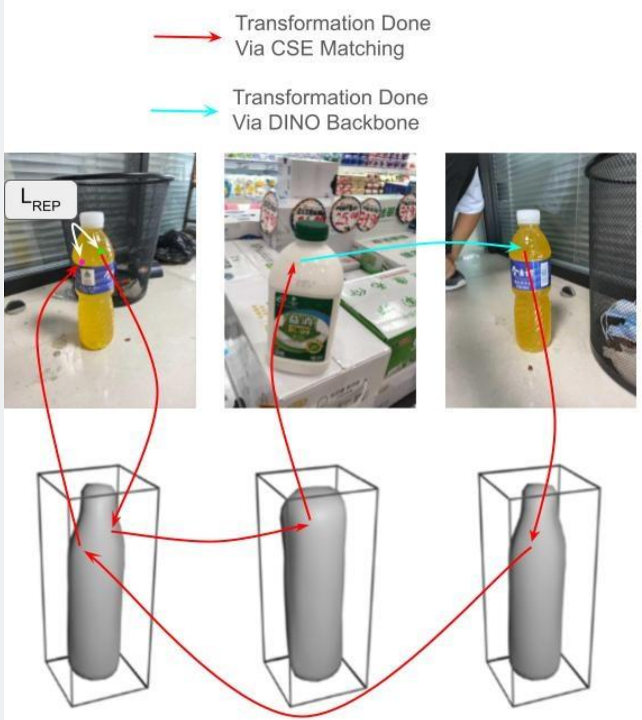
Model Accuracy				
Cycle-Loss-Weight	Metric	W/O augmentation	W/ augmentation	Improvement
0.02	3D IOU: 25% 3D IOU: 50%	0.854 0.109	0.864 0.097	0.170% -11.009%
0.05	3D IOU: 25% 3D IOU: 50%	0.844 0.118	0.851 0.101	0.829% -14.406%
0.08	3D IOU: 25% 3D IOU: 50%	0.855 0.098	0.857 0.091	0.233% -7.142%

Motivation

3. Cycle-consistency losses limitation
   (Don’t account for intermediate mapping inaccuracies)



RepLoss Function



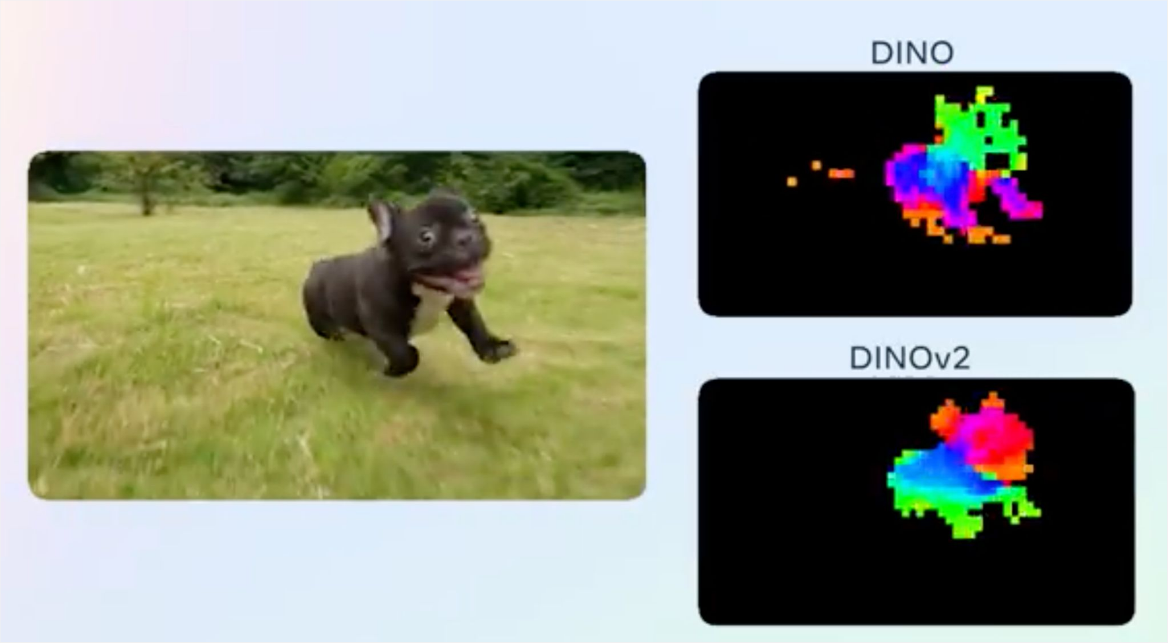
Model Accuracy			
Metric	W/O RepLoss	W/ RepLoss	Improvement
3D IOU: 25%	0.999	0.998	-0.090%
3D IOU: 50%	0.864	0.896	+3.702%
5°, 2cm	0.024	0.024	-0.752%
5°, 5cm	0.028	0.043	+55.844%
10°, 2cm	0.083	0.126	+52.042%
10°, 5cm	0.167	0.170	+1.927%

