The Curious Robot: Learning Visual Representations via Physical Interactions

Lerrel Pinto, Dhiraj Gandhi, Yuanfeng Han,
Yong-Lae Park, Abhinav Gupta

ECCV 2016

Presenter: Ginevra Gaudioso

R B

Problem

- Learning **visual representations** of objects
- By **actively interacting** with the objects

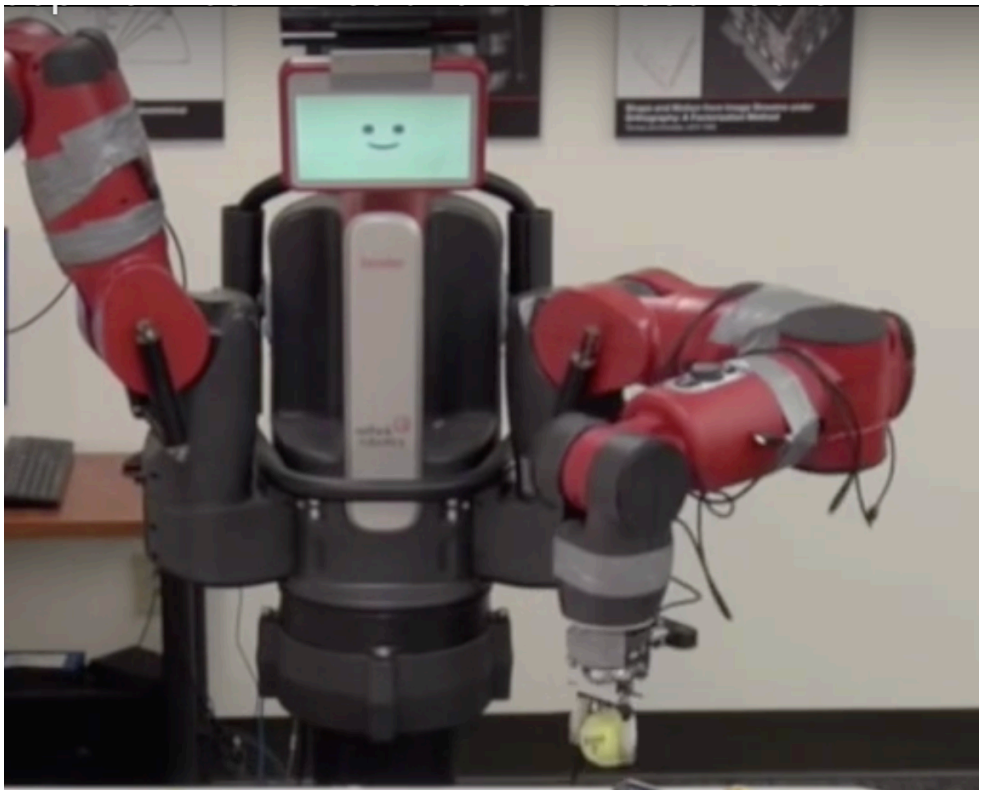


Image source: https://youtu.be/oSqHc0nLkm8?t=49

Why do we care?

- Robotics: robot needs to recognize objects
- Vision: classification

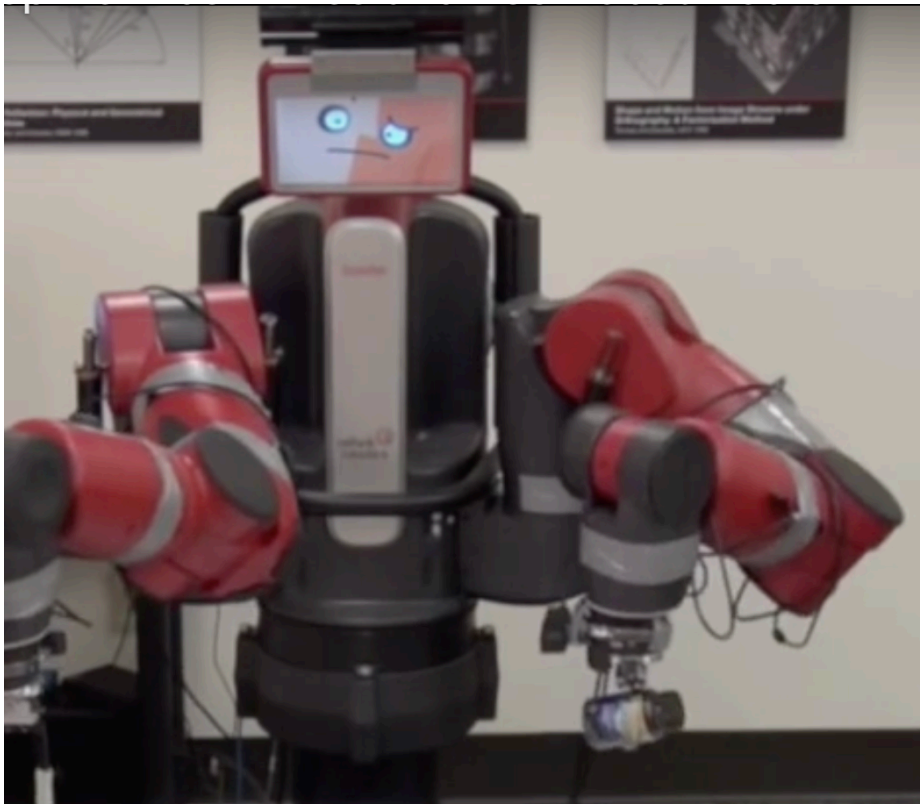


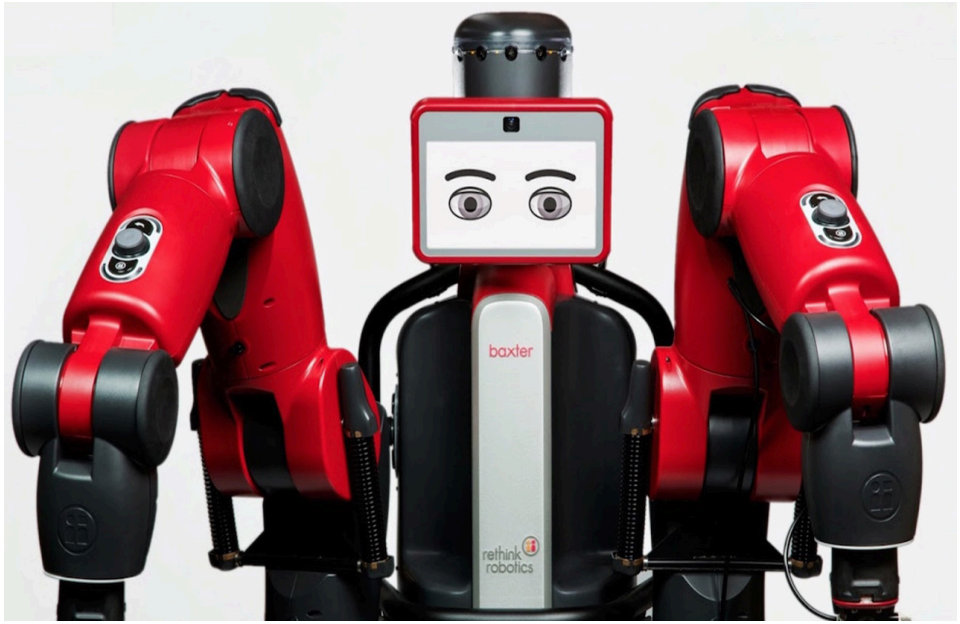
Image source: https://youtu.be/oSqHc0nLkm8?t=47

Related Work

- Unsupervised Learning
  - Other approaches use passive data
  - Here the robot actively plays with the objects
- Robotic Tasks
  - In robotics, we use vision to plan the best grasp
  - Here we use grasp data to classify the object

Approach

- Use robot's experiment results to label objects
- Train Network to predict experiment results given the picture of the object as input

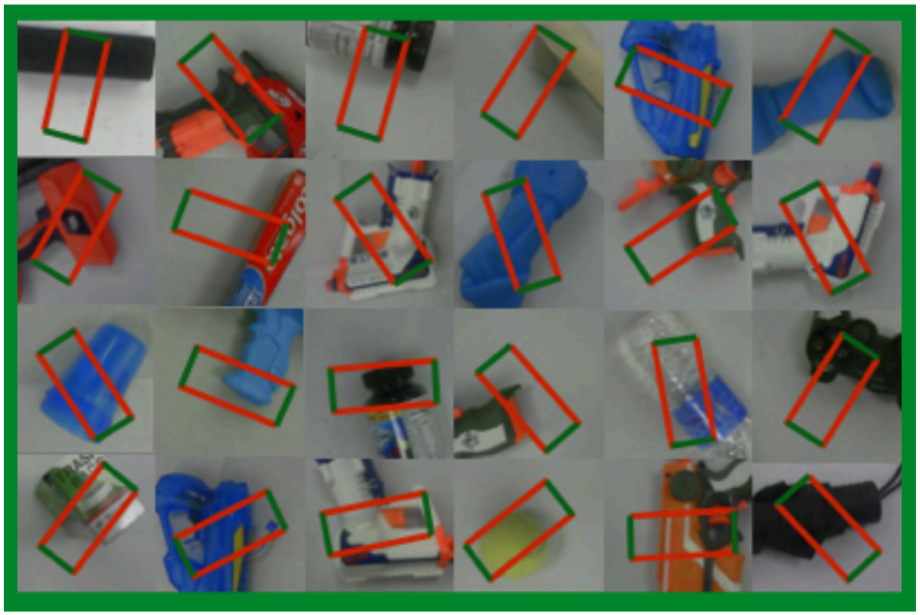


Experiments:
- Grasping
- Pushing
- Poking
- View at different angles

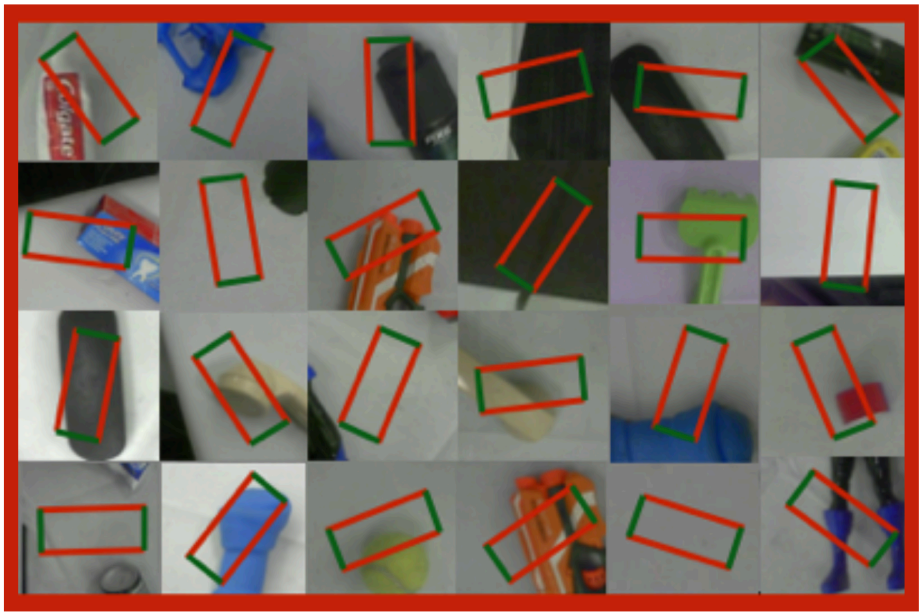
Image source: www.roboticsbusinessreview.com