Motivation

4. Outperforms its own backbone

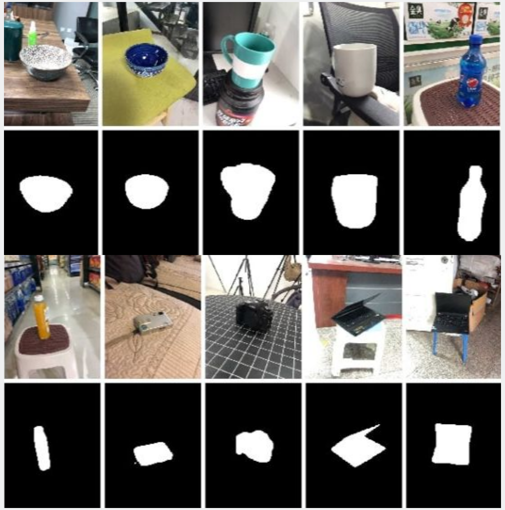
Method	3D/2D Transfer	PCK
VGG (Simonyan & Zisserman, 2014)	2D	17.2
DINO (Caron et al., 2021)	2D	60.2
Ours-2D	2D	**72.9**
Ours	3D	64.5

Replacing Backbone



Motivation

5. Ground truth requirements



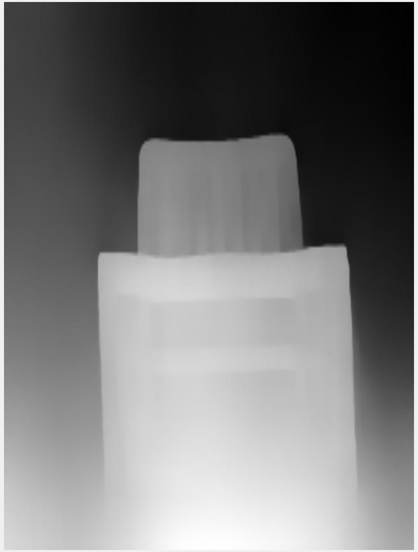
Segmentation Mask

Depth Map

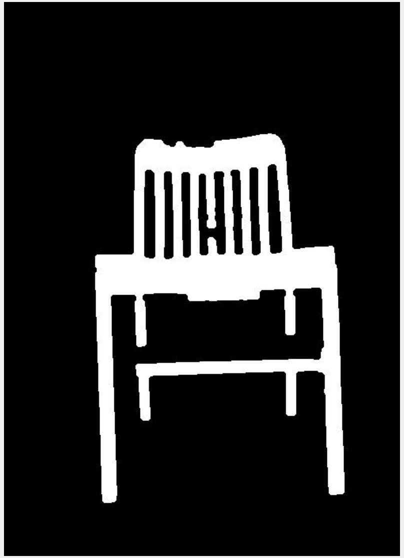
New Datasets

RGB-Image

Depth Map



Object Mask



W26 Final project webpage

(will continue to update)

https://deeprob.org/w26/projects/finalproject/

Timeline:

- Proposal ("lightning talk"):
  - March 31, 2026 during discussion section
- Final showcase:
  - FRB Atrium, April 27, 2026 1:30PM-3:30PM EST

Final Project team sign-up

https://docs.google.com/spreadsheets/d/1Jjv6vBQwAtCpUSQDAx0Gj5pDCWWqcvSQEeBFIF5prqo/edit?usp=drive_link

Great Lakes Computing Resources

Useful Links for Compute Resources

https://coerc.engin.umich.edu/intro-to-hpc/

- What is a cluster?
- Logging into a Cluster
- OpenOnDemand GUI access (interactive app - e.g., Jupyter Notebook)

All students enrolled should have received an email from Great Lakes.

$60.91 $\approx$ 371 GPU hours

Request placed in Queue. Depending on scheduler
Max time to request: two weeks (but may have to wait)

Useful Links for Compute Resources

Saving data:

- 60 days data purged (ARC will send reminder email. Be aware!)

- Saving data:

Your own folder

(both 430 and 599.430) #SBATCH --account=rob430w26s001_class

```
ls /scratch/rob430w26s001_class_root/rob430w26s001_class/YOUR_UNIQNAME
ls /scratch/rob430w26s001_class_root/rob430w26s001_class/shared_data
```

`my_accounts YOUR_UNIQNAME` (see your own account)

Shared data

```
#SBATCH --account=rob430w26s001_class
```

Submit batch job

Useful Links for Compute Resources

Submitting jobs:

```
sbatch python.sbat
```

Submit batch job, define time, job, etc.

```
module keyword torch
module avail
module list
pip install numpy -user
module load python/3.10.4
```

#(currently loaded modules)

#(example)

ROBOTICS

Useful Links for Compute Resources

Note about OpenOnDemand

there is a 'viz' partition in Great Lakes, that is only accessible form Open OnDemand: https://documentation.its.umich.edu/arc-hpc/open-ondemand
It is set aside for interactive jobs and has a 2 hour wall clock limit. The nodes have NVIDIA Tesla P40 configured for accelerating OpenGL graphics using VirtualGL. This means that OpenGL application can get accelerated graphic with in the web interface.

Useful Links for Compute Resources

Transferring Data/Files (GLOBUS):

https://coerc.engin.umich.edu/globus/

Contact:

coe-research-computing@umich.edu
arc-support@umich.edu

ROBOTICS

Useful Links for Compute Resources

Additional resources on Great Lakes tutorial

Great Lakes “Cheat sheet”
https://docs.google.com/document/d/1wsr3yzkkojUMBCCneCz-l413xBzU-SZFAqcFrAAjttk/edit?tab=t.0#heading=h.kquo6lavnl0f
ARC training events
https://ttc.iss.lsa.umich.edu/ttc/sessions/tag/arc/
Great Lakes User Guide
https://documentation.its.umich.edu/arc-hpc/greatlakes/user-guide
Great Lakes User Login
https://its.umich.edu/advanced-research-computing/high-performance-computing/login

Midterm Prep

(some previous examples)
https://drive.google.com/file/d/1ajuXX8whNlH4V190Yaig1GnysEeDRqpW/view?usp=drive_link
Will discuss in Discussion section on Tuesday March 10, 2026

DeepRob Midterm

Wednesday March 11, 2026 12PM-1:25PM @1303 EECS (Lecture time and location)

10% overall grade. Individual.

- Please bring your own pen/pencil!
- No phone/computer/internet/genAI.
- You may bring 1 A4/Letter-size note sheet (front and back).
- All course content (lecture, discussion, projects, canvas quizzes, etc.) up until March 11, 2026.
- If you need accommodations please contact instructors.