Grasp

Successful grasps



Unsuccessful grasps



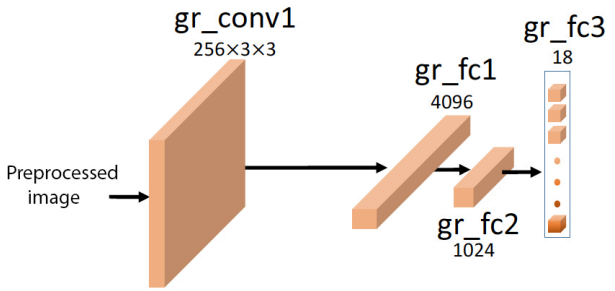
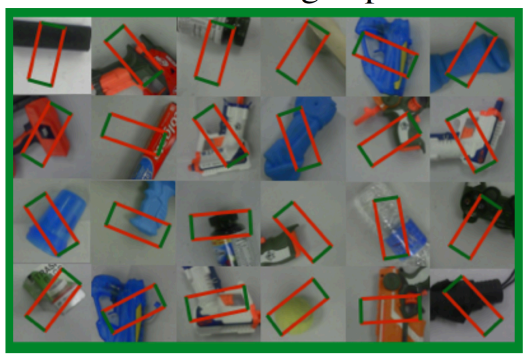


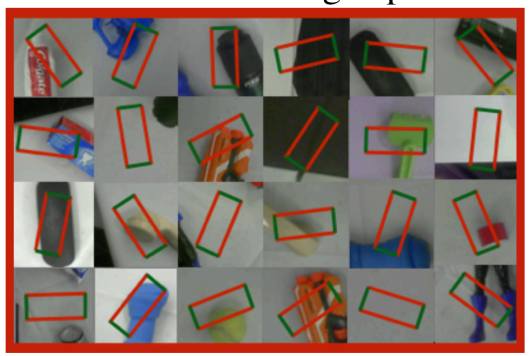
Image source: paper

Grasp

Successful grasps



Unsuccessful grasps



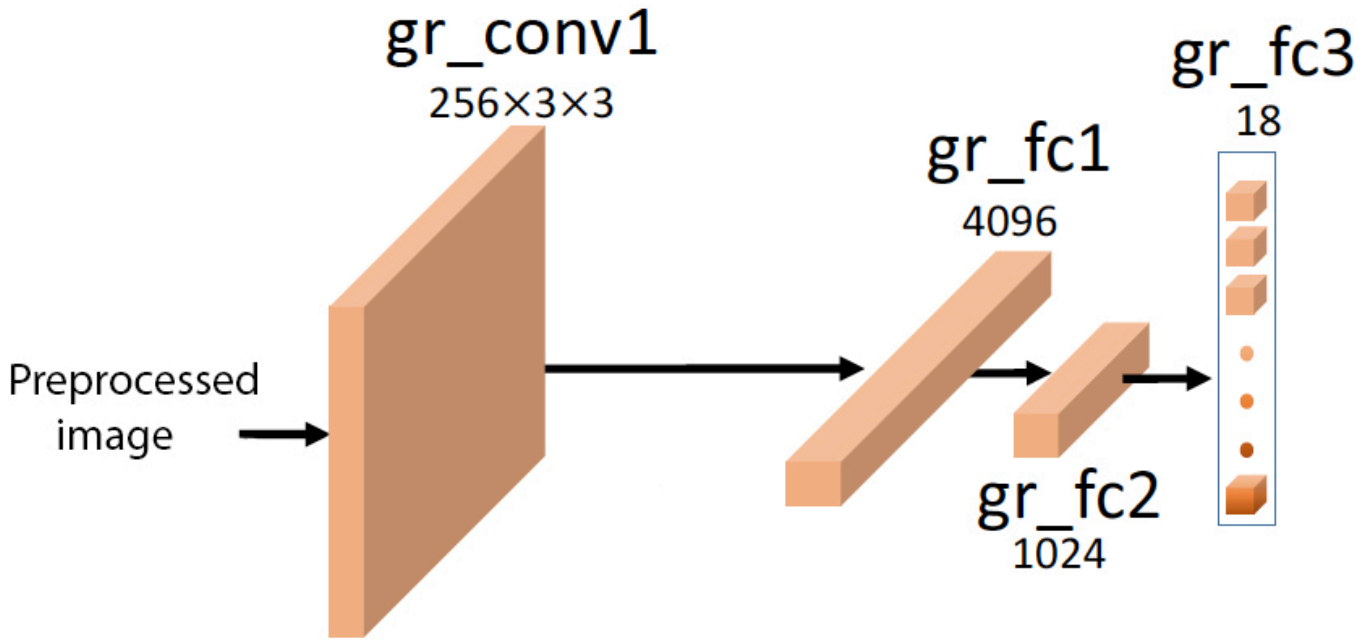
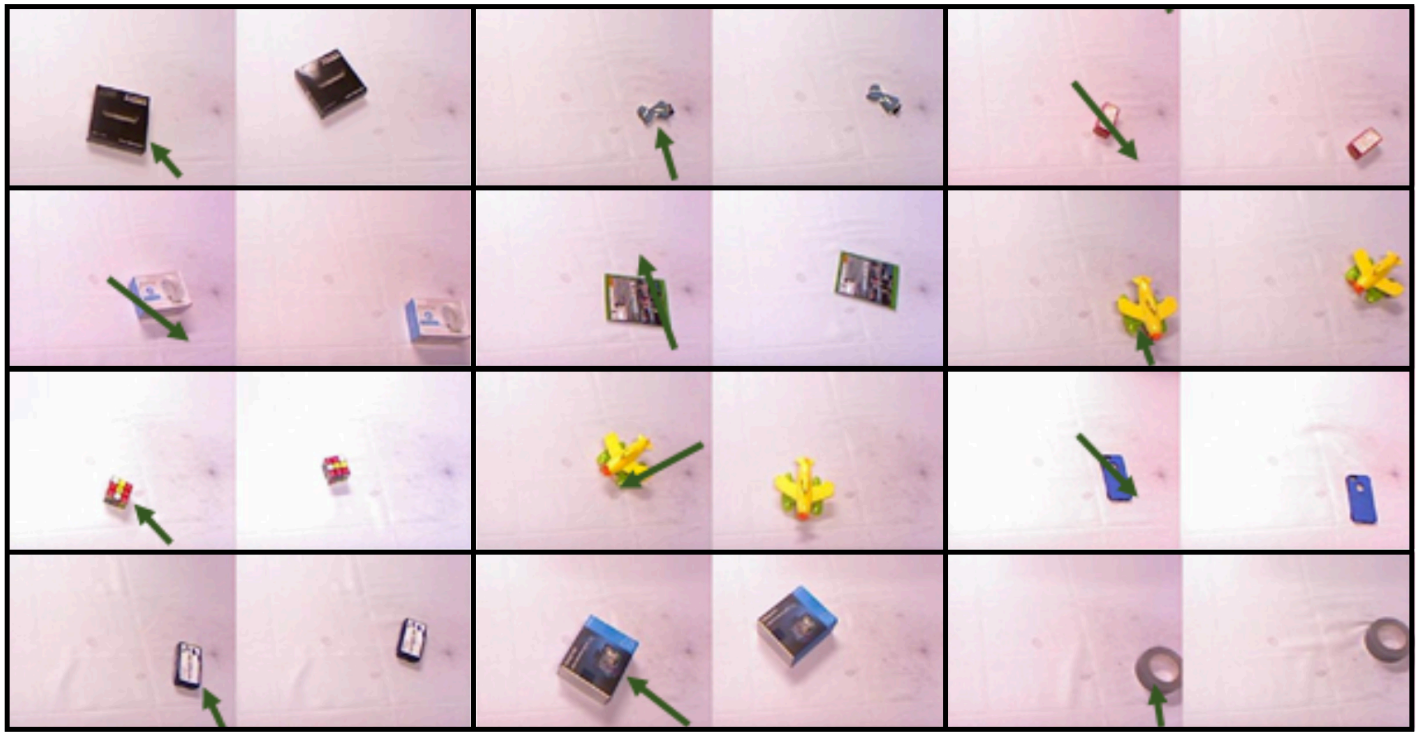


Image source: paper

Push

Objects and push action pairs



Initial state Final state Initial state Final state Initial state Final state

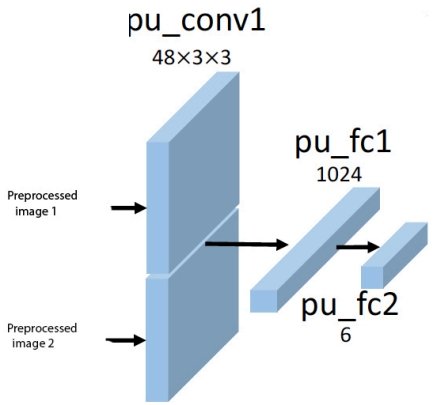
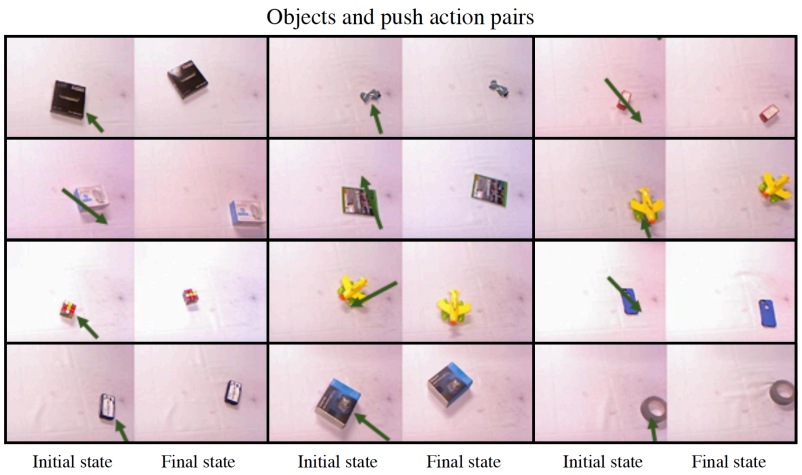


Image source: paper

Push



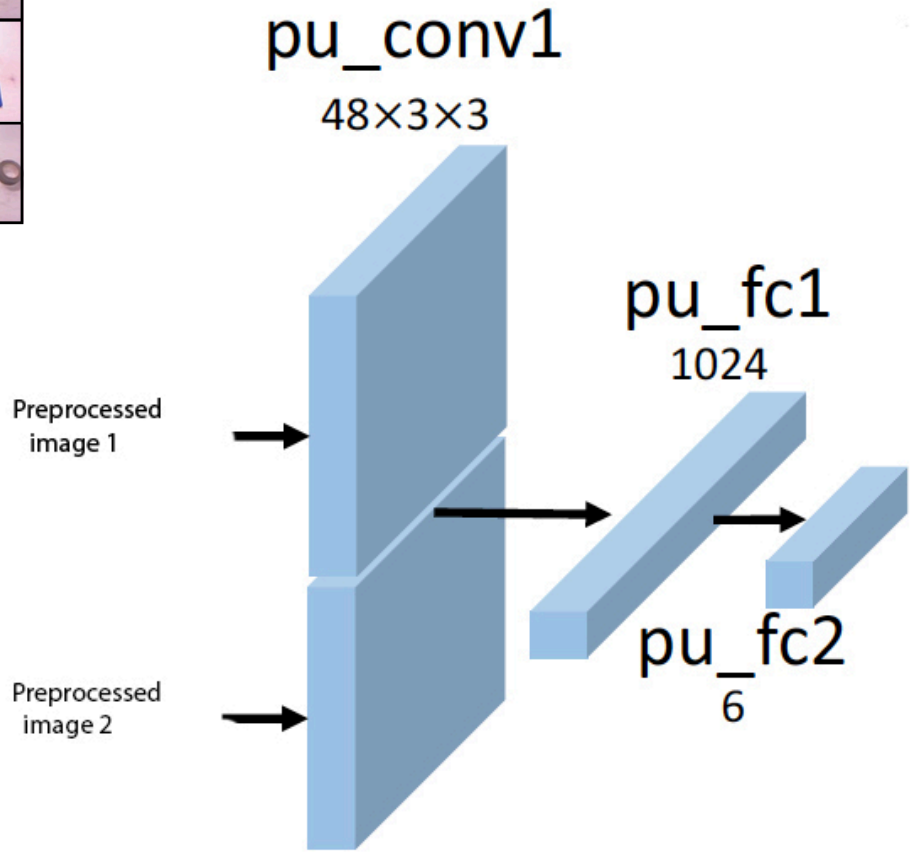
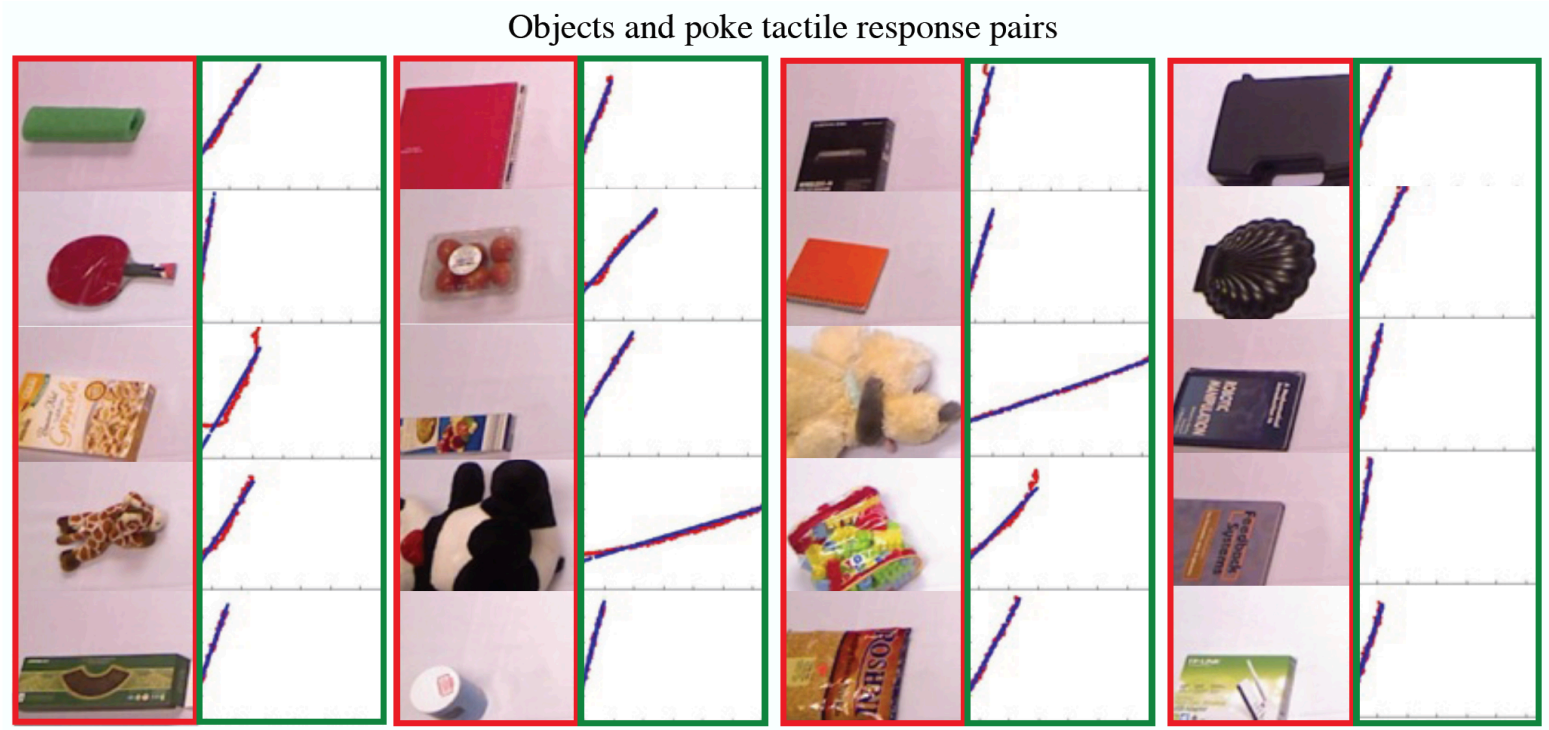


Image source: paper

Poke



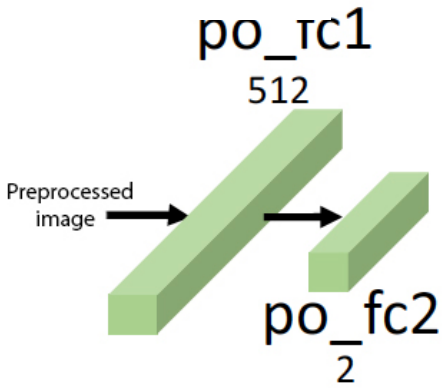
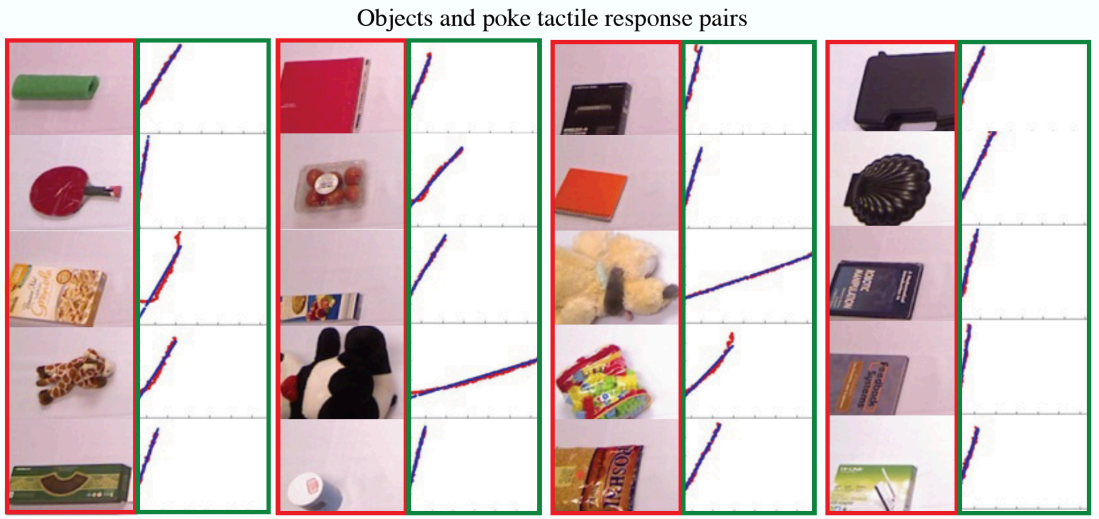


Image source: paper

Poke



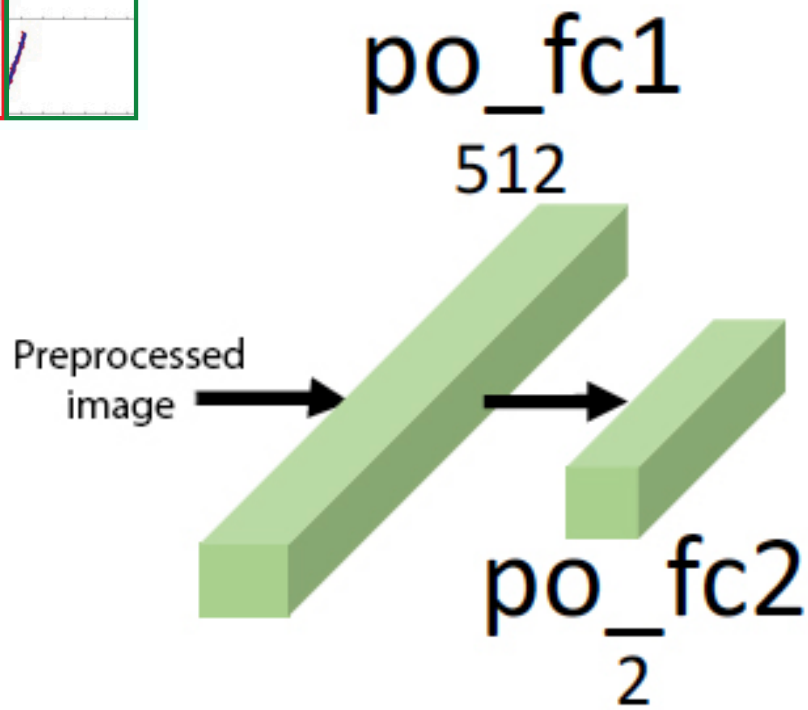
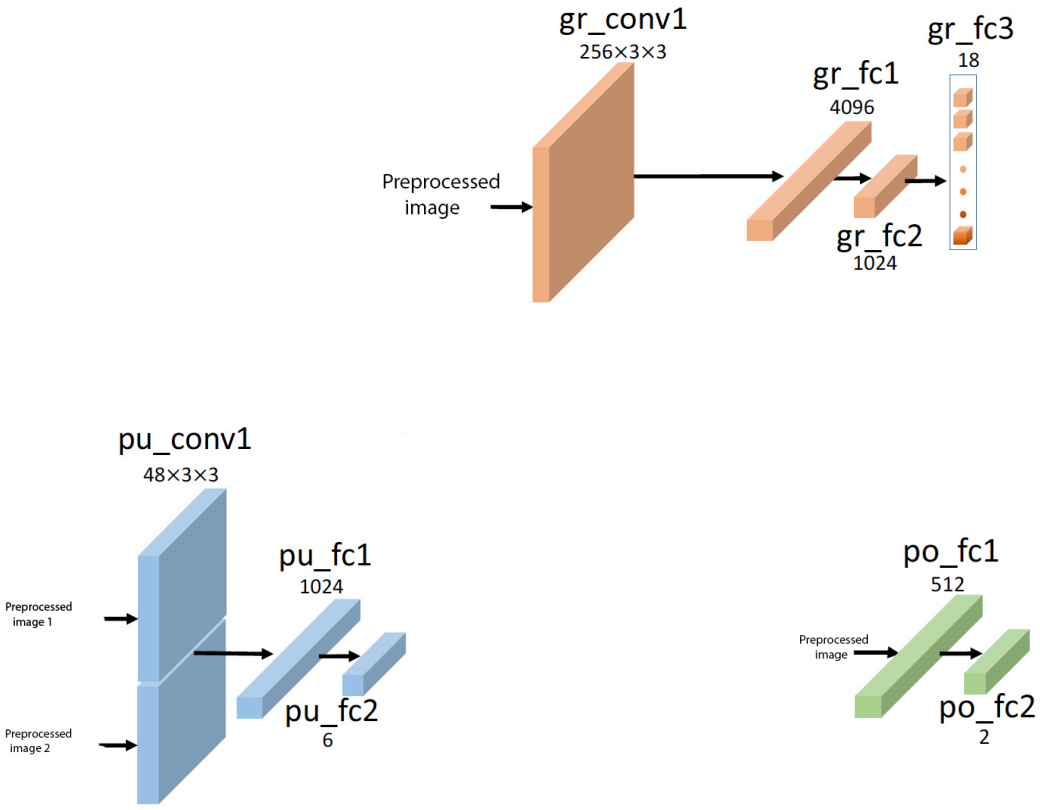


Image source: paper

Complete Network



Complete Network

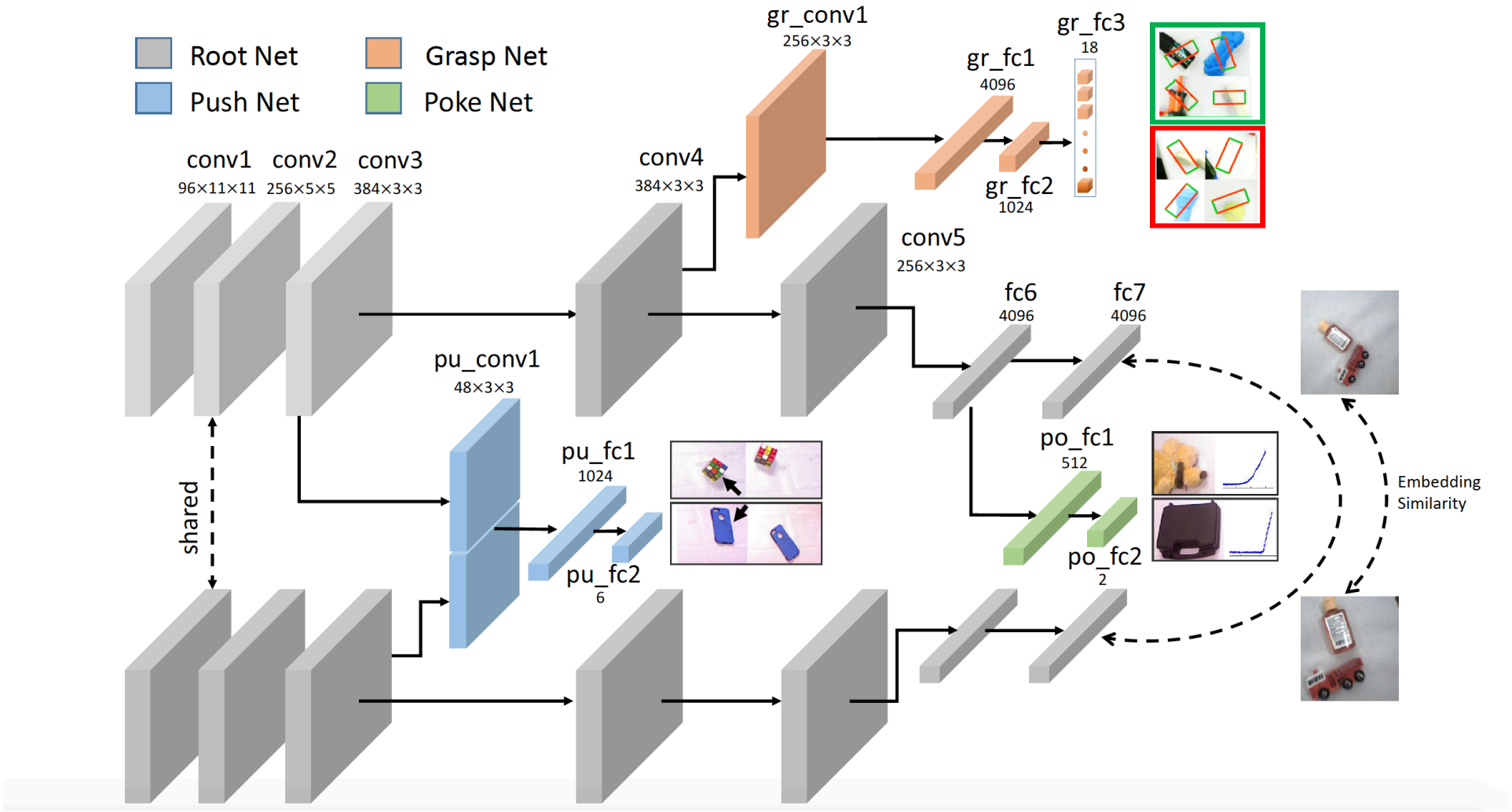


Image source: paper

Complete Network

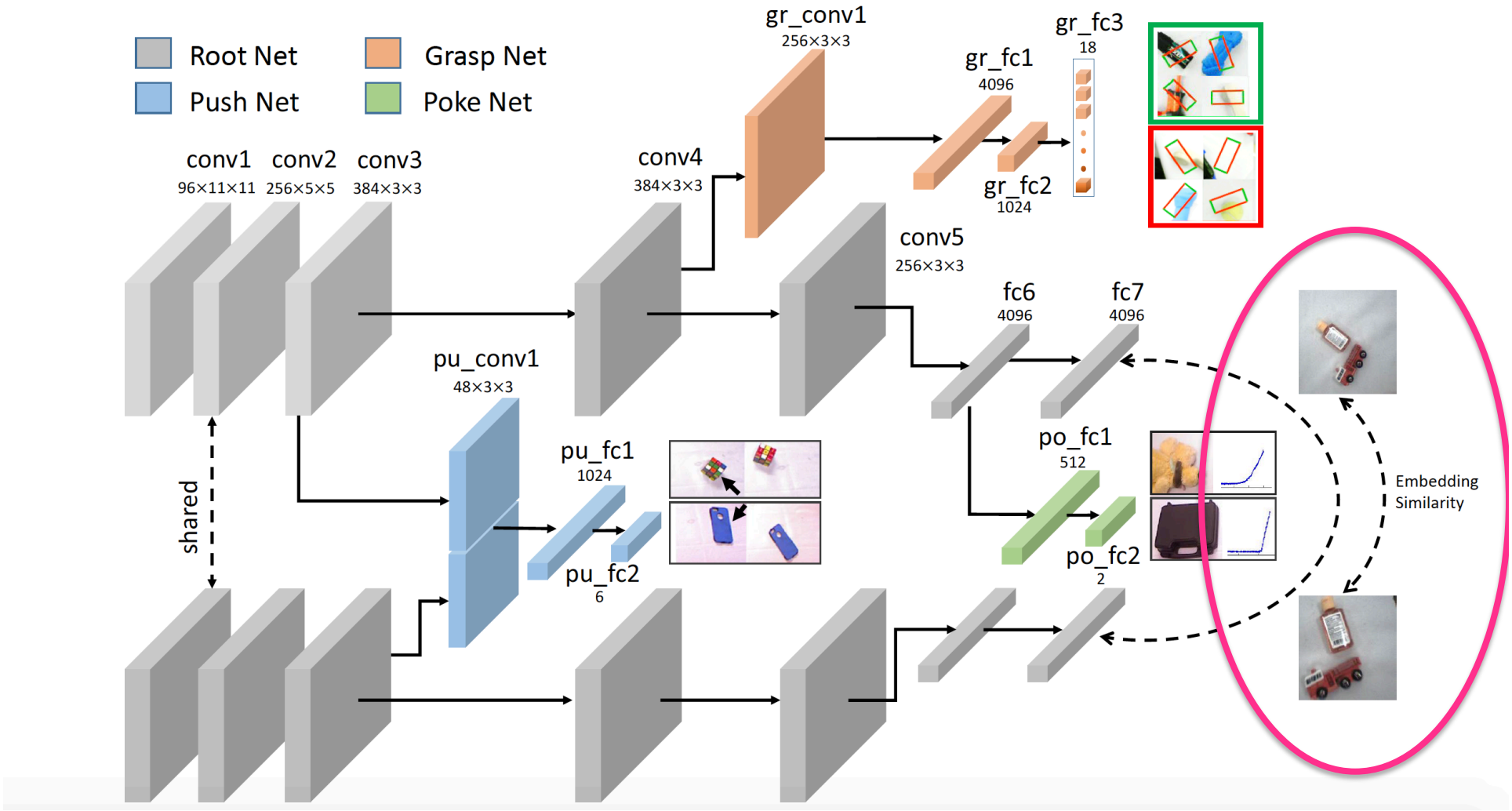


Image source: paper

Complete Network

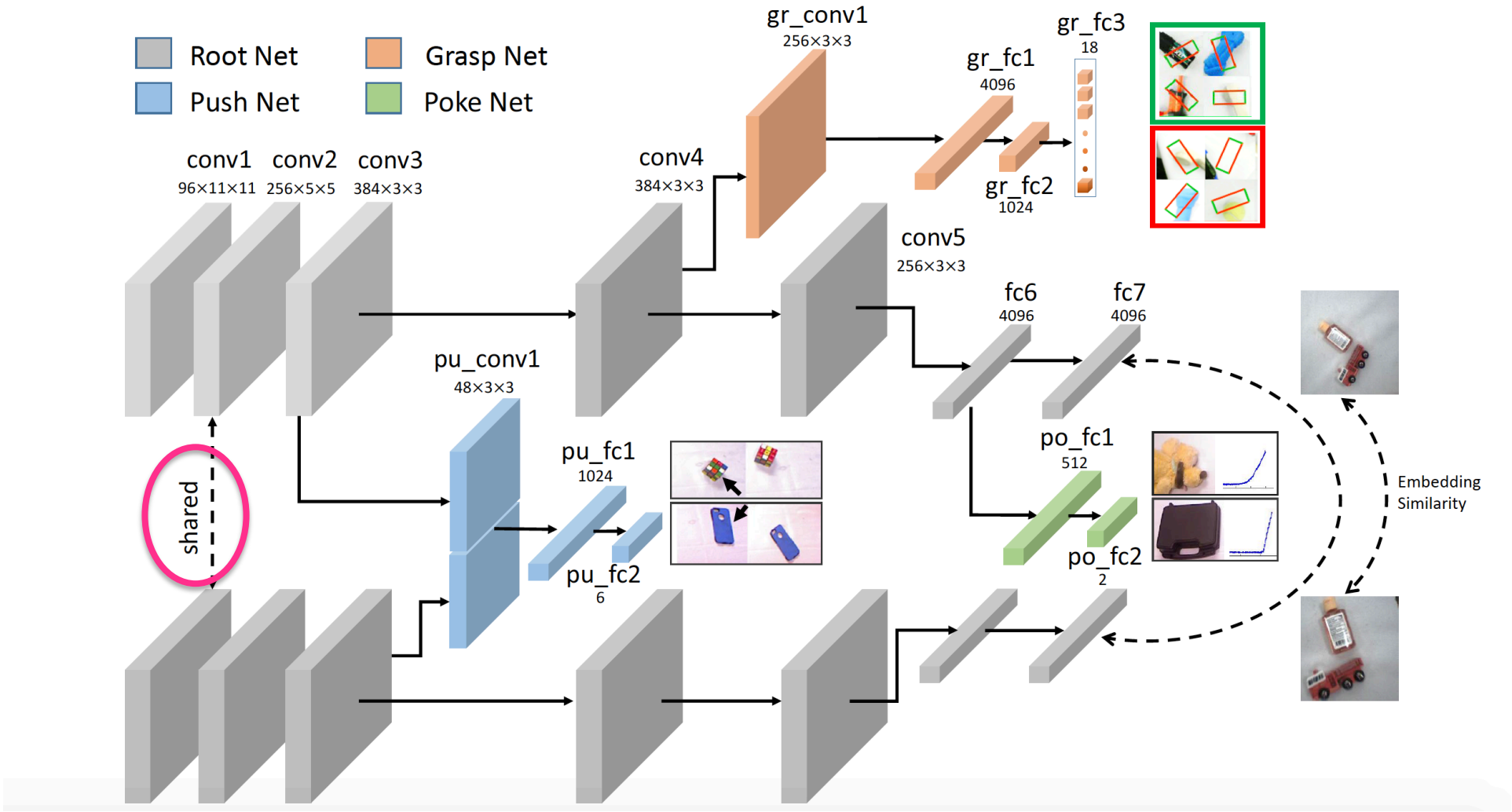


Image source: paper

Experiments – Root Network

Which images generate similar activation patterns in the Root network?

Experiments – Root Network

Network learns high level features of objects, such as shape.

Image source: paper

Experiments – Image Retrieval

	Instance level				Category level			
	k=1	k=5	k=10	k=20	k=1	k=5	k=10	k=20
Random Network	0.062	0.219	0.331	0.475	0.150	0.466	0.652	0.800
Our Network	0.720	0.831	0.875	0.909	0.833	0.918	0.946	0.966
AlexNet	0.686	0.857	0.903	0.941	0.854	0.953	0.969	0.982

Recall@k : this approach leads to good retrieval levels

Image source: paper

Experiments – Image Retrieval

Nearest Neighbors relies mostly on shape

Image source: paper

Discussion

- Strengths
  - Robot learns all by itself
  - Able to learn meaningful features (shape)
  - Very good retrieval results
- Weaknesses
  - Physical interaction is expensive
  - Network is heavily handcrafted
  - Limited to objects physically available in training

Extensions

- Allow for passive data to enrich dataset
- Actively choose which training data to gather
- Would this work with simpler network?